Accelerated Quality Discovery through Sponsored Search Advertising in Online Marketplaces

Hongkai Zhang*
MIT
Job Market Paper [latest version]

November 15, 2016

Abstract

This paper analyzes the dynamic effect of sponsored search advertising on quality discovery on the Taobao.com retail platform. A stylized model spells out the dynamic from a theoretical perspective: A new product’s boosted exposure from sponsored search ads could help the platform to infer how well the product converts exposure to sales (the quality measure). Because the platform awards top organic search ranks to products inferred of high quality, sellers with higher private quality signals would bid more aggressively for sponsored ads, which accelerates the platform’s discovery of these high-quality products. An empirical specification, connecting search ranks to sales through exposure, is estimated using scraped panel data on sales of hundreds of products and their sponsored and organic search ranks under thousands of keywords on both PC and mobile interfaces. A platform-specific feature with individual organic ranks affected by asynchronous weekly cycles provides exogenous variation for identification. The estimation results echo the stylized model’s characterization of the informational landscape and reveal a synergy between the platform’s PC and mobile interfaces: PC consumers browse extensively and generate significant exposure for sponsored products. Observations of sponsored sales to PC consumers could improve the organic rankings in the mobile app, where consumers highly concentrate on top products.

*zhanghk@mit.edu. I would like to thank Glenn Ellison, Muhamet Yildiz and Sara Ellison for invaluable advice and support. This paper benefited from discussions with Nikhil Agarwal, Jie Bai, Andrey Fradkin, Yan Ji, Chris Nosko, Michael Whinston, Kevin Williams and Giorgios Zervas. I also thank participants in MIT IO lunch and IO seminar for their very helpful comments and suggestions. The data collection work is supported by the George and Obie Shultz Fund. All errors are mine.
1 Introduction

E-commerce platforms facilitate transactions between consumers and independent suppliers. Given the drastically reduced cost of listing and transacting, they often carry an enormous number of products. Consumers, typically limited by cognitive capacities or the form factor of the interface, can only investigate a few products carefully. Platforms could help the consumers to narrow their searches to high-quality products that they are likely to purchase, possibly through an informative search function that returns a ranked list of products. Both the platform and consumers would appreciate and benefit from information channels that improve the platform’s knowledge about product quality.

This paper examines a mechanism of quality discovery on an e-commerce platform, Taobao.com. To fulfill its informational role, when a consumer expresses her interest by searching for a keyword, the platform ranks the relevant products on the search result page. A prominent feature of the search result page on Taobao is that a list of sponsored products is displayed alongside the main, or organic, results and sellers can bid on per-click fees to be ranked in the sponsored list. The goal of this paper is to examine the effect of this feature on the platform’s effectiveness as an information intermediary.

I look at the women’s apparel categories where scores of small sellers retail numerous varieties of horizontally differentiated, nonbranded, seasonal goods, and consumers rely heavily on the keyword search. Although consumers can learn some basic information (product title, picture, price and monthly sales) with an effortless glance at the search result page, they incur search costs to investigate product detail pages. The information on those pages determines how likely a consumer viewing the product page will be convinced to make a purchase. This probability, called the conversion rate, is the primary quality measure of interest in this paper, so the word "quality" will refer to the conversion rate most of the time. The platform can learn very limited quality-relevant information from the picture-rich product pages, so it primarily relies on its data on observed consumer browsing and purchasing behaviors: the conversion rate of a product can be recognized by the platform only after sufficient consumer exposure. However, given the large number of new products, consumer exposure is scarce.

I set up a stylized model in Section 3 to illustrate that the sponsored ads system can help the platform allocate consumer exposure and discover high-quality products more efficiently by incorporating the sellers’ private information in an incentive-compatible way. I consider a market with \( N \geq 2 \) new products each period with conversion rates drawn from a known distribution, and assume each product lasts for two periods. The platform has two slots to display to consumers. In any period, the platform sells the sponsored slot on the right to an entrant through a second price auction with simultaneous bids. For the organic slot on the
left, the platform infers from last period data to identify and display the incumbent with the highest expected conversion rate.

A key element of my model is that the sellers have private signals about the quality of their products. An entrant’s bid for the sponsored slot depends on how many extra sales the sponsored slot can bring, conditional on the entrant’s private signal. In the current period, the entrant could receive additional sales from the exposure of the sponsored slot, and it expects this benefit to increase with the private signal. Furthermore, the entrant has a dynamic incentive, as its sales from the sponsored exposure are observed by the platform to infer its expected conversion rate: an entrant with a higher private signal would be more optimistic about the sponsored exposure this period leading to the award of the organic slot in the next period. Section 3.2 shows entrants use bidding functions increasing in their private quality signals, and an entrant with a high private quality signal would bid much higher than the current-period benefit of the sponsored slot.

This monotonic bidding equilibrium for the sponsored slot has positive effects on the average qualities of both the sponsored slot and the organic slot. For the sponsored slot, because the entrant with the highest private signal wins the auction, the average quality of the sponsored slot increases in the accuracy of the private signals. Furthermore, because the sponsored exposure is directed to a new product with a high private signal, the platform’s experiment starts with a very promising candidate and is more likely to confirm a high-quality product for the organic slot in the next period.

To demonstrate these effects, Section 3.3 considers a benchmark environment, in which the platform randomly chooses a fixed number of entrants to display in the right-side slot without seller input, and use their observed sales to determine the product for the left-side slot in the next period. Section 3.4 compares the outcomes of the two environments and varies the private signal accuracy, the number of entrants and the number of consumers. The sponsored ads environment performs better when private signals are accurate and the number of entrants is high, because the total information of the entrants depends on these two factors. The performance of the benchmark environment depends heavily on the number of consumers, which caps the platform’s information power, and would not benefit from an increasing number of entrants as much as the sponsored ads environment.

To understand this quality discovery story from an empirical perspective in the setting of Taobao.com, I collect daily panel data on sponsored ads display, organic search ranks and 30-day sales for hundreds of products and thousands of keywords on Taobao.com by scraping the relevant data from Zhibi365.com\(^1\), an independent data service for sellers. The

\(^{1}\) Many existing papers on Taobao.com (for example, Chu and Manchanda (2016), Ju et al. (2013) and Fan et al. (2016)) use proprietary data obtained directly from the Taobao/Alibaba company to study suitable
platform provides different organic and sponsored ranks on its PC website and mobile app, so I have recorded all four types of ranks. In Section 4, I develop an empirical specification that relates a product’s all types of search ranks under various keywords to consumer exposure and measures its relative conversion rate by matching exposure to sales.

A unique feature of Taobao.com perturbs the organic rankings with asynchronous weekly cycles and helps to identify the empirical model with exogenous variation. For historical reason, the platform nominally relists each product weekly at a time of week chosen by its seller. As the relisting time approaches, the product is considered to expire soon and given some priority in all relevant organic rankings. The priority disappears when the product is automatically relisted, and a new seven-day cycle starts. This feature, retained by Taobao.com as a way to disperse exposure across products, creates exogenous and frequent shuffling in organic ranks, and has a more significant role on the website than in the mobile app.

The estimation reveals some interesting contrasts of consumer behaviors on the mobile app and the desktop website. While consumer exposure is quite dispersed across organic ranks on the website, in the mobile app, consumer concentrates at the top ranks and diminishes very fast for lower ranks. This is consistent with the two interfaces’ difference in the physical factors: mobile users are limited in their abilities to scroll over the list of products and to jump back and forth between the list and individual products, so they rely on the organic ranking much more heavily.

A more striking comparison is that the organic rankings on the website have contributed only a quarter as many sales as their counterparts on the mobile app. This is partly due to the popularity of the mobile app exceeding the desktop website recently. However, this contrast is also consistent with a quality story: the organic rankings on the website shuffle significantly according to each product’s weekly cycle, contrasting to the much more stable organic rankings on the mobile app. Therefore, the mobile app offers more informative and reliable organic rankings so that the consumers expect the top organic products on the mobile app to have high quality, and are more willing to investigate those products, which leads to more sales.

In contrast to the case of organic rankings, the sponsored rankings on the website are much more important than those on the mobile app, most likely due to the ample space and easier browsing for the consumers. In fact, the sponsored rankings on the website
contribute more sales than the organic rankings next to them: in this extreme case, because the sponsored rankings are determined through an informative auction and are much more stable, consumers are more willing to click the sponsored products than the organic results.

These pieces of evidence in Section 5 suggest the reality of the women’s apparel category on Taobao.com is quite close to the stylized model in several aspects. First, since the dominant organic rankings are ones on the mobile app with consumer exposure concentrated at top ranks, they resemble the single top organic slot in the stylized model, which implies that new products have little organic exposure and the dynamic incentive to purchase sponsored ads is strong. Second, the high exposure received by the sponsored products on the website suggests the sponsored list can indeed of high quality and the consumers would respond to that. In contrast, the significant shuffling of desktop rankings resemble the random experiments in the benchmark environment, which degrades the quality of the ranking and suppresses consumer clicks.

My empirical analysis suggests a strong complementarity between the desktop interface and the mobile interface, which is of interest for platform strategies. The deep browsing pattern and the ample advertising capacity on the desktop interface enable quality-relevant information to emerge quickly. The mobile interface, given its strong influence on consumer behaviors, has made great use of this information from the desktop interface to improve efficiency. Essentially, this represents a positive externality generated by the consumers with lower search costs on the consumers with higher search costs. As most platforms today have multiple interfaces that generate heterogeneity in search costs, this complementarity is worth exploring so that platforms can improve the integration and differentiation of consumer experiences on multiple interfaces.

The estimation also yields a relative measure of the products’ qualities and allows testing of the dynamic patterns postulated by the stylized model in Section 5.3. Products in the data set demonstrate aggressive initial advertising and enjoy high exposure from organic search rankings later.

