Marketing Science

Matching Value and Market Design in Online Advertising Networks: An Empirical Analysis

Chunhua Wu

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Matching Value and Market Design in Online Advertising Networks: An Empirical Analysis

Chunhua Wu

Sauder School of Business, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada,
chunhua.wu@sauder.ubc.ca

Advertising networks have recently played an increasingly important role in the online advertising market. Critical to the success of an advertising network are two mechanisms: an allocation mechanism that efficiently matches advertisers with publishers and a pricing scheme that maximally extracts surplus from the matches. In this paper, we quantify the value and investigate the determinants of a successful advertiser-publisher match, using data from Taobao's advertising network. A counterfactual experiment reveals that the platform's profit under a decentralized allocation mechanism is close to the profit level when the platform centrally assigns the matching under perfect platform knowledge of matching values. In another counterfactual experiment, we explore the effect of platform technology and revenue model on the strategic choice of the pricing schemes of list price versus generalized second price (GSP) auction pricing. We find that platforms that profit from the advertiser side may have less incentive to adopt GSP auction than platforms that profit from the publisher side.

Keywords: advertising network; matching game; maximum score estimation; generalized second price auction; platform design

History: Received: December 24, 2012; accepted: June 18, 2015; Preyas Desai served as the editor-in-chief and Nitin Mehta served as associate editor for this article. Published online in *Articles in Advance* September 15, 2015.

1. Introduction

Advertising networks, which provide marketplaces for advertisers and publishers,[1] are changing the game in online advertising. In a *Businessweek* article, Hof (2009) reports that, in 2008, 30% of the $8 billion online display advertising spending was through advertising networks. Google's Display Network, the world's largest ad network, contributed one quarter of Google's $51 billion advertising revenue in 2013.[2] An online advertising network differs from traditional advertising markets in that it has an independent intermediary that facilitates transactions between advertisers and publishers. In traditional advertising markets such as TV, print, and online banners, publishers and advertisers get connected by themselves, with high information asymmetries and high search costs. In the Web 2.0 age, well-known platforms, including Google and Taobao, create advertising networks that bring advertisers and publishers together. Numerous online publishers can sell spaces on these platforms that otherwise would go unreached by advertisers. Advertisers can also easily advertise through more publishers and reach a bigger audience from various fragmented segments at low costs.

The advertising platform plays a central role in an advertising network, with the core task of matching advertisers with publishers. A successful match may be determined by many factors, such as product category match, geographical match, demographic match, as well as content semantic match. Our first objective in this paper is to quantify the value of advertising and investigate the determinants of a successful advertiser-publisher match.

As an economic agent, the platform also extracts the economic surplus from the matched transactions to maximize its own profit. In practice, platforms differ in their revenue sources. A common business model for platforms is to share revenue with publishers. For example, Google gains 49% of the advertising revenue from its participating publishers in its display network; major Web 2.0 sites that rely on user-generated content (UGC), such as Facebook and Twitter, harvest 100% of the revenue. On the contrary, Taobao's Alimama, an advertising network that we study in this paper, aims to maximize advertisers' revenue. Instead of sharing profit with publishers, Taobao's model is to generate profit from advertisers' sales revenue in the retail marketplace while providing free services to advertisers and publishers in the advertising market. Conditional on the chosen revenue model, every advertising platform needs to make strategic decisions on two key mechanism designs that have direct impacts on its own profitability: the allocation mechanism that efficiently

---

[1] In this paper, "publishers" refers to online websites that provide content for the online-browsing audience and advertising spaces for advertisers.

[2] http://investor.google.com.

matches advertisers with publishers and the pricing scheme that determines the transfers in the matches.

Our second research objective relates to the mechanism designs of the advertising network. We explore whether and how an advertising platform's revenue model has direct impact on its market designs. In practice, different networks have divergent choices. Taobao adopts a simple market-based model in which publishers first announce list prices for each of their ad slots, and advertisers then self-select to purchase particular slots. By contrast, Google's display network is more complicated. On the publisher side, it uses the AdSense portal to assign specific keywords to each publisher based on the content of the publisher's Web pages. On the advertiser side, it sells the keywords-indexed publishers' ad spaces to advertisers under generalized second price (GSP) auction at the AdWords portal.[3] Taobao's ad network and Google's display network thus differ significantly on two dimensions: first, Taobao uses a purely decentralized allocation mechanism, whereas Google adopts a more centralized one. This is because Google centrally decides the associations between keywords that advertisers bid and publisher pages that would feature related ads at AdSense. In addition, Google centrally assigns advertisers' quality scores, which largely determine the equilibrium bids and allocations under the GSP auction. As a result, advertisers cannot directly select the exact Web pages on which they would like to place their ads on Google's display network. Second, the pricing mechanisms are also different: Taobao chooses a list price scheme, whereas Google chooses the GSP auction at AdWords in its display network. In this paper, we specifically investigate the profitability implications for Taobao under a more centralized allocation, which corresponds to the role Google plays in its display network. We also research the optimality of different pricing schemes and study how they are linked with the platform's revenue model. This exercise corresponds to the GSP mechanism that Google adopts at AdWords in its display network.

We estimate a structural model of a two-sided matching market using data from Taobao's advertising network. We first prove that, despite the complicated competitive relationship among advertisers and publishers in the advertising network, under general conditions an advertiser-publisher stable equilibrium exists in

Taobao's pricing scheme. We then estimate the matching function between advertisers and publishers using necessary conditions derived from this stable equilibrium in a maximum score estimator approach. Our results show that advertisers have very heterogeneous valuations on advertising, and the main determinants of the matching value function are product category and demographic matches, followed by content semantic match, while geographical match plays a minimal role.

Based on the model estimates, we use counterfactual simulations to address our second research objective. Given that ad allocation on Taobao is determined by each advertiser's and each publisher's self-interest, we first study the allocation outcome compared to a situation in which Taobao centrally allocates advertisers to publishers—a situation that explicitly maximizes platform profit. We find that the market-based allocation has generated a similar level of platform profit. A second policy simulation studies the optimality of various pricing schemes and the role of platform targeting technology. We manipulate the platform's technology as its ability to uncover the matching values in a GSP auction setting. We find that total advertiser revenue is more sensitive than total publisher revenue to the technology. In addition, total publisher revenue could be higher under GSP than under list price. Our results suggest that it is better for platforms such as Taobao with the objective consistent with maximizing total advertiser revenue to choose a list price scheme; however, platforms with the objective of maximizing total publisher revenue are better off under GSP when the targeting technology is better. This finding resonates with and provides a potential explanation on the different pricing strategies pursued by Taobao and Google display network.

This paper contributes to the literature on Internet advertising. Early papers on online display advertising study the performance of banners (Drèze and Hussherr 2003, Chatterjee et al. 2003, Manchanda et al. 2006). Recent papers investigate targeting issues in display advertising (Goldfarb and Tucker 2011, Zhang and Katona 2012). With the strong growth in search advertising, researchers recently have paid more attention to this area. Theoretical papers focus mainly on advertisers' bidding strategy for keywords and search engines' platform designs (Edelman et al. 2007, Varian 2007, Katona and Sarvary 2010). Empirical research explores diverse topics, such as the spillover dynamics among keywords (Rutz and Bucklin 2011), the complementarity of organic and sponsored search (Yang and Ghose 2010), the interplay of users, advertisers, and the search engine (Yao and Mela 2011), the competition among advertisers (Chan and Park 2015), and the value of customer acquisition through search advertising (Chan et al. 2011). Powerful advertising networks have emerged as a new phenomenon in the past few years.

---

[3] AdWords and AdSense are two portals for advertisers and publishers, respectively, under Google Ads. People conventionally use AdWords to refer to Google's search advertising business and AdSense to refer to Google's display advertising business. However, Google's display advertising also involves AdWords, because the ad spaces provided by publishers in its display network get sold under the GSP auction at the AdWords portal. AdWords thus serves both search advertising and display advertising. The same GSP mechanism is used for both markets under AdWords.

This paper, to the best of our knowledge, is the first empirical study on the matching effects and platform designs in an advertising network.

This paper is also closely related to the literature on matching. Theoretical papers on matching games date back to the 1960s. Examples include the "Gale-Shapley" algorithm for the college admissions problem (Gale and Shapley 1962), studies on matching between plants and locations (Koopmans and Beckmann 1957, Shapley and Shubik 1979), and matching in the marriage market (Becker 1973). Because of the complexity in matching games, empirical methods have been developed only recently. A few papers use the maximum likelihood approach in various settings (Choo and Siow 2006, Sørensen 2007, Hitsch et al. 2010). In general cases, a full likelihood approach can be computationally intractable, or the likelihood function may even be not well defined because of multiple equilibria. Fox (2010a) proposes a maximum score estimator method that uses only necessary conditions from equilibrium and can handle multiple equilibria. Fox (2010b) further gives identification proofs on the estimator. This method is applied in several recent studies on diverse topics, ranging from FCC spectrum auctions (Fox and Bajari 2013), to the NBA athletes market (Yang et al. 2009) and to office selections (Baccara et al. 2012). Our paper also uses the maximum score estimator method.