This paper adds to the growing literature on sponsored search. Edelman and Ostrovsky (2007), Edelman et al. (2007) and Varian (2007) are the pioneering papers that examine the auction mechanisms and the bidding strategies on leading search engines. Athey and Nekipelov (2012) analyzes the equilibrium of the auction game with uncertain click-through rates. Chen and He (2011) and Athey and Ellison (2011) relate the advertisers’ heterogeneous valuation of consumer clicks to their probability of converting clicks to transactions, and both show that sellers with the higher quality would bid more because they can achieve more transactions from sponsored exposure. Furthermore, Athey and Ellison (2011) endogenizes the number of consumer clicks received by each slot in the sponsored list, and captures that a
sponsored ranking informative of quality would encourage more consumer clicks. This paper builds on these elements, and augments the sellers’ bidding incentive by the intertemporal link between sponsored lists and organic lists, so that the dynamic patterns of sponsored ad advertising can be understood.

While the aforementioned papers treat the sponsored list in separation, a few papers explore the interaction of the organic list and the sponsored list. White (2013) considers both complementarity and substitution between the goals of the two lists: on the one hand, the higher quality of the organic list can attract more users and therefore allow higher revenue from the sponsored list; on the other hand, the high-quality merchants in the organic lists may compete with merchants interested in the sponsored slots and reduce their willingness-to-pay for the sponsored ads. Katona and Sarvary (2010) takes on a different substitution effect: free exposure in the organic list undermines the marginal benefit of sponsored exposure for the same advertiser.

This paper adds to above strand of research with two dynamic complementarities: First, the sellers’ private information stimulates the competition for the sponsored slots, and the outcome of the sponsored slots helps to improve the quality of the organic ranking. Second, the higher quality of the organic list would attract more consumer exposure for top organic lots, so that the high-quality entrants have stronger incentives to bid for the sponsored slot in order to earn the top organic lots later.

This work also belongs to the literature on platforms and two-sided markets, as it considers a model of facilitated interactions between the two sides that enriches the generic match functions modeled in the pioneering work (for example, Rochet and Tirole (2003) and Armstrong (2006)). The platform’s ability to extract information from early interactions and use it to improve future matchings is a new type of network effect, and its implications for welfare and competition should be explored in future work.

The empirical exercise in this paper connects to the literature on consumer search behaviors and the effectiveness of sponsored ads. Ellison and Ellison (2009) studies the effect of price rankings on consumer search and the sellers’ obfuscation strategies in response. Blake, Nosko and Tadelis (2016) analyzes consumer search patterns using a dataset from eBay and find that consumers on average conduct 36 searches over three days before making a purchase, which echoes the low conversion rates in my data and the delayed effects of exposure in my estimates. Through a large-scale field experiment on major search engines, Blake, Nosko and Tadelis (2015) shows the sponsored ads are informational instead of persuasive, which I assume in my model. Their work also highlights the endogeneity that experienced consumers searching for brand keywords would have purchased the relevant products regardless of their advertising. This endogeneity issue is also studied by Edelman (2013) and Lewis
and Reiley (2014). By focusing on categories with non-branded, horizontally differentiated products and small sellers, I minimize this endogeneity problem.

The paper is organized as follows: Section 2 describes the empirical setting. Section 3 proposes a stylized model. Section 4 describes the data. Section 5 specifies and estimates an empirical model that relates ranks to sales. Section 6 concludes.

2 Empirical Setting

In this section, I first describe the search function of Taobao.com on both desktop and mobile interfaces to give the context of analysis. Then I discuss the factors affecting organic and sponsored rankings. A special feature of the platform with individual organic ranks affected by asynchronous weekly cycles is covered here and will then have an important role in the empirical section.

2.1 Search on desktop and mobile interfaces

Taobao.com contains numerous products in almost every possible category listed by independent sellers. Given the vastness of the selection, consumers most often try to locate the products of interest through keyword search. After entering a keyword (or a combination of keywords) and hitting the search button, a search result page appears. An example, for the keyword "sweater", is given below:

![Desktop search result screenshot](image)

---

2 Consumers can also reach some products by clicking through the curated promotions on the homepage, but this channel covers very few products. Consumers can also directly go to sellers’ virtual storefronts, but most small stores do not have many such consumers who are willing to start their shopping from the store.
On the left is the organic search results, where the relevant products are ranked by Taobao.com’s proprietary algorithm. On the right is the sponsored search results, where a seller has to pay a per-click fee whenever a consumer clicks on its product. Each product in the lists has its picture and title shown, as well as its price, location, the reputation score of the seller and 30-day sales of the product. There are about 40-44 products in the organic list, occupying 4 columns and 10-11 rows. There are 17 products in the sponsored list. Twelve of them occupy the rightmost column, while 5 of them are located at the last row of the page. On a normal computer screen, about the first two rows of the organic results and the top 4 slots in the sponsored column are shown before the user scrolls down. One additional complication is that the first 3 slots in the organic results area are sometimes allocated to sponsored products. When this is the case, the sponsored products will be designated with a small label. Both the small label and the header of the sponsored list says "sellers’ recommendation," and most consumers understand they are influenced by the sellers’ payment to the platform.

Parallel to search results on the website, consumers can also search within the mobile app:

Figure 2. A mobile search result screenshot

The mobile result is a single list of products presented in two columns that the consumer can scroll down to load more and more products. For every ten slots in the stream, there is one slot used to display a sponsored product. The mobile organic rankings are correlated with, but different from the organic rankings on the desktop, as the platform adopted different algorithms to adjust for the different consumer behavior patterns. Also, most sellers are allowed to submit different bids for the mobile sponsored slots.

From any search result page, consumers can click the product picture or title to reach a product page with more details about the product. Given the product’s subcategory, the
platform mandates a set of possible attributes which are presented in a plain and modest way. Sellers use most space of the product page to feature even finer information, including product-specific attributes, additional pictures of the product and clarification of service terms. To ensure uniform rendering on all computers and mobile phones, most sellers choose to organize these additional details into a series of pictures, with all texts embedded using picture editing software. On a desktop, the product detail pictures usually span several screens. The picture-rich format of the product page makes it hard for the platform to parse the information automatically.

Figure 3. A product Page

2.2 Organic rankings

The ranking of the organic results aims to bring the most relevant, reliable and popular products to the front places. For relevance, the platform would check the product title, attributes allowed for the subcategory, and the seller’s primary category. For reliability, the platform takes into account of seller reputation/service scores and product review scores.
For popularity, the platform uses the product’s sales in the past 30 days and the observed conversion rate, which is the ratio of the number of sales to the number of product page views.

Generating up-to-date and high-quality organic rankings is a challenging task. For a popular keyword, there could be hundreds of thousands of listings that qualify for the organic ranking. In categories where products are mostly horizontally differentiated and non-branded, a significant proportion of products would appear very close in relevance and reliability measures. Hence, the popularity measure is the most differentiating element in the determination of the top ranks.

Among the factors in the popularity measure, the conversion rate is weighed heavily by the platform in its algorithm for the organic rankings for several reasons. First, the conversion rate is a key measurement that directly affects the volume of transactions. Second, it has a strong role in consumers’ shopping experience. Consumers are more willing to click products for details when they often find the products investigated are worth purchasing, and would feel more satisfied overall. Finally, the conversion rate is more costly for sellers to manipulate in unfair ways. The challenge for using conversion rates is time pressure: the platform can only have a reliable observation of a product’s conversion rate after the product receives enough clicks, and many products are only relevant in the medium term and become obsolete in the long term. Notably, this challenge is more or less also faced by other platform with horizontally differentiated time sensitive products, such as Youtube for videos.

**Asynchronous weekly cycle.** A unique and very useful feature of the platform is a shuffling of organic ranking according to the listing time. While a product can exist on Taobao.com for as long as the seller wishes, the listing nominally expires every seven days and is automatically reinstated immediately afterwards. The product’s sale records and review records are not interrupted by this nominal process. The only effect of this pseudo expiration is that products are given a greater advantage in the organic rankings as they are close to the expiration time, and this advantage disappears entirely at the beginning of the next seven-day cycle. On the desktop website, this force can boost a product’s organic ranks drastically on the last day or the last hour of the cycle. On the mobile app, this force is weaker and the organic rankings do not shuffle as much. Below is an example of a product’s desktop and mobile ranks under a given keyword. The desktop rank often improves significantly on the dates noted by horizontal ticks, while the mobile rank shows some mild
shocks on those dates.

The sellers can choose the expiration time by changing the starting time of their listings, and it is popular among sellers to choose an expiration time during evenings and weekends. The pseudo expiration time is not shown to consumers.

2.3 Sponsored rankings

The ranking of the sponsored results is determined through a generalized second price auction. The sellers submit bids for the pay-per-click charge, and the platform ranks the products by their bids multiplied by quality scores in the sponsored list. A product’s quality score (QS) depends on its match to the keyword, the seller’s historical effectiveness in sponsored ads promotion, the seller’s service scores and the product’s click-through-rate (CTR) under the keyword, which is calculated from its recent displays in the sponsored list. The product at the \( n \)th slot is actually charged a pay-per-click (PPC) equals to the bid of the product at the \((n + 1)\)th slot adjusted by the ratio of their click-through rates plus 1 cent.

\[
PPC^{(n)} = b^{(n+1)} \times QS^{(n+1)}/QS^{(n)} + 0.01
\]

Figure 4. An example of a product’s organic ranks under a given keyword
The sellers can see a normalized version of their QSs, on a scale of 1-10, where 10 means a product is among those with the highest QSs under the keyword. For experienced sellers, the first three components of the quality score are not very differentiating, and it is mainly the CTR that affects their bids and auction outcome. Even for the same seller, the CTRs of its products could vary and the seller could experiment with different products to find the one gives the highest QS. Sellers can also experiment with different titles and pictures for same products to improve its QS.

The evolution of the sponsored ads service, which is one of the platform’s main revenue channels, has been an iterative process with revisions aiming at improving the consumer experience. Besides the click-through-rate (CTR), the platform has also included in the quality score (QS) other factors that do not affect the advertising revenue but reflect the relevance of the product to the keyword searched, in order to deter sellers from hurting the consumer experience with irrelevant sponsored ads. Sellers are also encouraged to pay attention to the conversion rates of their sponsored products, and they are educated that the conversion rates they demonstrate following clicks on sponsored ads are recorded and can affect their organic ranks.