The rest of the paper is organized as follows. Section 2 describes the data and the business background of our empirical application. Section 3 introduces the model and estimation strategy. Section 4 shows the main estimation results, followed by policy experiments in §5. Section 6 concludes with directions for future research.

2. Data

2.1. Advertising Network on Taobao

Our data comes from Alimama.com, the advertising network affiliated with Taobao.com. Taobao is the world's largest online C2C retail platform for consumers and retailers. Although often perceived as China's eBay, Taobao is quite different from eBay in its business model. Taobao charges neither listing fees nor commissions. Instead, its revenue comes from two sources: sponsored keywords and display advertising on Taobao.com, and the interest return from the transaction revenue of its participating online retailers. To control online transaction fraud, Taobao has established a subsidiary payment company Alipay. Alipay withholds the retailer's revenue from each transaction until the buyer receives the product and confirms the transaction to be valid. The process usually takes a week or longer, enabling Alipay to gain interest from transaction revenue in the financial market.

To help its retailers attract more traffic from other websites, Taobao established the advertising network Alimama.com. Advertisers on Alimama are mainly the participating retailers on Taobao.com across all of the product categories. Publishers are mostly small- to medium-sized websites, including personal blogs, interest group pages, discussion forums, and small news portals. Publishers provide detailed descriptions of their websites and list daily prices of each slot that they are selling. Advertisers then make purchase decisions based on the list prices. If no one purchases a particular advertising slot, Taobao automatically assigns its own advertisement for that slot. Taobao provides rich and transparent information, including the publishers' characteristics, their daily traffic statistics, as well as the past transaction history for each slot (transacted advertisers and prices). Moreover, every publisher and advertiser can be reached easily through a real-time communication tool.

2.2. Sample and Summary Statistics

We collect information from the advertising network on January 1, 2011. The data record both advertiser and publisher characteristics, as well as ad slot prices. We select those advertisers and publishers who have been on the platform for more than one year and we further choose only those publishers with every slot priced at a minimum of 1.0 Chinese yuan (CNY hereafter) per day. This ensures that our sample consists of relatively more-experienced publishers who know how to set the right market price and advertisers who have high value for advertising. This selection process leaves us with a subset of 295 publishers with 992 slots and 483 advertisers. Total monthly retail revenue generated from those advertisers is 25.7 million CNY, which accounts for 80% of the total monthly revenue of the 880 advertisers who have purchased at least one advertisement on the platform that day. It is ideal to use data for the whole market in the estimation. However, since the estimation approach we use in this paper is based on the assumption of a stable market equilibrium, we want to exclude those relatively inexperienced advertisers and publishers. Besides, the estimation burden increases quadratically with the number of agents on each side. Our estimation on the selected sample represents the inferences on a submarket in which experienced advertisers and publishers transact.

Table 1 provides summary statistics on the size and characteristics of each side. On average, an advertiser advertises on 1.3 publishers, and a publisher provides 3.4 ad slots. Two-thirds of the slots are sold, resulting in a total transaction amount of 3,832 CNY. For the advertiser side, we have information on geographical location, product category based on Taobao's classification, and retail store performance (assortment size, average price, and monthly sales). Advertisers

Table 1   Summary Statistics of Data Sample

	Min	Median	Mean	Max	Total
Advertisers					483
Number of publishers advertised on	1	1	1.3	11	295
Number of slots advertised on	1	1	1.3	11	655
Total cost (CNY)	1.0	4.2	7.9	95.7	3,832
Number of items in the store	1	28	59	282	—
Average price of items	1	90	197	3,869	—
Monthly sales (thousand CNY)	0.06	13.3	53.2	3,019	25,720
Publishers					295
Number of slots provided	1	3	3.4	13	992
Total revenue from slots sold (CNY)	1.0	8.8	19.5	207.7	3,832
PageRank	0	1	1.4	6	—
Daily unique visits (thousand)	0.03	4.1	7.1	157	2,087
Average number of page views	1.43	3.13	3.52	10.92	—
Slots					992
Price	1.0	3.0	5.9	50.0	5,764
Area (10,000 square pixels)	0.7	3.9	5.0	9.6	4,998
Number of competing slots	1	7	7.7	18	5,647
Slot position	0.02	0.35	0.44	0.99	—

differ significantly in their store characteristics. The median assortment size of the stores is 28 items, and the median average price across these stores is 90 CNY. Advertisers' median monthly revenue is 13.3 thousand CNY, whereas the maximum is 3 million CNY, 230 times larger than the median. For the publisher side, our variables include geographical location, publisher category based on Taobao's classification, targeted demographics of the website (gender, age, and income), website visits metrics (PageRank, daily unique visits, average page views), and advertising slots (number of slots, position of slots, prices). Publishers differ significantly in their ability to attract visitors: the minimum number of daily visits for a publisher is 30, whereas the maximum is as high as 157,000. The listed slot prices range from 1.0 CNY to 50.0 CNY per day. Ad slots also differ in sizes, number of competing slots, and positions, as summarized in Table 1.

2.3. Data on the Matching of Advertisers and Publishers

We now present some data on the factors that influence an advertiser's purchasing decision for advertising slots. We first explore whether semantic correlations between advertising messages and publishers' Web content would determine advertiser-publisher matches and find a marginal effect based on a measurement using the latent semantic analysis algorithm (Landauer et al. 1998). We also investigate the effect of the bilateral geographical distances (at the city level) between advertisers and publishers, and we find no significant differences between the transacted advertiser-publisher matches and the rest of the potential matches. Finally, we look at the frequency distribution of matches based

on other characteristics of advertisers and publishers. We specifically look at how the advertiser's product category relates to the publisher's targeted demographics in a transaction. Table 2 provides the tabulation. In this table and in subsequent analysis, we group the advertisers into five product categories: men's products (mainly clothes and shoes), women's products (mainly clothes and shoes), digital products, foods, and household items (mainly furniture). We also group publishers into five categories based on the content of their websites: fashion, life information, news portal, online shops/services, and entertainment/other.[4] Strong evidence of selective matching is shown in this table. First, at the category level, advertising slots on fashion-related websites are sold mainly to advertisers selling women's products. Also, digital products retailers tend to purchase advertisements on entertainment-related websites. Evidence of matching based on demographics is also observed: women's products are rarely targeted to websites with mainly male users; retailers of men's products, women's products, and digital products prefer the younger population, whereas the older population is preferred by sellers of foods and household products. These statistics provide evidence on selective matching between advertisers and publishers through the market mechanism at Taobao. To fully quantify the impacts of advertiser and publisher characteristics on the market outcome, we develop a structural matching model that quantifies the economic value created by the matching of advertisers and publishers.

3.   Model

We begin this section by discussing the equilibrium concept in an advertiser-publisher matching game under general functional forms and prove that a stable equilibrium exists. We then specify a functional form that can be used for empirical estimation and also discuss the estimation strategy.

3.1.   A Conceptual Framework of Stable Matching Equilibrium

We model the market transactions of advertising slots as a many-to-many matching game with transferable utility, in which advertisers (A) and publishers (P) compete among themselves on each side of the market. This game is complicated because it involves numerous differentiated advertisers and publishers. Advertising on a publisher's website may bring one advertiser a high profit but another advertiser a low profit, depending on the matching value. Market equilibrium may not exist, and even when it does, it may not be unique.

---

[4] Taobao has 52 main categories and more than 4,000 subcategories for stores (advertisers in our setting) and 23 categories and 125 subcategories for publishers.

Table 2   Tabulation for Observed Matches by Publisher and Advertiser Characteristics

	Advertiser category						
	Men's products	Women's products	Digital products	Foods	Household items	Unmatched	Total
Publisher category							
Fashion	38	165	9	16	14	62	304
Life information	10	39	10	12	50	66	187
News portal	16	25	15	14	9	66	145
Online shops/services	15	36	16	12	15	60	154
Entertainment/other	11	15	41	14	38	83	202
Demographics							
Gender							
Mainly male	18	8	31	10	18	45	130
Mainly female	27	172	17	19	24	100	359
Undifferentiated	45	100	43	39	84	192	503
Age							
≤ 25 years old	60	161	55	27	46	167	516
>25 years old	30	119	36	41	80	170	476
Income							
$\leq$CNY 3,000	51	135	54	31	51	163	485
$>$CNY 3,000	39	145	37	37	75	174	507
Total	90	280	91	68	126	337	992

Note. Numbers in cells are frequency counts; a pair of match is defined at advertiser-slot level; and "unmatched" means the slot is not sold in this period (January 1, 2011).

We will formulate the conditions under which the existence of a market equilibrium can be proved.

We define an allocation to be the matching of advertisers to advertising slots, with the constraint that each advertising slot can be matched to at most one advertiser. Denote the set of advertisers as A, the set of publishers as P, and the set of advertising slots as S. Let $\mathbb{M}$ be the collection of all possible allocations of advertisers to advertising slots. An element $\mathcal{M} \in \mathbb{M}$ is a specific allocation, where $\langle i, k_j \rangle \in \mathcal{M} \subset A \times S$ denotes the specific match of advertiser i to publisher j, where the kth slot is assigned in the allocation. A corresponding vector $\mathcal{P}$ denotes the listed daily prices of all advertising slots from publishers, where an element $p_{k_j} \in \mathcal{P}$ is the price paid to publisher j for the kth slot. Denote $V_i(\mathcal{M})$ as advertiser i's total revenue function from advertising on the slots it obtains in the allocation $\mathcal{M}$. We define an advertiser's total profit as the difference between $V_i(\mathcal{M})$ and the total advertising cost of the slots it purchases. That is,

$$\pi_i = V_i(\mathcal{M}) - \sum_{k_j: \langle i, k_j \rangle \in \mathcal{M}} p_{k_j}. \qquad (1)$$

A publisher's profit is the sum of prices for each slot it sells: $\Pi_j = \sum_k p_{k_j} I[\sum_i I[\langle i, k_j \rangle \in \mathcal{M}] = 1]$, where $I[\cdot]$ is an indicator that takes value 1 if the statement is true and 0 otherwise. If $I[\langle i, k_j \rangle \in \mathcal{M}] = 1$, slot k_j is sold to advertiser i. Any slot can be sold to only one advertiser; therefore, $\sum_i I[\cdot]$ can be at most equal to 1, and if k_j is not sold, the sum is 0.

We assume that $V_i(\mathcal{M})$ satisfies three properties:
- Independent: $V_i(\mathcal{M}) = V_i(\mathcal{M}_i)$,

$\mathcal{M}_i = \{\langle i, k_j \rangle \mid \langle i, k_j \rangle \in \mathcal{M}\}$.
- Monotonic: $V_i(\mathcal{M} \cup \mathcal{M}') \geq V_i(\mathcal{M})$, $\forall \mathcal{M} \cap \mathcal{M}' = \varnothing$.
- Nonincreasing marginal return:

$V_i(\mathcal{M}' \cup \mathcal{M}) - V_i(\mathcal{M}') \geq V_i(\mathcal{M}'' \cup \mathcal{M}) - V_i(\mathcal{M}'')$, $\forall \mathcal{M}' \subset \mathcal{M}''$, $\mathcal{M} \cap \mathcal{M}' = \varnothing$, $\mathcal{M} \cap \mathcal{M}'' = \varnothing$.

The independent assumption implies that an advertiser's value for the set of slots purchased does not depend on the allocation of other slots. An advertiser's value should not depend on another's if none of the advertisements from the two would be shown together; however, the value may be impacted by the competing slots on the same publisher's webpage. We assume that the value is independent of the *identity* of advertisers matched to the same publisher.[5] Yet, we allow potential demand effects from neighboring slots, which could be captured in slot-level attributes. The assumption is consistent with the literature on search advertising, where it is commonly assumed that valuation is affected by the position of the advertisement, but not by the identity of the competitors (Varian 2007, Edelman et al. 2007).

The monotonic property is a natural assumption that requires nonnegative marginal valuation. The nonincreasing marginal return property is a general

---

[5] In an advertising network, there are numerous players on each side making simultaneous decisions. Thus, who the direct competitors are is hard to tell a priori. Any coordination to avoid direct competition in the face of the substitution effect or any coordination to bundle when there is a complementarity effect would be difficult.

condition that captures two institutional realities. First, it is a common practice that an advertiser can occupy only one position on the same publisher's page in most of the advertising networks.[6] Mathematically, this is equivalent to assuming perfect substitution between slots on the same publisher's page—i.e., $V_i(\langle i, k_j \rangle \cup \langle i, k'_j \rangle) = \max\{V_i(\langle i, k_j \rangle), V_i(\langle i, k'_j \rangle)\}$. Second, each publisher could be viewed as a unique market that refers its specific audience to the advertisers. Thus, the marginal return from advertising on a new publisher is likely to be a constant (neither increasing nor decreasing) that does not depend on the current matching set. In other words, the utility function has a linear additive specification, $V_i(\mathcal{M} \cup \langle i, k'_{j'} \rangle) = V_i(\mathcal{M}) + V_i(\langle i, k'_{j'} \rangle)$, $\forall k$, $\langle i, k_{j'} \rangle \notin \mathcal{M}$. Our nonincreasing marginal return assumption is a more general condition that captures both of the above effects. It is also mathematically attractive in proving the existence of equilibrium. In our empirical specification, we assume away potential competition across publishers and use a constant marginal return function for model tractability purposes.

We also assume that every advertiser has an advertiser-specific outside option "*o*" for each purchase decision. The outside option refers to other marketing opportunities, such as search advertising, email advertising, and other offline channels. For each rational advertiser to purchase a slot k_j, the marginal contribution from this slot must be higher than the price plus the value of the outside option, V_{i0}—that is, $V_i(\mathcal{M}_i) - V_i(\mathcal{M}_i \backslash \langle i, k_j \rangle) \geq p_{k_j} + V_{i0}$.

The information structure of the game is as follows. We assume that the properties of the publishers are common knowledge to advertisers and that each advertiser knows its own value function $V_i(\mathcal{M}_i)$. In addition, there are no search costs for advertisers. In our specific context, the platform provides detailed information about each advertising space from each publisher, including the average number of impressions and clicks, characteristics of the audience, and the transaction history. An advertiser can easily click into the link to examine potential content and audience matches. In addition, it is quite easy for an advertiser to experiment with purchasing different slots and examine the performance. Moreover, in our empirical application, we use a sample of more-experienced advertisers, who presumably know even more about their individual advertising value and have a lower search cost because of the familiarity with the platform's functions and possible past experiments. Thus, we believe that our information assumption is a good approximation to the reality.

We propose an equilibrium concept of *advertiser-publisher (A-P) stable allocation*. We define an allocation as advertiser-publisher stable if the advertisers and publishers in the current allocation have no incentives to deviate.[7] An advertiser-publisher stable allocation is in the core of this advertiser and publisher many-to-many matching game. Two types of deviation from $\mathcal{M}$ and $\mathcal{P}$ can potentially be profitable. First, as in the classic many-to-many matching game, either the advertiser or the publisher in a current match can deviate by partially exiting the market. For example, the advertiser can stop purchasing any slot from the publisher, and the publisher can refuse the purchase from the advertiser.[8] Second, an advertiser can reallocate its budget to a specific set of advertising slots if it can make every publisher who sells those slots strictly better off. We define the stable allocation equilibrium based on the conditions that neither of the above deviations would be profitable.

DEFINITION 1 (ADVERTISER-PUBLISHER STABLE ALLOCATION). An A-P stable allocation consists of an advertiser-slot allocation $\mathcal{M}$ and a price vector $\mathcal{P}$ that satisfies two conditions:

• *Individual rationality*. Under the stable allocation $\mathcal{M}$, no advertiser can be better off by partially exiting the market, and no publisher can be better off by partially exiting the market. Formally, $\forall \langle i, k_j \rangle \in \mathcal{M}$, $V_i(\mathcal{M}_i) - V_i(\mathcal{M}_i \backslash \langle i, k_j \rangle) - p_{k_j} \geq V_{i0}$ and $\forall k_j, p_{k_j} \geq 0$.

• *Incentive compatibility*. There does not exist an advertiser i, a new allocation $\mathcal{M}'$, and a price vector $\mathcal{P}'$, such that

—$V_i(\mathcal{M}'_i) - \sum_{\delta:\langle i, \delta \rangle \in \mathcal{M}'} p'_\delta \geq V_i(\mathcal{M}_i) - \sum_{\delta:\langle i, \delta \rangle \in \mathcal{M}} p_\delta$.

—$\forall j \in \{j \mid \langle i, k_j \rangle \in \mathcal{M}'\}$, $\sum_{k:\exists a, \langle a, k_j \rangle \in \mathcal{M}'} p'_{k_j} \geq \sum_{k:\exists a, \langle a, k_j \rangle \in \mathcal{M}} p_{k_j}$.

—Strict inequality holds for at least one condition.

That is, when advertiser i purchases the set of advertising slots specified in the allocation $\mathcal{M}'$ at prices specified in the price vector $\mathcal{P}'$, the advertiser gets better off, and all of the publishers who the set of slots corresponds to make higher profits.

The individual rationality condition says that neither the advertiser nor the publisher in a stable match can get better off by partially exiting the market. The incentive compatibility condition states that for any new allocation $\mathcal{M}'$ and price $\mathcal{P}'$, there must exist an advertiser-publisher matched pair, such that either the advertiser or the publisher is worse off than in

---

[6] On Taobao, although an advertiser is not constrained to purchase at most one slot from a publisher, in practice, less than 2% of them purchase multiple slots from the same publisher. We exclude these advertisers and the corresponding publishers from our sample.

[7] It is worth noting that although the allocation is defined at the advertiser-slot level, each slot should not be treated as an independent player. This is because publishers maximize total profits from slot bundles rather than maximize profit from each individual slot.

[8] In practice, even if the publisher does not have the option to refuse any purchase request, he can always achieve this by raising the price to a high enough level.

the current stable allocation. The following existence result can be proved for a general V_i function, which is independent, monotonic, and has nonincreasing marginal return.