Taobao.com started its business without charging proportional fees from sales, and it has remained so. The platform today generates revenue mainly through promotional services including sponsored advertising. However, like many other platforms, Taobao.com’s business strategies have been prioritizing the total volume of transactions on the platform and the satisfactory of the consumers and the sellers in general over the platform’s own current period revenues. A primary concern of the platform today is to compete with other online retail websites for consumers and sellers so that it can maintain its dominant place in the Chinese online retail market. Taobao.com has also been creating externalities that are captured by other branches of its parent company, Alibaba. During its early days, Taobao.com was mainly for small sellers to market products that they purchased from wholesalers on Alibaba.com, and Alibaba.com can make money from those wholesale transactions. This is still the practice for many sellers today. Also, Taobao.com is contributing consumer flow to Tmall.com, a sibling platform on which large sellers including manufacturers sell branded products and pay the platform 1%-3% transaction fee. Both externalities depend on the popularity of Taobao.com.

3 A stylized model

In this section, I set up a stylized market in which the platform has two slots to show. In the main analysis, one slot is for sponsored display. I argue that sponsored advertising
in equilibrium encourages high-quality entrants to bid aggressively for the sponsored slot, which allows the platform to leverage on the seller’s private information to accelerate quality discovery. I then compare this outcome to a benchmark case where both slots are organic so that the platform determines the products displayed without seller input, and discuss what features in the setup would make a strong case for using one slot for sponsored advertising.

3.1 Setup

I consider a discrete time model with an infinite number of periods. Consider a simplified market with a continuum mass of $M$ consumers searching for a specific keyword. In this market, at the beginning of each period, $N$ new products enter and each product lasts 2 periods, so there are always $2N$ available products in total. Each product is solely characterized by its conversion rate, $\gamma_j$, which is the probability for a consumer having landed on the product page to purchase the product. On the search result page, the platform has two slots to display two products at any time.

![Figure 5. Diagram of the search result page](image)

Upon landing on the search result page, each consumer $i$ has to decide whether to click the products’ links to read the product page for detailed information. Doing so incurs a search cost $c_i$, but if the consumer later decides to purchase the product, she will get a benefit $B$ and the seller will get a profit, normalized as 1. The consumer’s clicking and purchasing decisions are independent across products.

The consumers have heterogenous search costs uniformly distributed on $[0, 1]$. In equilibrium, the consumers have rational expectations about the conversion rates of the two products on the result page. Hence, a consumer with search cost $c_i$ would only click a product’s link on the left/right of the search result page if she believes the product on the left/right has an expected conversion rate of at least $c_i/B$. The two slots on the result page may receive different numbers of clicks from consumers if the products at those slots are chosen by different rules. I assume the consumers’ search costs are independent from their purchase behaviors conditional on clicking the product page.
The platform, sellers and consumers share a common prior that for each product, $\gamma_j$ is drawn from a gamma distribution with shape $\alpha_0$ and mean $\bar{\gamma}_0 = \alpha_0/\beta_0$. Before entry, each seller receives a private imperfect signal $q_j$ about its own conversion rate. The signal $q_j$ follows a Poisson distribution with mean $\gamma_j\pi_0$, so the seller has a gamma posterior for its conversion rate, with shape $\alpha_0 + q_j$ and mean $\bar{\gamma}_j^E = (\alpha_0 + q_j) / (\beta_0 + \pi_0)$, according to Bayes rule.

### 3.2 Platform with sponsored slot

In this subsection, I consider an environment with one sponsored slot on the result page. In particular, the platform assigns the right slot on the result page to an entrant through a second price auction with simultaneous bids, and allocates the left slot to the incumbent whom the platform believes to have the highest expected conversion rate among incumbents. The platform uses observed sales from incumbents’ last period to compute their expected conversion rates. Under this set of rules (denoted as $O\mid S$ as the platform has one organic slot and one sponsored slot), I solve for the equilibrium strategies of consumers and sellers.

![Diagram of a platform with one organic slot and one sponsored slot](image)

In this environment, an entrant would bid for the sponsored slot with his posterior after observing his private signal. If the entrant loses the auction, it gets zero sales during the first period and the platform does not update belief about its conversion rate. If the entrant wins the auction, it pays the second highest bid, and receives consumer clicks $\Pi^s$ from the sponsored slot during its first period. Its sales $Q_j$ during this period, which follow a Poisson distribution with mean $\gamma_j\Pi^s$, are observed by the platform. The platform uses this observation to update its belief according to the Bayes rule, so its posterior about the incumbent’s conversion rate $\gamma_j$ at the beginning of the next period is a gamma distribution with shape $\alpha_0 + q_j$ and mean $\bar{\gamma}_j^P = (\alpha_0 + Q_j) / (\beta_0 + \Pi^s)$.

The platform places the incumbent of the highest expected conversion rate at the top organic place with random tie breaking. Hence, if the platform finds the incumbent who
had the sponsored slot last period has an updated expected conversion rate greater than 3 than \( \bar{\gamma}_0 \); the expectation of its prior, this incumbent would be placed at organic slot; otherwise, a randomly chosen other incumbent gets the top organic slot. The single sponsored slot is allocated through a second price auction with simultaneous sealed bids.\(^4\)

An equilibrium of this market consists of three components: the consumer expectation of the conversion rate of the top organic slot \( \hat{\gamma}^o \), the consumer expectation of the conversion rate of the sponsored slot \( \hat{\gamma}^s \), the bidding strategy of entrants. I consider an equilibrium where all entrants use the same bidding strategy, which can be denoted as \( b^E(\hat{\gamma}^E_j) \). In an equilibrium, the consumers’ expectations are consistent with the long run average of the products at the respective slots, therefore, the numbers of clicks for the organic slot and the sponsored slot are \( \Pi^o(\hat{\gamma}^o) = \hat{\gamma}^o BM \) and \( \Pi^s(\hat{\gamma}^s) = \hat{\gamma}^s BM \), respectively.

The bidding strategy is an equilibrium strategy of the auction game, so it must satisfy the incentive compatibility constraint. In a second price auction, this generally means an entrant’s bid equals to its incremental benefit from the sponsored slot. The intuition discussed in the introduction section postulates that entrants with higher quality products would benefit more from the sponsored slot, because they not only can convert more exposures to sales in the current period, but also bear higher probability to earn the organic slot in the next period. This intuition is to be formalized into a proposition.

I define \( W^s(\bar{\gamma}^{e_1}, \hat{\gamma}^s) \) as the probability perceived by entrant \( e_1 \) for him to be allocated the organic slot in the next period given: (1) entrant \( e_1 \) has posterior expected conversion rates \( \bar{\gamma}^{e_1} \) after observing own private signals; (2) in the first period, entrant \( e_1 \) holds the sponsored slot (3) the number of clicks of the top sponsored slot is \( \hat{\gamma}^s(\hat{\gamma}^s) \).

The platform’s rules for slot allocation leads to the following lemma:

**Lemma 1** \( W^s(\bar{\gamma}^{e_1}, \hat{\gamma}^s) \) increases in \( \bar{\gamma}^{e_1} \). \( W^s(\bar{\gamma}_0 + \frac{1}{\beta_0 + \pi_0}, \hat{\gamma}^s) > 0.5 \).

**Sketch of Proof.** See Appendix 1 for the full proof and the exact formula for \( W^s(\bar{\gamma}^{e_1}, \hat{\gamma}^s) \).

The condition for entrant \( e_1 \) to earn the organic slot in the next period is that its sales from the sponsored exposure in this period, \( Q^{e_1} \), is greater than than \( \Pi^s(\hat{\gamma}^s) \bar{\gamma}_0 \), the expected sales according to the platform’s prior. Conditional on \( \bar{\gamma}^{e_1} \), \( Q^{e_1} \) has a negative binomial distribution that increases in \( \bar{\gamma}^{e_1} \) in the sense of first-order stochastic dominance, hence the probability for \( Q^{e_1} \) to exceed a given threshold increases in \( \bar{\gamma}^{e_1} \). When \( \bar{\gamma}^{e_1} = \bar{\gamma}_0 + \frac{1}{\beta_0 + \pi_0} \),

\(^3\)I only consider for generic parameters such that equality is impossible.

\(^4\)It is assumed that the platform does not infer any information from the seller’s bid per se, which is a realistic characterization of Taobao.com. The platform has not stated that the bid itself has any effects on the organic rankings, and the seller communities have been always working on tricks to pay as less as possible for a given number of clicks. Because of the significant heterogeneity among the sellers and their products, if the platform did relate the sellers’ bids to their organic rankings in the future, it would be too much a burden for sellers to figure out their bidding strategies.
according to Van de Ven and Weber (1993) (see the full proof for details), the median of the distribution of $Q^{e_1}$ is bounded from below by $\Pi^s (\hat{\gamma}^s) \bar{\gamma}_0$, so $Q^{e_1}$ is more likely to exceed $\Pi^s (\hat{\gamma}^s) \bar{\gamma}_0$ than not. The term $\frac{1}{\beta_0 + \pi_0}$ is needed because the negative binomial distribution is right-skewed.

Now I can show the following proposition:

**Proposition 1** the entrant’s equilibrium bidding function $b^E (\cdot)$ is increasing.

**Sketch of Proof.** See Appendix 1 for the full proof. I consider an entrant $e_1$ with posterior expectation $\bar{\gamma}^{e_1}$. Denote his equilibrium bid as $b^{e_1}$. Take any alternative bid $b' < b^{e_1}$, the entrant’s expected payoff following bids $b^{e_1}$ would be higher than that with $b'$, given other players sticking to their equilibrium strategy. Holding the value of $b^{e_1}$ and $b'$ constant, I show this difference in payoffs would be higher for an entrant with higher posterior expectation $\bar{\gamma}^{e_1}$, which implies that for this entrant, $b'$ cannot be its equilibrium bid, so its equilibrium bid must be at least $b^{e_1}$.