PROPOSITION 1. *The A-P stable allocation exists for the advertiser-publisher many-to-many matching game.*

Furthermore, when an advertiser's value from different advertising slots is linear additive (constant marginal return)—that is, $V_i(\mathcal{M} \cup \langle i, k'_{j'}\rangle) = V_i(\mathcal{M}) + V_i(\langle i, k'_{j'}\rangle)$, $\forall k, \langle i, k_{j'}\rangle \notin \mathcal{M}$. We obtain the following result:

PROPOSITION 2. *Under linear additive value function, an A-P stable allocation is Pareto optimal.*

The proofs of the propositions are outlined in Online Appendix A (available as supplemental material at http://dx.doi.org/10.1287/mksc.2015.0944). Proposition 1 shows that an equilibrium A-P allocation exists. But in general, it may not be unique. A stable equilibrium is assumed to be the market outcome observed in our data, perhaps after an initial period of experimentation with pricing and matching between advertisers and publishers at the platform. Proposition 2 argues that an A-P stable allocation is Pareto optimal under a linear-additive value function assumption.[9] However, it is worth noting that Pareto optimality does not necessarily imply maximization of market efficiency—i.e., joint profit for advertisers and publishers. Alternative mechanisms that improve the joint profit may exist, and we explore the possibility in our empirical analysis.

3.2. Advertiser Valuation Function
As mentioned in the discussion of the advertiser value function assumptions, in our empirical application, we constrain the marginal return of an additional advertising slot from a new publisher to be constant and from an existing publisher to be decreasing. Formally, $V_i(\langle i, k_j\rangle \cup \langle i, k'_{j'}\rangle) = V_i(\langle i, k_j\rangle) + V_i(\langle i, k'_{j'}\rangle)$ and $V_i(\langle i, k_j\rangle \cup \langle i, k'_j\rangle) = \max\{V_i(\langle i, k_j\rangle), V_i(\langle i, k'_j\rangle)\}$. We next specify the functional form of an advertiser's value. Let $V_{ik_j} = V_i(\langle i, k_j\rangle)$ be the value when advertiser i advertises through the kth slot of publisher j. The value comes ultimately from the revenue of product sales referred by advertisements, which is specified as

$$
\begin{aligned}
V_{ik_j} =\ & \text{Impressions} \times \Pr(\text{click} \mid \text{impression}) \\
& \times \Pr(\text{purchase} \mid \text{click}) \times E(\text{value} \mid \text{purchase}) \\
=\ & EI_j \times CTR_{ijk} \times CR_{ij} \times v_i \\
=\ & EI_j \times PE_k \times \overline{CTR}_{ij} \times CR_{ij} \times v_i,
\end{aligned}
\tag{2}
$$

where EI_j is the expected impressions of publisher j. We use expected impressions instead of raw number

of visits because a visit is not necessarily equal to an impression, especially because people may intentionally avoid advertisements when they browse Web pages (Cho and Cheon 2004). The next component, CTR_{ij} is the click-through rate, which generally depends on the match between audience (j) and advertisement (i) and also the position of the advertisement. Thus, we further model it as the product of two components: a baseline click-through rate $\overline{CTR}_{ij}$ and a positional effect PE_k. This formulation implicitly assumes that the positional effect does not depend on the identity of the advertisers, which is the same assumption used in the sponsored search advertising literature (Varian 2007, Edelman et al. 2007). The next component, CR_{ij} is the conversion rate, which we assume depends on audience (j) and product (i). Finally, v_i is the expected transaction value conditional on conversion. Since we observe neither click-through rate nor conversion rate, we combine the CTR_{ij} and CR_{ij} parts that both depend on the match between the advertiser and the publisher and relabel it as m_{ij}, which stands for the *matching effect*. It represents the value created when a specific pair of advertiser and publisher matches together. This results in the reformulated value function of $V_{ik_j} = EI_j \times PE_k \times m_{ij} \times v_j$, with each component modeled in a log-linear fashion

$$
\begin{aligned}
\ln EI_j &= Z_j \gamma + \nu_j, \\
\ln PE_k &= Y_k \delta + \kappa_k, \\
\ln m_{ij} &= W_{ij} \alpha + \varepsilon_{ij}, \\
\ln v_i &= X_i \beta + \mu_i.
\end{aligned}
$$

Thus, the final valuation equation is

$$
V_{ik_j} = \exp(Z_j \gamma + Y_k \delta + W_{ij} \alpha + X_i \beta + \nu_j + \kappa_k + \mu_i + \varepsilon_{ij}). \tag{3}
$$

Variables included in each part of the value function are listed in Table 3. For advertiser effect X_i, we include the log of assortment size, the log of average prices, and the log of monthly sales. For publisher effect Z_j, variables include a PageRank score,[10] the log of daily unique IP visits, and the log of the average number of page views. Unique IP visits is an approximation for the number of unique individual exposures to the website. Average page views measures how many pages people would view on the particular website, with a high value for paying more time and also possibly a larger marketing opportunity. The advertisement's positional effect Y_k includes variables of advertisement size, an indicator of whether the advertisement would appear on the main page, and the log number of pages on which

---

[9] The concept of Pareto optimality in this paper refers to the welfare of advertisers and publishers, but not the consumers.

[10] PageRank is a measurement for the importance of websites on the World Wide Web. We use the scores calculated by Google, which are integers from 0 to 10, with a high value representing high importance. For a detailed description, see http://en.wikipedia.org/wiki/PageRank.

Table 3   Variables Included in the Analysis

Variable	Description
X_i	
items	Log of advertiser's assortment size
avg_price	Log of advertiser's average price across products
month_sales	Log of advertiser's monthly product sales (in thousand CNY)
Z_j	
pagerank	PageRank score as defined by Google (0–10)
unique_ip	Log of daily unique IP visits (in thousand)
page_view	Log of average number of page views
Y_k	
ad_size	Log of the square root of advertisement area (in 100 pixels)
ad_mainpage	Indicator of whether the advertisement is on the main page
ad_pages	Log of the number of pages the advertisement would appear
n_slots	Log of the number of publisher's advertisement slots
ad_position	The position of the advertisement, measured as the ratio of the line number to total lines in the raw html file (0–1)
m_{ij}	
content	Contextual correlation between a publisher's website and an advertiser's website from latent semantic analysis (−1, 1)
geography	Log of city level geographical distance between a publisher and advertiser
category	Indicators for bilateral relationship between advertiser's category and publisher's category (5×5 indicators)
demographics	Indicators of bilateral relationship between advertiser's category and publisher's targeting demographics (gender 3×5, age 2×5, income 2×5 indicators)

advertisement will show up. To capture the potential competition effect of neighboring advertisements, we use a reduced-form fashion by including the log number of slots the publisher offers and also a relative position measurement that indicates the location of the ad slot. Because the value of advertising comes mainly from the attention of browsers, we believe that these two measurements directly affect the attention level and thus capture the most important essence of the potential competition effect.[11] Last, the matching effect m_{ij} includes those variables discussed in the previous section: a correlation measuring semantic relevance between the advertiser and the publisher, the log of the geographical distance between the two, and the dummy variables for bilateral relationship categorization between product categories and audience demographics. The stochastic components (ν_j, κ_k, ε_{ij}, μ_i) are unobservable to researchers. Advertisers know these values perfectly when they purchase ad slots. We assume that each stochastic component is

---

[11] Although it is difficult to verify the degree to which the potential competition effect could have been captured by these measurements, our model can accommodate more variables that are not specific to an advertiser's identity. Extensions to allow for an advertiser-dependent effect could easily make the equilibrium intractable.

independently distributed from all of the observed attributes and is also independent from each other. We further assume that the distribution of each stochastic term is symmetric about 0, which results in the mean and median being 0. The assumption on the stochastic components enables us to use the maximum score estimator.

3.3.   Estimation Method

Although we have established the conditions under which the advertiser-publisher stable equilibrium exists, in a general empirical setting with many advertisers and many publishers on both sides of the advertising network, it is very difficult to fully specify the sufficient conditions for equilibrium outcomes. Furthermore, unique stable equilibrium in general does not exist. Therefore the standard maximum likelihood estimation approach cannot be applied without imposing additional restrictive assumptions (e.g., Sørensen 2007). We adopt in the model estimation the maximum score approach (Manski 1975, Fox 2010a), which uses only the necessary conditions derived from equilibrium.

Three sets of inequalities can be derived from the advertiser-publisher stable equilibrium of allocation $\mathcal{M}$ and price $\mathscr{P}$:

• *Across-publisher pairwise stability*: $V_{ik_j} - p_{k_j} \geq V_{ik'_{j'}} - p_{k'_{j'}}$, $\forall \langle i, k_j \rangle \in \mathcal{M}, \langle i, k'_{j'} \rangle \notin \mathcal{M}$. This condition is derived as follows: consider the case in which i purchases the k'th slot from j' but not the kth slot from j. Given the price of each slot as fixed, the individual rationality condition in Definition 1 implies that $V_i(\mathcal{M}_i^- \cup \langle i, k'_{j'} \rangle) - p_{k'_{j'}} - \sum_{\delta: \langle i, \delta \rangle \in \mathcal{M}_i^-} p_\delta < V_i(\mathcal{M}_i^- \cup \langle i, k_j \rangle) - p_{k_j} - \sum_{\delta: \langle i, \delta \rangle \in \mathcal{M}_i^-} p_\delta$, where $\mathcal{M}_i^- = \mathcal{M}_i \setminus \langle i, k_j \rangle$. Putting our assumptions regarding the substitution effects between advertising slots, $V_i(\mathcal{M}_i^- \cup \langle i, k_j \rangle) = V_i(\mathcal{M}_i^-) + V_i(\langle i, k_j \rangle)$, into the condition implied by Definition 1, we get the pairwise stability condition $V_{ik_j} - p_{k_j} \geq V_{ik'_{j'}} - p_{k'_{j'}}$, $\forall \langle i, k_j \rangle \in \mathcal{M}, \langle i, k'_{j'} \rangle \notin \mathcal{M}$. It is valid regardless of whether or not the alternative advertising slot $k'_{j'}$ is currently occupied. If it is not occupied and the value is larger than the current position of k_j, the advertiser can be better off by simply purchasing $k'_{j'}$ instead of k_j; if it is occupied by advertiser i', and advertiser i also prefers the slot to the current one, then the publisher can at least increase the price of p_{k_j} slightly to get better off.

• *Within-publisher pairwise stability*: $V_{ik_j} + V_{i'k'_j} \geq V_{ik'_j} + V_{i'k_j}$, $\forall \langle i, k_j \rangle \in \mathcal{M}, \langle i', k'_j \rangle \in \mathcal{M}$. This condition is a local production maximization condition on publisher j. If $V_{ik_j} + V_{i'k'_j} < V_{ik'_j} + V_{i'k_j}$, then advertisers i and i' would have the incentive to exchange the slots to make both better off under a certain transfer. The transfer can be realized when publisher j sets prices for both slots, with the sum no worse than the current condition.

• *Individual rationality*: $V_{ik_j} - p_{ik_j} \geq V_{i0}$, $\forall \langle i, k_j \rangle \in \mathcal{M}$. This is directly from the individual rationality condition under Definition 1.

We use the semiparametric maximum score estimator (Fox 2010a, Manski 1975) in the estimation. This estimator maximizes a score function over the parameter space. The score value is the total number of equilibrium inequalities that are satisfied under specific parameters. In our application, we use *all* of the inequalities implied by the above three sets of inequality conditions. If we define the deterministic part in the profit function as $\bar{\pi}_{ik_j} = \bar{V}_{ik_j} - p_{k_j} = \exp(Z_j\gamma + Y_k\delta + W_{ij}\alpha + X_i\beta) - p_{k_j}$ and denote θ to be the set of parameters to estimate, the estimator is defined as

$$Q(\theta) = \frac{1}{N}\sum_i\sum_j\sum_{j'}\sum_{k_j}\sum_{k'_{j'}}\{I[\langle i, k_j\rangle \in \mathcal{M}, \langle i, k'_{j'}\rangle \notin \mathcal{M}]$$
$$\cdot I[\bar{\pi}_{ik_j}(\theta) \geq \bar{\pi}_{ik'_{j'}}(\theta)] + I[\langle i, k_j\rangle \in \mathcal{M}, \langle i', k'_{j'}\rangle \in \mathcal{M}]$$
$$\cdot I[\bar{V}_{ik_j} + \bar{V}_{i'k'_{j'}} \geq \bar{V}_{ik'_{j'}} + \bar{V}_{i'k_j}]$$
$$+ I[\langle i, k_j\rangle \in \mathcal{M}]\cdot I[\bar{\pi}_{ik_j} > 0]\}, \tag{4}$$

where $I[\cdot]$ is the indicator function and N is the total number of inequality conditions. Given the independent symmetric about 0 assumptions on the stochastic components, we can derive that $Median(\bar{\pi}_{ik_j} - \bar{\pi}_{ik'_{j'}}) = Median(\pi_{ik_j} - \pi_{ik'_{j'}})$, $Median(\bar{V}_{ik_j} + \bar{V}_{i'k'_{j'}} - \bar{V}_{ik'_{j'}} - \bar{V}_{i'k_j}) = Median(V_{ik_j} + V_{i'k'_{j'}} - V_{ik'_{j'}} - V_{i'k_j})$, and $Median(\bar{\pi}_{ik_j}) = Median(\pi_{ik_j})$. This property is equivalent to the assumption that the median of stochastic components conditional on observed attributes is 0 in Manski (1975), which is required for maximum score estimation.

The identification and consistency properties are discussed in detail in Manski (1975) and Fox (2010b). We use the equilibrium transfer data in the profit function, which helps us to quantify the value function to a real monetary term. Using Monte Carlo experiments, Akkus et al. (2012) and Fox and Bajari (2013) both verified that maximum score estimator with equilibrium price transfers performs extremely well and is robust to different distributional assumptions of the error terms. Because our many-to-many matching context is different from the one-to-one and one-to-many settings in the above two studies, we use Monte Carlo experiments to verify the finite sample performances. Results show that the estimator performs quite well when the standard deviation of the stochastic component is relatively small. Moreover, estimates for matching effects are quite consistent even when the standard deviation is large. We outline details of the simulation exercise in Online Appendix B.

The score function defined above is a step function. Thus, we cannot use the derivative-based optimization routines. Instead, we use the differential evolution (DE) (Storn and Price 1997) method suggested in the literature. The confidence intervals of parameter estimates from this score estimator are difficult to derive analytically, so we rely on subsampling to compute them.

Based on the work of Politis et al. (1999), Fox (2010b) shows that subsampling yields consistent estimates for those standard errors. We randomly sample 200 publishers and those correspondingly transacted advertisers in each subsampling iteration. We subsample 360 iterations to derive the 95% confidence intervals.

4. Estimation Results

Point estimates, along with the 95% confidence intervals, are reported in Table 4. Bold numbers indicate significance at the 5% level. For the advertiser effect, the average price is marginally positively significant and log of monthly sales is positively significant. This implies that popular stores selling high-valued items would in general find a higher value in advertising, probably because they have higher conversion rates for referred visits and higher profit margins. For the publisher effect, the significant variables are the log of unique IP visits and average page views of visitors. This is intuitive, as more visits will generally lead to a larger number of exposures, and more page views presumably lead to a higher chance of advertising exposure for each potential consumer. The effect of PageRank is small and insignificant, indicating that the Google PageRank score does not well reflect the true value of a website from an advertiser's profitability perspective in our context.

Four variables are significant in terms of the position effect. We find a strong positive effect of advertisement size and whether the advertisement would appear on the main page. This is as expected, since large size advertisements on a center page would naturally attract more attention. The same advertisement appearing on the main page would, on average, increase its value by 16% over its appearance on other pages. The position of the advertisement on the Web page is significantly negative, implying that putting an advertisement at the bottom of the page would reduce its value by an average of 22% compared to placing it at the top of the page. Finally, the number of ad slots, which captures competition between advertisements, is significantly negative. For example, the value of an ad slot on a page with two competing slots is, on average, 15% less than the value if it faces no competition. In terms of the value from matching between advertisers and publishers, we find that the contextual correlation variable is positively significant, whereas the geographic distance measurement is insignificant. This suggests that advertisers are more likely to consider content matches than geographical matches when making purchase decisions. For the match of advertiser and publisher categories, several dummies are significant: advertisers selling men's products and women's products find higher value when they match with a website dedicated to fashion topics, whereas stores selling

Table 4   Parameter Estimates for the Matching Model

	Estimates	Conf. Int.
Advertiser effect		
items	−0.08	(−0.18, 0.13)
avg_price	**0.16**	(0.00, 0.33)
month_sales	**0.35**	(0.27, 0.44)
Publisher effect		
pagerank	0.01	(−0.00, 0.05)
unique_ip	**0.18**	(0.11, 0.24)
page_view	**0.11**	(0.04, 0.17)
Position effect		
ad_size	**0.32**	(0.18, 0.42)
ad_mainpage	**0.15**	(0.05, 0.21)
ad_pages	0.08	(-0.02, 0.19)
n_slots	**−0.08**	(−0.13, −0.02)
ad_position	**−0.25**	(−0.29, −0.12)
Matching effect I		
content	**0.30**	(0.17, 0.68)
geography	0.01	(−0.00, 0.06)