Because $b^{e_1} > b'$, there is a positive probability that bid $b^{e_1}$ wins the auction while bid $b'$ cannot. The difference in payoffs is caused by this situation and has three parts. The first part, winning instead of losing the sponsored slot makes current period profits, which increases in the entrant’s posterior expected conversion rate proportionally. The second part, winning the sponsored slot in this period leads to the organic slot in the next period with probability $W^s (\bar{\gamma}^{e_1}, \hat{\gamma}^s)$, which by Lemma 1 increases in $\bar{\gamma}^{e_1}$. Finally, the entrant also needs to pay for the sponsored slot, and the payment does not depend on $\bar{\gamma}^{e_1}$. Hence, the difference in payoffs following bids $b^{e_1}$ and $b'$ has all three components increase or hold constant in $\bar{\gamma}^{e_1}$, which completes the proof.

Given the increasing bidding function, the sponsored slot always goes to the entrant with the highest private signal. Given the common prior, the entrants’ private signal follows a negative binomial distribution $NB \left( \alpha_0, \frac{\pi_0}{\pi_0 + \beta_0} \right)$. The $N$ private signals of all entrant form a sample of size $N$ of this distribution. Denote the maximum value in this sample as $\bar{q}_0$. Then the expected quality of the sponsored slot in equilibrium is

$$\hat{\gamma}^s = E \left[ \frac{\bar{q}_0 + \alpha_0}{\pi_0 + \beta_0} \right] = E \left[ \bar{q}_0 \right] + \alpha_0 \frac{\pi_0}{\pi_0 + \beta_0} \tag{1}$$

Given $\hat{\gamma}^s$ and $\Pi^s (\hat{\gamma}^s)$, the stochastic procedure that generates the incumbent at the organic slot in equilibrium is fully specified, hence its expected conversion rate, $\hat{\gamma}^o$, is well defined in terms of the model parameters, although the exact formula is too complex to
present here. For the purpose of presentation, I denote \( \hat{\gamma}^o \) as a function:

\[
\hat{\gamma}^o = \Gamma^o (M, B, N, \alpha_0, \beta_0, \pi_0)
\]  

(2)

In Section 3.4, both \( \hat{\gamma}^o \) and \( \hat{\gamma}^s \) will be numerically computed to compare against counterparts from a different environment.

I now use the local incentive compatibility (LIC) constraint to derive the last component of the equilibrium \( b^E (\cdot) \). The entrant considers how marginally change his bid would affect his payoff: in equilibrium, an entrant should be indifferent about changing its bid marginally conditional on event \( \Lambda(b) \). Because it is a second price auction, a small change in bid has no effect on any outcome unless the highest bid among the other entrants’ bids, denoted as \( b^E (\tilde{\gamma}^e_h) \), is the same as the entrant’s bid \( b \). Denote this tipping event as \( \Lambda(b) \).

Denote entrant \( e_1 \)’s first period profit as \( V_1^{e_1} \) and second period profit as \( V_2^{e_1} \). Denote the winner of the sponsored slot at the first period as \( w_1 \). The local IC constraint gives that entrant \( e_1 \)’s equilibrium bid \( b^E (\tilde{\gamma}^{e_1}) \), which is also his payment for the sponsored slot at event \( \Lambda(b^E (\tilde{\gamma}^{e_1})) \), should be just as high as the expected benefit of winning conditional on \( \Lambda(b^E (\tilde{\gamma}^{e_1})) \), so that the entrant is indifferent about just winning or just losing the auction.

\[
b^E (\tilde{\gamma}^{e_1}) = \frac{E[V_1^{e_1} | w_1 = e_1, \Lambda(b^E (\tilde{\gamma}^{e_1}))] - E[V_1^{e_1} | w_1 \neq e_1, \Lambda(b^E (\tilde{\gamma}^{e_1}))]}{+E[V_2^{e_1} | w_1 = e_1, \Lambda(b^E (\tilde{\gamma}^{e_1}))] - E[V_2^{e_1} | w_1 \neq e_1, \Lambda(b^E (\tilde{\gamma}^{e_1}))]} \quad \text{([LIC])}
\]

Because \( V_1^{e_1} \) is independent of \( \Lambda(b^E (\tilde{\gamma}^{e_1})) \) given \( w_1 \), \( E[V_1^{e_1} | w_1 = e_1, \Lambda(b^E (\tilde{\gamma}^{e_1}))] = \Pi^s \tilde{\gamma}^{e_1} \) and \( E[V_1^{e_1} | w_1 \neq e_1, \Lambda(b^E (\tilde{\gamma}^{e_1}))] = 0 \). Now I focus on the parts with \( V_2^{e_1} \).

If entrant \( e_1 \) wins the sponsored slot, his probability of getting the organic slot in the next period is \( W^s (\tilde{\gamma}^{e_1}, \tilde{\gamma}^s) \). Conditional on winning, which means that his sales in the first period is above the threshold, he update his expected quality to \( \tilde{\gamma}^{e_1,w} (\tilde{\gamma}^{e_1}) \), which is a function of \( \tilde{\gamma}^{e_1} \). Hence,

\[
E[V_2^{e_1} | w_1 = e_1, \Lambda(b^E (\tilde{\gamma}^{e_1}))] = W^s (\tilde{\gamma}^{e_1}, \tilde{\gamma}^s) \Pi^o (\hat{\gamma}^o) \tilde{\gamma}^{e_1,w} (\tilde{\gamma}^{e_1})
\]

If entrant \( e_1 \) just loses the sponsored slot (i.e. at event \( \Lambda(b^E (\tilde{\gamma}^{e_1})) \)) the winning entrant must have posterior expectation \( \tilde{\gamma}^{e_h} \) the same as \( \tilde{\gamma}^{e_1} \). Hence, the probability for entrant \( e_1 \) to get the top organic slot in the next period is \( (1 - W^s (\tilde{\gamma}^{e_h}, \tilde{\gamma}^s)) / (N - 1) \) so

\[
E[V_2^{e_1} | w_1 \neq e_1, \Lambda(b^E (\tilde{\gamma}^{e_1}))] = \frac{(1 - W^s (\tilde{\gamma}^{e_1}, \tilde{\gamma}^s)) \Pi^o (\hat{\gamma}^o) \tilde{\gamma}^{e_1}}{N - 1}
\]
Plug in the four terms into (LIC), I have the equilibrium bidding function:

\[
b^E(\bar{\gamma}^{e_1}) = \Pi^s(\bar{\gamma}^s)\bar{\gamma}^{e_1} + W^s(\bar{\gamma}^{e_1}, \bar{\gamma}^s)\Pi^o(\bar{\gamma}^o)\bar{\gamma}^{e_1,w}(\bar{\gamma}^{e_1}) - \frac{(1 - W^s(\bar{\gamma}^{e_1}, \bar{\gamma}^s))}{N - 1}\Pi^o(\bar{\gamma}^o)\bar{\gamma}^{e_1}
\] (3)

I summarize the solved equilibrium by the following proposition.

**Proposition 2** On a platform with rule set $\mathcal{O}|\mathcal{S}$, the unique symmetric equilibrium consists of three elements: the expected conversion rate of the sponsored slot $\bar{\gamma}^s$, the expected conversion rate of the organic slot $\bar{\gamma}^o$ and the bidding function of entrants $b^E(\bar{\gamma}^{e_1})$, and they are specified by (1), (2) and (3), respectively.

Now I discuss the bidding function with more details. The first component in $b^E(\bar{\gamma}^{e_1})$ is the static incentive from profits in the current period. The other two components are the dynamic incentives from profits in the next period. Regardless of the number of entrants $N$, if $W^s(\bar{\gamma}^{e_1}, \bar{\gamma}^s) > \frac{1}{2}$, the dynamic incentive is positive. Combining Lemma 1 and Proposition 2 gives the following corollary

**Corollary 1** An entrant with expected quality of at least $\bar{\gamma}_0 + \frac{1}{\beta_0 + \pi_0}$ would always bid higher than the current-period benefit for the sponsored slot.

If $N$ is small, an entrant would weigh the third term substantially. For an entrant with very low expected quality, he understands that if he could win the sponsored slot, it must be that the other entrants also have very low expected quality, so for the purpose of earning the organic slot in the next period, it is better to let another entrant demonstrate its low quality in the sponsored slot. Hence, the dynamic incentive could be negative for a low quality entrant and cause bid-shading.

When $N$ is large, the third term is insignificant due to the fact that an entrant losing the auction has very small probability to be randomly assigned the organic slot in the next period. Then the entrant is only concerned with earning the top organic slot in the next round through winning the sponsored slot in this round, and the dynamic incentive is always positive.

In this environment, I have restricted the incumbents from bidding for the sponsored slot. If they were allowed to bid, for incumbents who did not had the sponsored slot in the previous period and therefore have not updated their beliefs, their bids as incumbents would simply equal to their static incentives in the previous period, which corresponds to the first term in (3). If an incumbent have not earned the organic slot after a period in the sponsored slot, it should bid even lower due to the downward information updating. Hence, when $N$
is large, the incumbents are most likely outbid by the top entrant, who has a very strong
dynamic incentive on top of its static incentive. Therefore, the restriction that only entrants
can bid is not very restrictive after all for large \( N \).

3.3 Platform with two organic slots

In this subsection, I consider an alternative environment where the platform does not take
any input from the seller and determine all the rankings with information it gains from
observed sales. In particular, the platform commits to rank the incumbent with the highest
posterior expected quality at the top organic slot on the left of the first page. The right
slot on the first page is now also an organic slot at the platform’s disposal. Assume that
the platform constantly randomizes between an incumbent and \( n \) randomly chosen entrants
within a period, and each entrant receives a \( \delta \leq \frac{1}{n} \) proportion of the total exposure of
the second organic slot, and the observed sales of the entrants are used by the platform to
compute posterior expectations for these entrants in the same Bayesian fashion as in the
previous subsection. Then for all entrants, the platform determines the entrants with the
highest and the second highest quality (with random tie breaking), which will be placed in
the top organic slot full-time and the second organic slot part-time, respectively. Denote
this set of rules as \( O(n, \delta) \).