Matching effect II	Advertiser category				
	Men's products	Women's products	Digital products	Foods	Household products
Main effect	1.05	0.14	**1.43**	0.21	0.83
	(−0.81, 1.63)	(−0.46, 0.38)	(0.36, 1.83)	(−0.16, 1.09)	(−0.04, 1.48)
Publisher category					
Fashion	**1.28**	**0.70**	**−1.48**	0.04	**−0.98**
	(1.00, 2.06)	(0.51, 1.00)	(−2.13, −1.24)	(−0.11, 0.18)	(−1.89, −0.34)
Life information	−0.02	**0.34**	**−0.76**	−0.01	**0.33**
	(−0.53, 0.67)	(0.23, 0.59)	(−1.31, −0.37)	(−0.17, 0.14)	(0.07, 0.50)
News portal	**0.69**	**0.33**	**−0.45**	**0.19**	**−0.99**
	(0.15, 1.37)	(0.18, 0.56)	(−0.81, −0.06)	(0.04, 0.39)	(−1.51, −0.88)
Online shops/services	**1.11**	**0.42**	−0.33	0.18	−0.34
	(0.15, 1.53)	(0.27, 0.68)	(−0.83, 0.16)	(−0.06, 0.29)	(−1.00, 0.02)
Entertainment/other	—	—	—	—	—
Demographics					
Gender					
Mainly male	0.31	**−0.45**	**0.22**	0.50	−0.08
	(−0.35, 0.71)	(−0.87, −0.29)	(0.27, 1.06)	(−0.14, 0.47)	(−0.23, 0.06)
Mainly female	**−1.19**	**0.13**	0.30	0.14	−0.04
	(−1.57, −0.82)	(0.03, 0.31)	(−0.27, 0.53)	(−0.14, 0.36)	(−0.23, 0.08)
Undifferentiated	—	—	—	—	—
Age					
≤25 years old	0.21	−0.07	0.21	**−0.25**	**−0.37**
	(−0.05, 1.44)	(−0.16, 0.04)	(−0.16, 0.49)	(−0.38, −0.11)	(−0.66, −0.19)
>25 years old	—	—	—	—	—
Income					
≤CNY 3,000	−0.19	−0.05	0.10	0.87	−0.22
	(−0.37, 0.35)	(−0.19, 0.03)	(−0.14, 0.49)	(−0.07, 0.19)	(−0.54, 0.12)
>CNY 3,000	—	—	—	—	—

Note. Numbers in parentheses are 2.5% and 97.5% confidence quantiles from subsampling of 360 times.

household products such as furniture find less value in advertising on fashion websites and news portals. For digital product sellers, the value of advertising is highest on entertainment-related websites and lowest on fashion websites. Our estimates also suggest the importance of demographic matches: there is little value in advertising men's products on websites that have a mostly female audience, and vice versa. For digital products, it is more valuable to advertise on websites with a concentration of male viewers. In terms of age, sellers of food and household products find higher value in an old population. Finally, the income effect across advertiser categories is not significant.

Based on the estimation results, we can calculate the value of every advertising slot for every advertiser. The numbers are reported in Table 5. The value for all

Table 5 **Distribution of Values from Main Estimates**

	Count	Min	Median	Mean	Max
Estimated median value (V_{ijk}) in CNY					
Full sample	479,136	0.07	21.01	44.60	1,506.00
Matched sample	655	0.33	32.24	70.20	984.80
Unmatched sample	478,481	0.07	21.00	50.90	1,506.00
t-statistic for difference of mean is 6.25, *p*-value is 0.00					
Advertiser score	483	0.16	3.91	4.92	25.87
Publisher score	295	0.39	1.40	1.41	2.73
Position score	992	0.76	1.15	1.17	1.69
Matching score	142,485	0.21	3.69	4.79	37.63

potential advertiser-slot level matches ranges from less than 0.07 CNY to 1,506 CNY per day, with the median and mean at about 21 CNY and 45 CNY, respectively (first row of "full sample"). This implies the importance of correct matches. The value for the observed advertiser-publisher matches ("matched sample") in the data is significantly higher, with the median at about 32 CNY and the mean at 70 CNY. This is consistent with the sorting behavior in the economics literature: advertisers and publishers who can create high value self-select to match with each other, while those who are unable to create value will be left out of the market.

The rest of Table 5 reports the scores due to advertiser, publisher, position, and matching effects. The scores are median estimates of each effect in the value function, which are calculated by putting the estimates into each corresponding component of the value function in Equation (3). For our specific value function, each score can be interpreted as a multiplier. For example, if a median advertiser i (advertiser score 3.91) advertises on a median publisher j (publisher score 1.40) in median slot k (position score 1.15), with the median matching value between the advertiser i and publisher j (score 1.61), then the expected median value is $V_{ijk} = 3.91 \times 1.40 \times 1.15 \times 3.69 = 23.23$ CNY. This is less than the median value in the matched sample (32.24 CNY), again suggesting the sorting behavior. The advertiser score ranges from 0.16 to 25.84, showing a tremendous variation among advertisers' base values. The publisher score varies seven times from the minimum to the maximum, and the position score varies 2.2 times. The matching score varies the most—182 times from the minimum to the maximum—suggesting that the effect of matching between advertisers and publishers is most likely the determining factor in driving market allocation.

5. Policy Experiments

Based on the structural model estimates, we now address our second research objective, which involves the mechanism designs of advertising platforms. In our first policy analysis, we examine whether and to

what degree Taobao can improve its profit using a more centralized allocation that matches publishers and advertisers that is similar to Google's role in its display network. We simulate an extreme scenario where Taobao determines the allocation in a completely centralized way. In a second policy experiment, we explore the profit implications of Taobao adopting Google's pricing scheme—GSP auction. We further provide discussions on how the revenue source might influence the platform's choices regarding GSP versus list price. Understanding the connection between platform revenue source and pricing mechanism provides insightful managerial implications for platforms.

One factor that greatly influences the platform design is which party would have better knowledge about the matching values, the advertisers or the platform? Each individual advertiser may have better knowledge of the advertising effect as compared to the platform, or vice versa. Which party possesses better knowledge depends on the design of the network and how experienced the advertisers are. Essentially, an advertiser's value of advertising, as modeled in Equation (2), depends on (a) the advertiser's general valuation for advertising v_i, (b) the publisher's characteristics (EI_j and PE_k), and (c) the value of a specific advertiser-publisher matching (m_{ij}). In general, advertisers would have private information regarding (a) and (c) because not every advertiser attribute is observed or can be easily quantified by the platform—e.g., the advertiser's outside option value, demand seasonality, and specific targeting audiences, to name a few. On the other hand, the platform might be better informed on (b) if not all of the attributes that critically determine the advertising effects are revealed on the platform. In our specific context, the platform provides detailed information about each advertising space from each publisher, including the average number of impressions and clicks, characteristics of the audience, and the transaction history. Thus, it is more likely that the platform would have lesser knowledge than the advertisers do. In the second policy analysis, we assume that the platform may have lesser knowledge. We manipulate the level of knowledge and explore the equilibrium outcomes. In the first policy analysis, we rule out this factor by assuming the platform to have the same knowledge as the advertisers collectively do.

The model estimates give us the median value function of advertiser-publisher matches—i.e., the deterministic part of the model—while the stochastic components and the outside option values remain unspecified. For the purpose of policy experiments, we need to assign them values such that all of the equilibrium conditions we derive are satisfied. We first rely on a quadratic linear programming approach to impute the stochastic unobservables. Specifically, denoting

$e_{ik_j} = \nu_j + \kappa_k + \mu_i + \varepsilon_{ij}$, we minimize the objective function of $Q(e) = \sum_{ik_j}(\exp(e_{ik_j}) - 1)^2$, with the constraint that all three sets of inequality conditions used in estimation are satisfied. Based on the imputed e_{ik_j}s, we further calculate the bounds of the outside option values, V_{i0}s, where the lower bound is the maximum profit from the set of feasible advertising slots that are not purchased, and the upper bound is the minimum profit in the set of slots that are currently purchased. Because the two bounds turn out to be very close to each other, with a mean ratio of lower bound to upper bound being 0.97, we use the average of the two in our policy experiments. Details on the approaches we use to assign the values for the stochastic components and outside option values are described in Online Appendix C. We also discuss the advantages of and caveats to our approach.

5.1. Centralized Allocation of Advertising Slots by Taobao

In the first policy simulation, we consider that Taobao allocates advertising slots to advertisers in a centralized fashion, instead of relying on the market mechanism. We are interested in the profit comparison between these two mechanisms. Given that Taobao's profit is a fraction of advertisers' retail revenue referred from advertisements, its profit-maximizing objective is largely consistent with maximizing the advertiser-publisher joint profit when advertisers' retail profit closely correlates to the sales revenue. This is because $\sum_{ik_j \in \mathcal{M}}(V_{ik_j} - p_{ik_j}) + \sum_{ik_j \in \mathcal{M}} p_{ik_j} = \sum_{ik_j \in \mathcal{M}} V_{ik_j}$. Whether or not the current decentralized allocation yields higher platform profit as compared to a centralized one depends on two factors. The first is the platform's knowledge level—i.e., the extent to which the platform possesses greater or lesser knowledge of advertising effects as compared to each individual advertiser. Intuitively, Taobao would benefit from a decentralized allocation if it has much less knowledge than the advertisers do. The second factor relates to the difference in the objective function that is being maximized across the two mechanisms: in a centralized mechanism, the platform explicitly maximizes the platform profit when determining the allocation and prices; however, in a decentralized market-based mechanism, where individual advertisers and publishers form matches to maximize their own profits, joint advertiser-publisher profit maximizing (and thus, platform profit maximizing) is not guaranteed.[12]

We assume that the platform has perfect knowledge regarding the advertising value, as the individual advertisers collectively do. Under this assumption,

centralized allocation always achieves a higher level of joint advertiser-publisher profit. We formulate the centralized allocation problem as a linear programming problem as follows:

$$\max_{d_{ik_j}} \sum_i \sum_j \sum_k d_{ik_j} V_{ik_j}$$

$$\text{s.t.} \quad d_{ik_j} \in \{0, 1\}; \quad d_{ik_j}(V_{ik_j} - V_{i0}) \geq 0;$$

$$\sum_k d_{ik_j} \leq 1, \quad \forall (i, j) \text{ pair}; \quad \sum_i d_{ik_j} \leq 1, \quad \forall k_j.$$

The first constraint represents the binary decision. The second constraint describes individual rationality. It enters because when a specific allocation violates this rule, the advertiser would have an incentive to exit the market. The remaining two conditions require that a slot could be sold only once and that an advertiser could purchase only one slot from a publisher.

The total advertiser-publisher joint profit achieved under the current market mechanism is 75,062 CNY. The joint profit under a centralized allocation when the platform has perfect knowledge about all of the specific value components is 76,029 CNY. The result shows that the current market-based allocation mechanism has achieved a very high level of advertiser-publisher joint profit—i.e., 99% of the level that could have been achieved under a centralized allocation. For a platform such as Taobao, which profits from advertisers' revenue, its profit objective is consistent with maximizing the advertiser-publisher joint profit. Thus, adopting a centralized allocation mechanism would not further improve its profit as compared to a market-based one, even when we assume that the platform has perfect knowledge about the advertising values of individual advertisers. The finding is also academically quite interesting. The competitive market force would lead to a Pareto optimal allocation in which it is impossible to further improve the joint profit without making anyone worse off. It is quite possible that because the large number of advertisers and publishers are maximizing own profit, the resulted A-P stable equilibrium is much worse than under the centralized allocation. Our result, however, shows that the competition among advertisers and the competition among publishers, for advertising slots, force any stable allocation to achieve a high joint advertiser-publisher profit level.

5.2. Generalized Second Price Scheme

Different from Taobao's cost-per-day list price format, cost-per-click (CPC) based GSP is embraced by other advertising networks such as Google's display network. GSP is an automated bidding system in which the platform plays a role in assigning advertiser-keyword specific quality scores. An advertiser submits a bid for each keyword it targets. Whether or not it can acquire slots associated with the specific keyword depends

---

[12] We prove in the conceptual framework that the A-P stable allocation is Pareto optimal. However, Pareto optimality does not necessarily imply that advertiser-publisher joint profit is maximized.

on the index "AdRank," which is the product of the assigned quality score and its bid. A winning advertiser pays a quality score adjusted amount. For example, if an advertiser i with a quality score of q_i submits the bid b_i (AdRank $= q_i b_i$) and is ranked one position above advertiser i' with quality score $q_{i'}$ and bid $b_{i'}$, then the CPC price advertiser i needs to pay $c_i = (q_{i'} b_{i'})/q_i$. Given the prevalence of GSP, the policy question we ask here is the following: why does Taobao adopt list price instead? Can Taobao improve its profit by switching to the GSP pricing format?

Quality score plays a critical role in determining the equilibrium outcome in the GSP bidding scheme (Varian 2007, Edelman et al. 2007). The rationale for the assignment of quality score is to balance the power of advertisers in getting the slots, such that advertisers who could generate higher values are at an advantage for getting a same position at a lower cost. This essentially helps the platform to achieve a local production maximization allocation under a bidding system. Google assigns quality scores based on the click-through rates in previous periods. In our context, the platform can assign quality scores based on the click-through rate and/or conversion rate when i advertises on publisher j, which is captured by the matching effect m_{ij} in the value function specification. In practice, quality scores are assigned at the advertiser-keyword level, and the platform determines which publishers each keyword is linked to. To simplify analysis, we do not consider how the associations between publishers and keywords are determined. Instead, we focus purely on the effect of the pricing mechanism here, by defining the quality score at the advertiser-publisher level.[13] Previous studies on sponsored search advertising treat the quality score as the true click-through rate. However, given the large number of possible matches between advertisers and publishers, to assign the correct quality score for each advertiser when matching with specific publishers is a challenging task, which crucially depends on the platform's knowledge regarding the value of m_{ij} for all i and j. We use a variable "technology" to represent the ability for the platform to assign the correct quality scores, which is defined as

$$\lambda_{Tech} = \text{cor}(\ln(m_{ij}), \ln(q_{ij})),$$

where m_{ij} is the matching score between advertiser i and publisher j, and q_{ij} is the corresponding assigned quality score. A high level of technology implies that the platform has much knowledge regarding the matching values and, thus, can assign the quality scores optimally.

[13] This practice should not impact the findings in this analysis, because we are essentially creating replicated markets when a keyword is associated with different publishers.

In previous studies (Varian 2007, Edelman et al. 2007), a perfect technology ($\lambda_{Tech} = 1$) is implicitly assumed. Our simulation can help us understand the importance of platform technology in the GSP pricing format.

In a GSP auction, the platform does not need to have full knowledge about the value function of each advertiser V_{ik_j}. The only part that it needs to know to maximize its own profit is the matching part m_{ij}, whereas the other parts are elicited in advertisers' equilibrium bids. The optimal bidding behavior based on full information assignment $q_{ij} = m_{ij}$ is derived in Varian (2007). In Online Appendix D, we derive the extension where quality scores may, in general, be different from matching scores. We also allow advertisers to have outside options. There are multiple equilibria of bids in this GSP bidding game, leading to a range of feasible total payments. We also give the minimum and maximum of total payments under these equilibria.

We vary the quality scores q_{ij} according to different levels of λ_{Tech} and explore the equilibrium allocation and prices from each advertiser's bidding behavior. Given each technology level, we first simulate a number of quality score assignments and then determine which advertisers would bid for which publisher slots under a specific quality score assignment. The equilibrium allocation and range of bids for each publisher market are then calculated according to the conditions in Online Appendix D. Online Appendix E outlines the simulation process step by step. Figure 1 reports the expected total advertiser revenue and Figure 2 reports the total publisher revenue—i.e., $\sum_{jk} p_{k_j}$. To be comparable, our analysis is based on those advertising slots that are sold under the Taobao market. The horizontal dashed lines represent the corresponding levels in the current observed matching achieved under list price format. The two solid lines in Figure 2 are the equilibrium lower

Figure 1   (Color online) Equilibrium Total Advertiser Revenue Under GSP
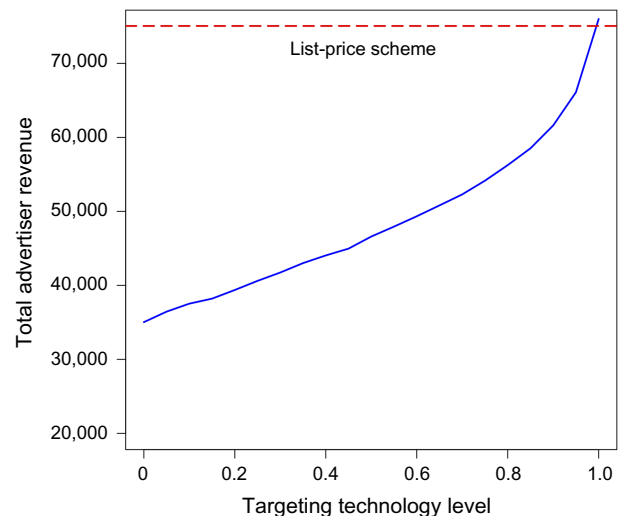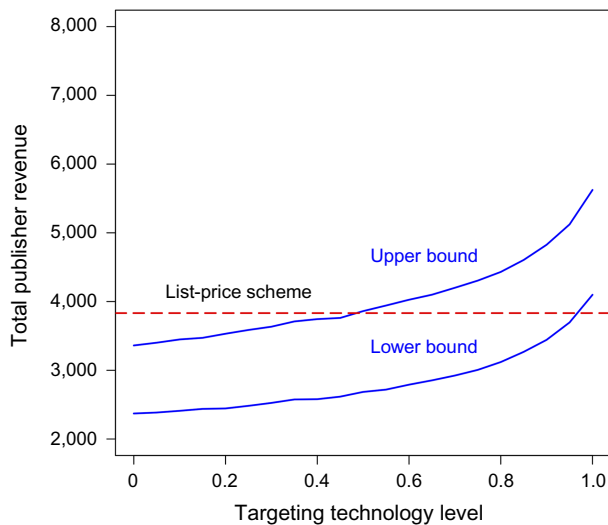
Figure 2 (Color online) Equilibrium Total Publisher Revenue Under GSP

bound and upper bound of total publisher revenue (equilibrium payment) as defined in Equations (D.1) and (D.2).