![Diagram of a platform with two organic slots](image)

Figure 7. Diagram of a platform with two organic slots

This benchmark setup aims to capture the essential trade-off for welfare comparison.
With the sponsored slot in the environment \( O_S \), the platform loses control of a precious
prominent space, and delegates the assignment of this slot to an auction mechanism so as to
use the private information of the seller. The benchmark case gives that control back to the
platform so that more information gained from last periods can be used. Because now the
platform needs to place two incumbents at the first page, I also give the platform the ability
to experiment with even more entrants so as to discover more than one good products for the
next period. Also, whenever the platform has two good incumbents and need to divert some
exposure from them to experiment with entrants, it is always more efficient to first divert exposure from the incumbent of lower quality. Hence, the assumption that the platform split the second organic slot instead of the first is reasonable.

Since now the sellers have no actions to take, the equilibrium of this environment only involves consumers’ rational browsing behaviors. Assuming consumers cannot distinguish the platform’s randomization, they would perceive the expected quality of the second organic slot as the weighted average of the $n + 1$ products. Hence, the equilibrium consists of the expected conversion rates of the first and second organic slots, $\bar{\gamma}^{o_1}$ and $\bar{\gamma}^{o_2}$, respectively. In this equilibrium, the first and second organic slots receive exposure $\Pi^{o_1}(\bar{\gamma}^{o_1}) = \bar{\gamma}^{o_1}BM$ and $\Pi^{o_2}(\bar{\gamma}^{o_2}) = \bar{\gamma}^{o_2}BM$.

Given $\Pi^{o_2}(\bar{\gamma}^{o_2})$, an entrant displayed will receive exposure $\delta \Pi^{o_2}$, and its number of sales, $Q_j$, according to the gamma prior of the conversion rates, follows a negative binomial distribution $NB\left(\alpha_0, \frac{\delta \Pi^{o_2}(\bar{\gamma}^{o_2})}{\delta \Pi^{o_2}(\bar{\gamma}^{o_2}) + \beta_0}\right)$ and its expected conversion rate will be updated to $\frac{Q_j + \alpha_0}{\delta \Pi^{o_2}(\bar{\gamma}^{o_2}) + \beta_0}$. Hence, the sample of the updated expected conversion rates of all $N$ entrants consists of $n$ draws from this distribution transformed from a negative binomial distribution, and $N - n$ fixed elements equals to $\gamma^0$. Denote the expectations of the first and second highest elements of this sample, defined in terms of $\bar{\gamma}^{o_2}$ and fundamental parameters, as $\bar{\gamma}^{(N)}(\bar{\gamma}^{o_2}, \alpha_0, \beta_0, B, M)$ and $\bar{\gamma}^{(N-1)}(\bar{\gamma}^{o_2}, \alpha_0, \beta_0, B, M)$.

The following proposition characterizes the equilibrium with the above notation:

**Proposition 3** On a platform with rule set $O(n, \delta)$, in equilibrium, the expected conversion rate of the second organic slot, $\bar{\gamma}^{o_2}$, is the fixed point of the following transformation:

$$\bar{\gamma}^{o_2} = (1 - \delta n) \bar{\gamma}^{(N-1)}(\bar{\gamma}^{o_2}, \alpha_0, \beta_0, B, M) + \delta n \gamma^0$$

and the expected conversion rate of the first organic slot, $\bar{\gamma}^{o_1}$, is

$$\bar{\gamma}^{o_1} = \bar{\gamma}^{(N)}(\bar{\gamma}^{o_2}, \alpha_0, \beta_0, B, M)$$

In the next subsection, I will numerically compute $\bar{\gamma}^{o_2}$ and $\bar{\gamma}^{o_1}$.

### 3.4 Welfare comparison

In this subsection, I compare the equilibria of the two environments from a welfare perspective. The total welfare is proportional to the total number of transactions. In the first environment, $O(S)$, the total number of transaction is

$$\Psi_{O(S)} = \Pi^{o}\bar{\gamma}^{o} + \Pi^{s}\bar{\gamma}^{s} = \left[\frac{(\bar{\gamma}^{o})^2 + (\bar{\gamma}^{s})^2}{\bar{\gamma}}\right]BM.$$
Similarly, in the second environment, $OQ(n, δ)$, the total number of transaction is

$$\Psi_{OQ(n, δ)} = \Pi^{o1} \hat{\gamma}^{o1} + \Pi^{o2} \hat{\gamma}^{o2} = \left[ (\hat{\gamma}^{o1})^2 + (\hat{\gamma}^{o2})^2 \right] BM$$

Hence, in the rest of this subsection, I will focus on comparing $\hat{\gamma}^{o}$ to $\hat{\gamma}^{o1}$ and $\hat{\gamma}^{s}$ to $\hat{\gamma}^{o2}$.

First, I focus on the first environment. The analysis of the bidding functions shows that the sellers’ private information are incorporated in the ranking mechanism to improve accelerate quality discovery, so the qualities of the organic and the sponsored slots depends on the amount of the entrants’ private information, which is measured by $\pi_0$.

Below is a graph depicting the equilibrium quality levels, $\hat{\gamma}^{s}$ and $\hat{\gamma}^{o}$, of the organic slot and the sponsored slot in the first environment $O|S$, as the power of the private information, $\pi_0$, varies. Other parameters are fixed at $N = 8, \alpha_0 = 0.1, \beta_0 = 10, B = 20, M = 1000$.

![Figure 8. Equilibrium quality levels in the environment with a sponsored slots](image)

The graph shows that both $\hat{\gamma}^{o}$ and $\hat{\gamma}^{s}$ increases in $\pi_0$. The improvement in $\hat{\gamma}^{s}$ is purely due to better private information of the entrants. The improvement in $\hat{\gamma}^{o}$ can be attributed to two factors: First, the higher is $\pi_0$, the more likely that it is the actually better entrant is allocated at the sponsored slot. The sponsored exposure thus can better help the platform to recognize the high quality. Second, because equilibrium exposure of the sponsored slot $\Pi^{s}$ is linear function of $\hat{\gamma}^{s}$, which increases in $\pi_0$, more consumers are willing to click the sponsored product, therefore the platform just receive better information regarding the quality of that product.

Second, I focus on the second environment with two organic slots. I plot the expected
quality of the two slots against $n\delta$ for $n = 2, 4, 8$ in Figure 9 and Figure 10. Other parameters are fixed at the same values, i.e. $N = 8, \alpha_0 = 0.1, \beta_0 = 10, B = 20, M = 1000$. Across the range of $n\delta$, the quality of the first organic slot increases in $n\delta$ in Figure 9 since more exposure is used to infer the qualities of the entrants.

![Figure 9. Equilibrium quality of the first organic slot in the environment with only organic slots](image)

The quality of the second organic slot is non-monotonic in $n\delta$ in Figure 10 due to the trade-off between two usages of the exposure of the second organic slot: obtaining information for entrants sacrifices the chance to utilize those obtained information. In fact, at the current set of parameters, because exposure is so scarce, the second best incumbent inferred from data is not so much better than the average\(^5\), so even the peaks of these curves are very close

\(^5\)For $n = 8$, if too little exposure is allocated to entrants such that the expected sales according to the prior is close to zero, the second best incumbent would likely have zero sales and therefore have an expected quality that is even lower than the mean of the prior, which explains the downward bend of the solid curve for $n = 8$. 

22
to the prior mean, $\gamma^0 = 0.01$.

Hence, in the interest of total number of transactions $\Psi_{O(n,\delta)}$, it is the best to allocate all second slot exposure for informational purpose by choosing $n\delta = 1$. Also, Figure 10 suggests it is beneficial to experiment with all entrants. These observations will not hold true when the number of products and the number of consumers are both very large. In that case information is no longer scarce, so the platform can easily identify a second best incumbent of very high quality, and the marginal informational benefit from more exposure for entrants is small, which warrants experimenting with a subset of entrants using partial exposure.

Finally, I compare the two environments. The expected quality levels in Figure 8, 9 and 10 show that, even at the optimal set of $n$ and $\delta$, the environment $O(n,\delta)$ is dominated by the environment $O(S)$. The second organic slot in the second environment cannot match the informativeness of the sellers’ private signal, so using the second slot for random experiments leads to very low expected quality in comparison to that of the auction winner.

The informativeness of the private signals is comparable to an amount of exposure of $\pi_0 N$, while the exposure received by the second organic slot in the second environment is slightly above $\gamma^0 BM$. The first environment would have an advantage over the second if $\pi_0 N$ is large or if $\gamma^0 BM$ is small.

The table below gives more comparisons for the expected qualities as the number of products $N$ and the number of consumers $M$ vary. For all entries, I fix $\alpha_0 = 0.1, \beta_0 = 10, B = 20$. For the entries for the environment $O(S)$, I fix $\pi_0 = 30$; for all entries for the environment $O(n,\delta)$, I choose $n = N$ and $\delta = 1/N$. To save space, the numbers are in
percentage.

| $N=4$  | $O|S$  | $\tilde{\gamma}^o:3.01$ | $\tilde{\gamma}^s:2.93$ | $M=500$ | $O|S$  | $\tilde{\gamma}^o:4.96$ | $\tilde{\gamma}^s:4.92$ |
|--------|--------|------------------|------------------|--------|--------|------------------|------------------|
| $O|(n,\delta)$  | $\tilde{\gamma}^{o1}:3.11$ | $\tilde{\gamma}^{o2}:1.00$ | $O|(n,\delta)$  | $\tilde{\gamma}^{o1}:4.14$ | $\tilde{\gamma}^{o2}:1.00$ |

| $N=8$  | $O|S$  | $\tilde{\gamma}^o:4.97$ | $\tilde{\gamma}^s:4.93$ | $M=1000$ | $O|S$  | $\tilde{\gamma}^o:4.97$ | $\tilde{\gamma}^s:4.93$ |
|--------|--------|------------------|------------------|--------|--------|------------------|------------------|
| $O|(n,\delta)$  | $\tilde{\gamma}^{o1}:4.79$ | $\tilde{\gamma}^{o2}:1.00$ | $O|(n,\delta)$  | $\tilde{\gamma}^{o1}:4.79$ | $\tilde{\gamma}^{o2}:1.00$ |

| $N=12$ | $O|S$  | $\tilde{\gamma}^o:6.50$ | $\tilde{\gamma}^s:6.49$ | $M=2000$ | $O|S$  | $\tilde{\gamma}^o:4.96$ | $\tilde{\gamma}^s:4.92$ |
|--------|--------|------------------|------------------|--------|--------|------------------|------------------|
| $O|(n,\delta)$  | $\tilde{\gamma}^{o1}:5.85$ | $\tilde{\gamma}^{o2}:1.00$ | $O|(n,\delta)$  | $\tilde{\gamma}^{o1}:5.22$ | $\tilde{\gamma}^{o2}:1.00$ |

The left panel of the table shows that the environment $O|S$ benefits more from increasing numbers of entrants. In this environment, more entrants means more draws of private signals, and the best private signal improves with the size of the pool. For the environment $O|(n,\delta)$, since the exposure that can be allocated to the entrants is limited, splitting the exposure to more entrants reduce the accuracy of the platform’s information on each entrant. Hence, at higher $N = 12$, the environment $O|S$ outperforms the other by a large margin, while at lower $N = 4$, the environment $O|S$ underperforms.

The right panel of the table shows that the performance of the environment $O|(n,\delta)$ depends on the number of consumers, which directly affects how much exposure the platform can use to test the qualities of the entrants. The environment $O|S$ does not react much to the number of consumers in the range of comparison, since the exposure of the sponsored slot is only used to determine whether that particular product is of good quality, which can be accomplished quite accurately with modest exposure.

### 4 Data

My data is obtained from a third party data service, Zhibi365.com, which uses a cluster of servers to record public information on Taobao.com on a daily basis. Every day, Zhibi365 conducts searches on Taobao.com and its mobile app using a list of 500,000 keywords provided by Taobao.com, and records all the information on the search result pages (top 6 pages for the PC website and top 100 products for the mobile app), including the organic and sponsored lists of products, and for each product, its title, price and sales in the past 30 days as displayed in the search results. The information collected by Zhibi365.com is sorted out by product ID and seller ID, so that users of Zhibi365.com can query all the relevant ranks and other daily records of a given product or all the products of a given seller. Zhibi365.com also allows to query the ranks by sales within subcategories defined by Taobao.com.

Given Taobao.com’s vast landscape and my research goal, I decided to draw data from the women’s apparel category, which features large number of nonbranded products with
frequently new product introductions, so that timely quality information is important to consumers and challenging for the platform. The seasonality guarantees a large number of new products in the early fall season.

Because Zhibi365.com organizes its data and queries based on seller ID and product ID, I first established a pool of products by recording the top 500 best selling products in the women’s sweater category on Dec 03 2015, on which day the total sales of women’s sweater category peaked. Then I tracked the sellers of these 500 products, and for each seller, I downloaded all the data relevant for this seller from Zhibi365.com, including the products it sold over time and the ranks of those products as discussed earlier. I kept the products that had appeared for more than 60 days in the ranks data, and ended up with 1500 products. From the user interface of Zhibi365.com, I collected daily data series covering seven months for these products.

The platform, Taobao.com, officially releases cross-sectional summary data on the 500,000 keywords. The list is intended for the sellers to find keywords to purchase sponsored ads for, so for each keyword, Taobao.com gives impression index (number of searches conducted for the keyword), click index (average number of clicks of a sponsored product), average click-through rate of sponsored product, average conversion rate of sponsored product, competitiveness for sponsored slots, and average per-click fee for sponsored ad. All of the rates are average values that aim to provide a relative idea about the effectiveness of the keyword in order to facilitate sellers purchasing sponsored ads. Zhibi365.com merges this data with its keyword search data and show them together, so I have collected this data from Zhibi365.com as well.

The descriptive statistics are presented in Table 1. The first two variables, Price and 30DaySales, are observed once per day for each product. The mean of prices (about $10) confirms that most products are not of premium brands and therefore the effect of marketing outside of the platform is minimal. Most products do not change price over time. Some observations of prices are missing, but I will not use price information due to the great variation among the products. Also, On average these products have very substantial sale numbers, so the platform should care about conversion rates and also has reasonable data to infer conversion rates.

I scraped four types of ranks from Zhibi365.com, including both organic and sponsored ranks on the PC website and the mobile app interfaces. The four types of ranks have different supports due to truncation in data collection done by Zhibi365.com. Observed ranks appear to be uniformly distributed in the data, even after controlling for popularity of keywords, which in fact suggests significant heterogeneity among products’ rankings: the products ranking in top slots in popular keywords are receiving much more exposure than
the products that not even show up in the observed range for those keywords.

Keyword-level cross-section variables are released by Taobao.com. While the impression index just measures how many times the keyword is searched, the click index and the average click-through rates and conversion rates are measured only for the sponsored products on the PC website. Taobao.com does not disclose the relevant length of the time interval used in computing the impression index, and the best guess is a week. The three variables specific to the sponsored products indicate that click-through rates are quite low and that the conversion rates are even lower.

Table 2. Variables and Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min.</th>
<th>Max.</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Product-level time-series variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>Price in CNY</td>
<td>62.4</td>
<td>47.2</td>
<td>9.9</td>
<td>518</td>
<td>43,279</td>
</tr>
<tr>
<td>30DaySales</td>
<td>Quantity sold in past 720 hours</td>
<td>758</td>
<td>2,195</td>
<td>0</td>
<td>39,109</td>
<td>71,556</td>
</tr>
<tr>
<td><strong>Product-keyword-level time-series variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(rank_{p,o})</td>
<td>Organic search rank on PC</td>
<td>97.6</td>
<td>57.5</td>
<td>1</td>
<td>243</td>
<td>812,711</td>
</tr>
<tr>
<td>(rank_{p,s})</td>
<td>Sponsored search rank on PC</td>
<td>51.76</td>
<td>33.9</td>
<td>1</td>
<td>102</td>
<td>(1 \times 10^6)</td>
</tr>
<tr>
<td>(rank_{m,o})</td>
<td>Organic search rank in mobile app</td>
<td>40.7</td>
<td>26.5</td>
<td>1</td>
<td>119</td>
<td>777,450</td>
</tr>
<tr>
<td>(rank_{m,s})</td>
<td>Sponsored search rank in mobile app</td>
<td>36.3</td>
<td>26.6</td>
<td>1</td>
<td>111</td>
<td>234,547</td>
</tr>
<tr>
<td><strong>Keyword-level cross-section variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impression-Index</td>
<td>Relative measure of displays during an interval of an undisclosed length</td>
<td>40,077</td>
<td>(7 \times 10^5)</td>
<td>0</td>
<td>(8 \times 10^7)</td>
<td>44,117</td>
</tr>
<tr>
<td>ClickIndex</td>
<td>Relative measure of clicks to sponsored products during the same interval</td>
<td>525</td>
<td>3,381</td>
<td>0</td>
<td>(2 \times 10^5)</td>
<td>44,117</td>
</tr>
<tr>
<td>AveCTR</td>
<td>Average percentage click-through rate</td>
<td>2.52</td>
<td>2.29</td>
<td>0</td>
<td>100</td>
<td>44,117</td>
</tr>
<tr>
<td>AveCR</td>
<td>Average percentage conversion rate</td>
<td>0.73</td>
<td>3.56</td>
<td>0</td>
<td>100</td>
<td>44,117</td>
</tr>
</tbody>
</table>

Two features of the data are of important empirical relevance. First, as mentioned in the empirical setting section, the platform has a mechanism to automatically vary each product’s ranks in the organic results by a 7 day cycle starting at a time specified by the seller. In particular, a product is gradually moved up in the organic ranks as time moves towards the end of the 7 day cycle, which marks a nominal expiration time of the listing. At the expiration time, the listing is immediately and automatically reinstated, and a new cycle starts as the product is moved down in the organic ranks abruptly. This exogenous rank change helps to identify the effect of organic ranks on sales.

Second, one limitation of the available quantity data is that it consists of daily records of 30DaySales, but the records are not necessarily taken at the same time on each day. In the empirical analysis I will use a flexible specification that allows imperfectly recovered daily sales to be properly related to the observed daily ranks.
5 Effects of sponsored and organic ranks on sales

In this section I connect search ranks to the sales they are responsible for. The unique 7-day asynchronous shuffling feature of the platform creates high-frequency, broad-support and, most importantly, exogenous variation in organic ranks, which plays a crucial role in the identification of the effect of the organic ranks. Because the platform gives the 7-day timer different influences on the PC website and the mobile app interface, the effects of the two types of organic ranks can be reliably separated. The sharp day-to-day changes also allow studying the delay between consumers seeing search results and making purchase. For the sponsored ranks, the exogenous variation is primarily supplied by the frequent involuntary rank changes due to other sellers’ bids.

Depending on the nature (organic or sponsored) and the interface (desktop or mobile) of the rank, the four types of ranks could have different effects on sales that reflect interesting patterns of consumer behavior. The estimation in this part is to be interpreted in tandem with the stylized model in Section 3. First, inspired by the stylized model, the empirical model in Section 5.1 specifies a linear relation between exposure and sales, with the coefficient representing a relative measure of the product’s quality, which will allow relating quality to rank dynamics in Section 5.3. Second, the estimated exposure measure by ranks can testify whether the stylized model, simplified with just two slots, is conceptually reflective of the actual market on Taobao.com. Section 5.2 discusses the relevant evidence supporting the setup of the stylized model.

5.1 Specifying effects of ranks on sales

Inspired by the conversion rate model, I explore a specification identifying the following relationship:

\[ Sales_{j,t} = \gamma_j E_{j,t} + \mu_{j,t} + \varepsilon_{j,t} \]

where \( Sales_{j,t} \) is the sale happens on date \( t \), \( E_{j,t} \) is the product’s exposure through searching results, \( \mu_{j,t} \) is the product’s expected sales from other unobserved channels, and \( \varepsilon_{j,t} \) is an error term reflecting randomness and variation in the sales process.