A few interesting findings emerge from these two graphs. First, there is a significant difference in the sensitivity to targeting technology: the total advertiser revenue is much more sensitive than the total publisher revenue to technology level. The reason for this finding is rooted in the GSP auction with heterogeneous advertisers. In this context, an advertiser's revenue (production function) is multiplicative in its valuation and position. A good quality score assignment essentially induces highly matched advertisers to occupy better positions under equilibrium, resulting in a higher total revenue (local production maximization). Because of the multiplicative revenue form, the total revenue is sensitive to the optimality of quality score assignment. On the other hand, the equilibrium payment is based on the revenue of the bidder ranked a position below, which is less responsive to position changes. Thus, the value created by improvement in quality score assignment is not proportionally captured in the payment, resulting in a less sensitive publisher revenue function. Second, comparing the observed revenue level under the list price scheme to the GSP equilibrium outcomes, we find that the current total advertiser revenue is quite high, at the same level as when Taobao has almost perfect information under GSP—i.e., $\lambda_{tech} = 1$. This is intuitive because at the perfect technology level, the assignment of quality score would be consistent with the advertiser matching value, and, thus, the equilibrium allocation under a GSP auction would be the same as under an efficient market mechanism. However, the current total publisher revenue level may be lower than the equilibrium levels under GSP when the platform has a high level of technology. We now

explain the intuition behind this interesting finding. Shapley and Shubik (1979) show that there exists a range of prices that could support an allocation to be the equilibrium under a market mechanism. The realized prices depend on the market power of each side and determine the surplus each side can get. In our context, the calculated advertiser and publisher surpluses based on the realized prices reveal that the market favors the advertiser side. However, under a GSP auction, when the platform improves quality score assignment, it not only increases the total advertiser revenue of a specific publisher market but also the competition among advertisers—i.e., advertisers with low quality scores need to bid more than before to occupy a good position. This shifts the market power toward the publishers. Thus, when the technology level is high, total publisher revenue is more likely to be higher under the GSP auction than under a market mechanism.

Our findings reveal a potential explanation for platforms' divergent pricing scheme designs. Since Taobao profits from the advertiser side, it does not have an incentive to use the GSP auction: the total advertiser revenue would not increase under GSP even with a perfect technology, but it could potentially be quite low when it has a low technology level. The story for other advertising networks that profit from the publisher side is different: they could potentially gain under a GSP auction when they invest in improving the technology. This could be a potential reason why networks such as Google use a GSP auction scheme in its display network. However, not every platform that profits from total publisher revenue would be better off under GSP. For those with low levels of technology, a list price scheme may still be a better option. For example, AOL uses a list price scheme in its network, advertising.com.

This simulation could also help quantify the value of having quality scores in GSP auctions, which has never been done in the literature. In the search advertising business, Yahoo! initially did not have the quality score component—i.e., the ranking of advertisers on a specific keyword is determined purely by their bids, which corresponds to $\lambda_{tech} = 0$ in our specification. On observing "the secret to Google's success," Yahoo! later adopted Google's innovation in the auction mechanism (Coy 2006). Industry experts estimate that "Google earns about 30% more revenue per ad impression than Yahoo! does" because of the quality score component in the auction mechanism.[14] From our simulation for this specific empirical context, the potential revenue gain from the quality score component could be as high as 70% when the platform has a high technology level.

[14] http://www.bloomberg.com/bw/stories/2006-03-05/the-secret -to-googles-success.

6. Conclusion

Advertising networks represent one of the fastest-growing changes in the online advertising industry. Understanding what drives advertiser-publisher matching behavior in the two-sided markets created by advertising networks is a first step investigation. Comparing the different business models employed by advertising networks is also of great interest to both practitioners and academics. We estimate advertisers' valuation function based on a data set from a leading Chinese advertising network and compare the optimality of the centralized versus market-based advertising slot allocations, as well as the optimality of different pricing schemes. Model estimates reveal that advertiser-publisher matching is the largest determinant in an advertiser's value function. We also show that the market-based allocation mechanism has achieved efficiency similar to that of a centralized allocation. Price setting and auction can be preferred to one another, depending on the platform's revenue source and targeting technology. Our findings are consistent with observations from the industry, where different platforms (Google versus Taobao) make different choices. We provide important guidance for the strategic mechanisms that any advertising network has to design.

To simplify the analysis, this paper abstracts away from many other interesting issues. For example, Taobao's advertising service is built on the success of its retail platform, which helps generate advertising demand from online retailers. How to attract advertisers and publishers represents an important issue to platforms. Similarly, the competition between advertising networks is becoming evident, as more and more players provide similar services and may have a direct impact on the optimality of allocation and pricing strategies. In addition, the evolutionary process of a matching platform is also of great interest.

From a technical perspective, we make assumptions that the experienced advertisers in our sample are perfectly informed about the advertising values, and there is no search cost in the matching process. Our empirical model also assumes away the potential substitution or complementary effects when an advertiser purchases slots from different publishers. These assumptions are likely to hold in our application but may be less likely to in other contexts. Although relaxing these assumptions would further enhance our understanding of the research questions, it is empirically challenging to incorporate either of them into the current model framework in a simple and tractable way, because the model estimation relies centrally on the equilibrium condition that cannot be directly derived under less restrictive assumptions. In addition, our model does not account for unobserved heterogeneity, which could potentially be identified if we had information across markets with very different characteristics (Fox and Yang 2012). These issues represent interesting directions for future research.

Supplemental Material

Acknowledgments

References

Akkus O, Cookson JA, Hortaçsu A (2012) The determinants of bank mergers: A revealed preference analysis. Working paper, University of Chicago, Chicago.

Baccara M, Imrohoroglu A, Wilson A, Yariv L (2012) A field study on matching with network externalities. *Amer. Econom. Rev.* 102(5):1773–1804.

Becker GS (1973) A theory of marriage: Part I. *J. Political Econom.* 81(4):813–846.

Chan TY, Park YH (2015) Consumer search activities and the value of ad positions in sponsored search advertising. *Marketing Sci.* 34(4):606–623.

Chan TY, Wu C, Xie Y (2011) Measuring the lifetime value of customers acquired from Google search advertising. *Marketing Sci.* 30(5):837–850.

Chatterjee P, Hoffman D, Novak T (2003) Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Sci.* 22(4):520–541.

Cho C-H, Cheon HJ (2004) Why do people avoid advertising on the Internet. *J. Advertising* 33(4):89–97.

Choo E, Siow A (2006) Who marries whom and why. *J. Political Econom.* 114(1):175–201.

Coy P (2006) The secret to Google's success. *Businessweek* (March 5). http://www.bloomberg.com/bw/stories/2006-03-05/the-secret-to-googles-success.

Drèze X, Hussherr F-X (2003) Internet advertising: Is anybody watching. *Marketing Sci.* 22(4):520–541.

Edelman B, Ostrovsky M, Schwarz M (2007) Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *Amer. Econom. Rev.* 97(1):242–259.

Fox JT (2010a) Estimating matching games with transfers. Working paper, Rice University, Houston.

Fox JT (2010b) Identification in matching games. *Quant. Econom.* 1(2):203–254.

Fox JT, Bajari P (2013) Measuring the efficiency of an FCC spectrum auction. *Amer. Econom. J.: Microeconomics* 5(1):100–146.

Fox JT, Yang C (2012) Unobserved heterogeneity in matching games. Working paper, Rice University, Houston.

Gale D, Shapley LS (1962) College admissions and the stability of marriage. *Amer. Math. Monthly* 69(1):9–15.

Goldfarb A, Tucker C (2011) Online display advertising: Targeting and obtrusiveness. *Marketing Sci.* 30(3):389–404.

Hitsch G, Hortaçsu A, Ariely D (2010) Matching and sorting in online dating markets. *Amer. Econom. Rev.* 100(1):130–163.

Hof RD (2009) Ad networks are transforming online advertising. *Businessweek* (February 18). http://www.bloomberg.com/bw/stories/2009-02-18/ad-networks-are-transforming-online-advertising.

Katona Z, Sarvary M (2010) The race for sponsored links: Bidding patterns for search advertising. *Marketing Sci.* 29(2):199–215.

Koopmans TC, Beckmann M (1957) Assignment problems and the location of economic activities. *Econometrica* 25(1):53–76.

Landauer TK, Foltz PW, Laham D (1998) Introduction to latent semantic analysis. *Discourse Processes* 25(2–3):259–284.

Manchanda P, Dubé J-P, Goh KY, Chintagunta PK (2006) The effect of banner advertising on Internet purchasing. *J. Marketing Res.* 43(1):98–108.

Manski CF (1975) Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* 3(3):205–228.

Politis DN, Romano JP, Wolf M, eds. (1999) *Subsampling* (Springer, New York).

Rutz OJ, Bucklin RE (2011) From generic to branded: A model of spillover dynamics in search advertising. *J. Marketing Res.* 48(1):87–102.

Shapley LS, Shubik M (1979) The assignment game I: The core. *Internat. J. Game Theory* 1(1):111–130.

Sørensen M (2007) How smart is smart money: A two-sided matching model of venture capital. *J. Finance* 62(6):2725–2762.

Storn R, Price K (1997) Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.* 11(4):341–359.

Varian HR (2007) Position auctions. *Internat. J. Indust. Organ.* 25(4): 1163–1178.

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

Yang Y, Shi M, Goldfarb A (2009) Estimating the value of brand alliances in professional team sports. *Marketing Sci.* 28(6): 1095–1111.

Yao S, Mela C (2011) A dynamic model of sponsored search advertising. *Marketing Sci.* 30(3):447–468.

Zhang K, Katona Z (2012) Contextual advertising. *Marketing Sci.* 31(6):980–994.