The parameters to be estimated are \( \{\gamma_j\} \), parameters generating \( \mu_{j,t} \) and parameters generating \( E_{j,t} \) from the product’s ranks. Now I specify the parameters leading to \( E_{j,t} \) and \( \mu_{j,t} \).

For each of a product’s four ranks (\{p(online), m(obile)\} by \{o(static), s(ponsored)\}), I estimate a rank function to measure how much exposure decreases as the rank number
increases. The rank functions have similar form

\[
g^{p,o}(\text{rank}^{p,o}_{j,k,t}) = \left[ (\alpha^{p,o} + \text{rank}^{p,o}_{j,k,t}) / (\alpha^{p,o} + 1) \right]^{-\beta^p}
\]
\[
g^{p,s}(\text{rank}^{p,s}_{j,k,t}) = \delta^{p,s} \left[ (\alpha^{p,s} + \text{rank}^{p,s}_{j,k,t}) / (\alpha^{p,s} + 1) \right]^{-\beta^p}
\]
\[
g^{m,o}(\text{rank}^{m,o}_{j,k,t}) = \delta^{m,o} \left[ (\alpha^{m} + \text{rank}^{m,o}_{j,k,t}) / (\alpha^{m} + 1) \right]^{-\beta^m}
\]
\[
g^{m,s}(\text{rank}^{m,s}_{j,k,t}) = \delta^{m,s} \left[ (\alpha^{m} + \text{rank}^{m,s}_{j,k,t}) / (\alpha^{m} + 1) \right]^{-\beta^m}
\]

The exposure at the top PC website organic slot is normalized to 1, and \( \delta^{p,s}, \delta^{m,o}, \) and \( \delta^{m,s} \) measures the exposure of top slots in other types of ranks. The logic of imposing some parameters to be common across rank functions is to respect the way webpage/smartphone app displays the search results. For the PC results, because the organic and sponsored products are displayed distinctively in two separated areas, they could potentially have very different rank functions. However, because they are grouped into pages, so for large rank numbers, exposures of these two types of results should decay in a proportional way. Hence, I impose that the two function shares their \( \beta^p \), but allows them to have different \( \alpha^{p,o} \) and \( \alpha^{p,s} \), so that for small ranks, the two function could have quite different shape. For the mobile results, because the sponsored product is blended into the flow of products, it is natural to use a single multiplicative factor to account for the effect of the sponsored products and keep the shape of the functions (determined by \( \alpha^{m} \) and \( \beta^{m} \)) the same.

To aggregate the exposure from multiple keywords, I use the relative average number of clicks, \( \text{ClickIndex}_k \), to weigh the thousands of keywords. So the exposure from the same type of ranks can be aggregated over keywords, for example:

\[
E^{p,o}_{j,t} = \sum_k \text{ClickIndex}_k g^{p,o}(\text{rank}^{p,o}_{j,k,t})
\]

I then aggregate for PC and mobile interfaces separately

\[
E^p_{j,t} = E^{p,o}_{j,t} + E^{p,s}_{j,t}
\]
\[
E^m_{j,t} = E^{m,o}_{j,t} + E^{m,s}_{j,t}
\]

Because consumers could delay purchase after viewing the product page, the sales in date \( t \) may be related to exposures in previous days, and this pattern of delay could be different for consumers searching through PC and those searching through mobile. I aggregate the
total effective exposure that is responsible for sales between \( t \) and \( t - 1 \) as

\[
E_{j,t}^{all} = E_{j,t}^p + \tau_{-t-1}^p E_{j,t-1}^p + \tau_{-t-2}^p E_{j,t-2}^p + \tau_{-t-3}^p E_{j,t-3}^p \\
E_{j,t}^m + \tau_{-t-1}^m E_{j,t-1}^m + \tau_{-t-2}^m E_{j,t-2}^m + \tau_{-t-3}^m E_{j,t-3}^m
\]

Ideally, if the data collecting actions are exactly aligned and every consumer who would purchase do so without delay, I would expect \( \tau_{-1}^p = \tau_{-1}^m = 1 \) while \( \tau_{-2}^p = \tau_{-2}^m = \tau_{-2}^m = \tau_{-3}^m = 0 \). However, due to consumers delay in purchase action, ranks from earlier days can potentially play a role here. This specification is also robust to misalignment in data collection time.

I also notice the possible effect of \( 30\text{DaySales} \) to consumer demand. First, it is possible that consumers are more likely to click/purchase if \( 30\text{DaySales} \) is higher, especially when the product is in the initial stage and \( 30\text{DaySales} \) is relatively small. Hence, both \( E_{j,t}^{all} \) and the constant term will be adjusted by \( 30\text{DaySales}^\kappa \) where \( \kappa \) is expected to be in \((0, 1)\). Second, because the product I am looking at are seasonal, so while my story is good at explaining how the product get advertised and discovered, I rely on \( 30\text{DaySales} \) to model the product’s obsolescence: after the day at which \( 30\text{DaySales}_{j,t} \) peaks (peakDay), I shrink the effect of exposure and the constant term exponentially. Hence, after these two adjustment, I have

\[
\mu_{j,t} = 30\text{DaySales}_{j,t}^\kappa e^{-\kappa \max(t-\text{peakDay}, 0)} \\
E_{j,t} = E_{j,t}^{all} \mu_{j,t}
\]

Next, I deal with the left-hand-side variable, the daily sales. The sales data available is the sum of sales made in the past 30 days, recorded daily. To transform this data into daily sales, I take the difference of consecutive observations. Ideally, if \( 30\text{DaySales} \) is collected every 24 hours, the difference between two consecutive observations of \( 30\text{DaySales} \) should just be the difference between the sales of two days that are 30 days apart.

\[
30\text{DaySales}_{j,t} - 30\text{DaySales}_{j,t-1} = Sales_{j,t} - Sales_{j,t-30}
\]

However, due to the randomness in data collection time, \( 30\text{DaySales}_{j,t} - 30\text{DaySales}_{j,t-1} \) could represent the difference between sales during two intervals that have equal lengths of \( \varphi_{j,t} \) days, where \( \varphi_{j,t} \) is a random number in \((0, 2)\) with \( E[\varphi_{j,t}] = 1 \). Hence,

\[
30\text{DaySales}_{j,t} - 30\text{DaySales}_{j,t-1} = \varphi_{j,t} (Sales_{j,t} - Sales_{j,t-30})
\]
Bring in the RHS variables, I have

\[ 30\text{DaySales}_{j,t} - 30\text{DaySales}_{j,t-1} = \varphi_{j,t} (\text{Sales}_{j,t} - \text{Sales}_{j,t-30}) + \gamma_j (E_{j,t} - E_{j,t-30}) + \mu_{j,t} - \mu_{j,t-30} \]

\[ = \gamma_j (E_{j,t} - E_{j,t-30}) + (\mu_{j,t} - \mu_{j,t-30}) + (\varphi_{j,t} - 1) (E_{j,t} - E_{j,t-30}) + (\varphi_{j,t} - 1) \mu_{j,t-30} + \epsilon_{j,t} + \epsilon_{j,t-30} \]

Because \( E[\varphi_{j,t}] = 1 \) and \( \varphi_{j,t} \) is independent from \( E \) and \( \mu \), \( (\varphi_{j,t} - 1) \gamma_j (E_{j,t} - E_{j,t-30}) + (\varphi_{j,t} - 1) (\mu_{j,t} - \mu_{j,t-30}) \) can be counted as part of the error term in a least square regression.

Now, the empirical model is completely specified and it can be estimated by a nonlinear least square regression of the differences in consecutive 30-DaySales on 30-day differences in exposure \( E_{j,t} \) and extra sales \( \mu_{j,t} \). Because \( (\varphi_{j,t} - 1) \gamma_j (E_{j,t} - E_{j,t-30}) + (\varphi_{j,t} - 1) (\mu_{j,t} - \mu_{j,t-30}) \) is treated as part of the error term, for observations with greater \( (E_{j,t} - E_{j,t-30}) \) and \( (\mu_{j,t} - \mu_{j,t-30}) \), the error term has higher variance, which creates a heteroskedasticity problem. I apply the feasible least square method, that is, I first estimate the model without weighting, and then weigh each observation by the inverse of the squared residuals from the first round of estimation and run estimation again.

### 5.2 Results

The final estimates are presented below. I now discuss the interesting pieces of evidence regarding the consumers’ usage of the platform, in connection to dynamic story told in the stylized model.

<table>
<thead>
<tr>
<th>Table 3. Estimated Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PC exposure parameters</strong></td>
</tr>
<tr>
<td>Organic Date Weight</td>
</tr>
<tr>
<td>( \tau_{p}^{0} )</td>
</tr>
<tr>
<td>0.60</td>
</tr>
<tr>
<td>( \alpha^{p,o} )</td>
</tr>
<tr>
<td>5.14</td>
</tr>
<tr>
<td>( \beta^{p} )</td>
</tr>
<tr>
<td>0.39</td>
</tr>
<tr>
<td>( \tau_{p}^{1} )</td>
</tr>
<tr>
<td>0.55</td>
</tr>
<tr>
<td>( \tau_{p}^{2} )</td>
</tr>
<tr>
<td>( \tau_{p}^{3} )</td>
</tr>
<tr>
<td>Sponsored</td>
</tr>
<tr>
<td>( \delta^{p,s} )</td>
</tr>
<tr>
<td>306.8</td>
</tr>
<tr>
<td>( \alpha^{p,s} )</td>
</tr>
<tr>
<td>4.99</td>
</tr>
<tr>
<td>( \kappa )</td>
</tr>
<tr>
<td>0.428</td>
</tr>
</tbody>
</table>
5.2.1 The rank functions: PC website vs. mobile

The shape of the rank functions are best understood through graphs. The graph below is the percentage distribution of exposures over the top 100 organic ranks for the two interfaces.

![Graph showing percentage distribution of organic exposures among top 100 ranks, website vs. mobile.](image)

There is a stark contrast between organic results on PC and on Mobile. The exposure is quite dispersed across organic ranks on the PC website, while in the mobile app, exposure diminishes very fast and most exposure is contributed by top ranks. This is consistent with the difference in consumer behaviors on the two terminals. Desktop consumers, with the aid of the monitor and mouse, can easily scroll through many products, and open multiple tabs to examine in detail. Mobile consumers who have to scroll up and down on small screen can hardly go too far down the list. Also on mobile the consumer has to examine the products sequentially, so patience may wear out fast.

5.2.2 Relative importance of organic ranks: PC website vs. mobile

The estimate of $\delta^{m,o}$ is quite large. After adjusting for the mobile organic ranking’s faster decline in exposure, it suggests that for a given keyword, the total exposure from the top 100 organic ranks in the mobile app is four times as high as the total exposure from the top 250 organic ranks on the PC website. One important reason is the boost of mobile app
usage. Because more people have mobile phones and mobile shopping can utilize fractured
time, the platform now indeed has more transactions on mobile than on PC. However, usage
ratio alone seems not enough to explain the huge difference, which may actually has to do
with how the platform arrange the organic results on PC website and the mobile app.

On the PC side, the platform does not rank the products very strictly by their quality,
potentially because consumers can look through large number of products easily. One piece
of evidence is that on the PC website, the nominal once-per-week expiration time has a much
stronger effect on products’ organic ranks than on the mobile app. Since the nature of this
mechanism is to intentionally shuffle most products’ ranking so that every product gets some
exposure, it certainly reduces the general quality of the organic ranking on the PC website.
By way of comparison, on the mobile side, since consumers have limited attention span over
the ranks, the platform is very careful about the ranks, and the ranks, despite having 7-day
cycles as well, is much more stable.

Another piece of evidence is that, despite Zhibi365.com collects data on about top two
hundred slots on the PC website, twice as many as on the mobile app, products in my sample
get similar number of records on the PC website and on the mobile app. Since my sample
consist of potentially more serious sellers, it is suggested that the PC website rankings are
more dispersed across products and sellers, and are therefore of lower expected quality.

If the mobile organic search ranking is more reliable and informative, consumers should
be more willing to click the results on the mobile app, which is indeed suggested by the
estimates. This is consistent with the equilibrium assumption in the stylized model that
consumers’ willingness-to-view increases with the expected quality of the products presented
at a given slot.

5.2.3 Relative importance of sponsored ranks: PC website vs. mobile

There are much fewer number of sponsored slots in the mobile app than on the website. Also,
the difference between $\beta^m$ and $\beta^p$ suggest that exposure diminishes much faster with ranks
in the mobile app. Hence, although the estimated scaling coefficient $\delta^{m,s}$ for the mobile
sponsored ranks is about three times as high as $\delta^{p,s}$ for the PC website sponsored ranks,
the estimates suggest that for a given keyword, the total sponsored exposure of the top 100
sponsored ranks is about six times as high as that of the sponsored ranks among the top 100
ranks in the mobile app. The desktop users, despite being outnumbered by mobile users,
are contributing more clicks to sponsored products, which is consistent with their greater
browsing capacity. Because the platform can use the data from the PC website sponsored
slots to learn about the products’ quality and improve the mobile organic rankings, the
desktop users have generated a positive externality on the mobile users.
5.2.4 Relative importance of organic and sponsored ranks

Given the very large estimate of the scaling coefficient for the sponsored ranks on the PC website, $\delta^{p,s}$, the estimates suggest the exposure related to the organic results on the PC website is dwarfed by that of the sponsored results. This is somewhat consistent with the consumers’ rational expectations of the qualities of the products in the two lists. While the organic results are shuffling, because the sponsored results are generated by auction, it is likely the top places are held by aggressive advisors, so it is very possible that the average quality of sponsored products is higher than that of the organic ones, so the consumers could pay even more attention to the sponsored products.

Since both $\delta^{p,s}$ and $\delta^{m,s}$ are very large even relative to $\delta^{m,o}$, it is possible that the sponsored exposure estimates are biased upwards due to an endogeneity issue. Because sponsored advertising is not the only paid promotional service on Taobao.com, when sellers purchase sponsored ads, they may simultaneously invest in other types of promotional service that boost sales. If that’s the case, the effects of other promotional services could be mistaken as part of the effect of the sponsored ranks and cause bias.

5.2.5 Purchase delay and the effect of 30DaySales

Other parameters appear to be intuitive. The $\tau$ parameters models consumers delaying purchase, and their estimates shows exposure today can convert to sales three days later, which echoes the finding of Blake, Nosko and Tadelis (2016) that consumers on ebay on average conduct searches over a period of three days for each purchase.

Finally, the first 30DaySales parameter $\kappa$ confirms that 30DaySales have a very concave effect on consumer purchase. When 30DaySales is small, consumers are skeptical and a boost in 30DaySales is very important; When the product is already well established, the consumers do not respond to 30DaySales much. The second 30DaySales parameter $\chi$ describes a reasonable fade-out stage possibly due to seasonality.

5.2.6 Relation to the stylized model

Several aspects of the estimates echo the theoretical model. First, the organic exposure across the platform is highly concentrated to the top ranks in the mobile app, highlighting that the emergence of the mobile interface brings the market closer to the narrative of the limited number of slots in the stylized model. The concentration means there is very limited exposure left for other organic ranks. Given the large number of products, exposure for new product is very scarce. According to the welfare comparison in Section 3.4, these features give edge to the environment with sponsored slots relative to the environment without those.
Second, the seven-day shuffling in the organic ranks, especially that on the PC website, is similar to the environment without sponsored slot described in Section 3.3. The estimates find exposure received by shuffled PC website organic rankings is dwarfed by that of the more stable mobile organic rankings; therefore, consistent with the simulation in Section 3.4, this shuffling seems to discourage consumer clicks on the PC website by lowering the information value in the rankings. The very quick shuffling on the PC website gives each product just a small amount of organic exposure, while for products that purchase sponsored ads, they receive far more exposure from sponsored slots than from their shuffled organic slots. These results supports that the sponsored slot mechanism has a much prominent role in new product discovery that the "random experiment" allowed by the seven-day cycles.

5.3 Dynamic Patterns

The empirical estimates provide a way to aggregate a product’s exposure of the same type across keyword-rank pairs. In the graph below, I translate a product’s appearance in keyword searches into its aggregate organic and sponsored exposure using the estimated empirical model. Note that each exposure type is normalized by its average for this product. The curves demonstrate the dynamics postulated in the stylized model: The product first purchased sponsored exposure (the dashed curve) aggressively. Meanwhile, its organic exposure (the solid curve) increases gradually. When the organic exposure stabilizes at a high level, the product stops purchasing sponsored exposure. While this is the pattern of a particular product, many other products share the same dynamic features.

![Figure 11. Dynamic exposure patterns of a product](image)

Figure 11. Dynamic exposure patterns of a product

Potentially due to the endogeneity problem for the sponsored exposure and the limited
data size, the relative conversion rates estimated from the empirical model demonstrate very large variance, making it difficult to relate the dynamic patterns of organic and sponsored exposure to the products’ conversion rates in a statistical framework.

6 Conclusion

In this paper, I have analyzed the sponsored ads scheme on a retail platform as an informational channel that incorporates the sellers’ private quality signals to accelerate the discovery of high-quality entrants. The theoretical model finds this channel to be prominent when the private quality signals are strong, the number of new products is high and consumer exposure is relatively scarce. Using data on publicly observable search ranks and 30-day sales, I have dissected the relation between various types of search ranks and the sales they are responsible for, which echoes the theoretical analysis’ characterization of the informational landscape.

Using data on both the PC website and the mobile app interfaces, my empirical analysis identifies striking differences and an interesting synergy between the two. For a given keyword, the major organic ranks on the PC website only contribute a quarter as many sales in total as the major organic ranks in the mobile app. In contrast, the major sponsored ranks on the PC website contribute six times as many sales in total as the major sponsored slots in the mobile app. The ranks functions suggest that consumers using the PC website are less affected by the rankings and click through a broad set of products, generating quality-relevant data through purchase decisions. The mobile consumers are found to rely heavily on the rankings and to only pay attention to the top results, so they represent the prize for top products that have demonstrated their qualities. This prize in return prompts sellers to compete for the sponsored exposure, which is primarily from the PC website.

Online platforms have revolutionized the supply side of many markets and enabled small or even individual suppliers to provide products or services at affordable transaction costs. On the demand side, online platforms connect to consumers more and more through convenient yet limited interfaces: the mobile app is a prominent example, and other rising venues with limited interactions include smart TVs, virtual assistants and connected devices taking voice commands. The clash of the vast selection of goods and the limited information processing ability of the users highlights the informational role of the platforms, which is to provide high-quality recommendations. The tech industry has already advanced very far in schemes collecting and analyzing information from the demand side, such as user reviews and personalized recommendations based on user histories. The example of sponsored advertising on Taobao.com, from a different perspective, suggests that encouraging the supply
side to provide information in an incentive compatible way could be an equally important frontier to explore.
References


A Appendix

Proof of Lemma 1. For the first part, given \( \hat{\gamma}^s \), entrant \( e_1 \) would earn the top organic slot in the next period if and only if the platform’s posterior expectation of \( e_1 \)’s conversion rate, \( \hat{\gamma}^{P,e_1} = (Q^{e_1} + \alpha_0) / (\Pi^s(\hat{\gamma}^s) + \beta_0) \), is greater than that of \( e_2 \), which is just the common prior \( \bar{\gamma}_0 = \alpha_0 / \beta_0 \). Hence, the threshold for \( Q^{e_1} \) such that entrant \( e_1 \) earn the top organic slot in the next period is \( Q(\hat{\gamma}^s) = \Pi^s(\hat{\gamma}^s) \bar{\gamma}_0 \).

Intuitively, the higher is \( \hat{\gamma}^{e_1} \), entrant \( e_1 \) would be more optimistic about \( Q^{e_1} \), and believe there is a higher probability for \( Q^{e_1} \) to exceed \( Q(\hat{\gamma}^s) \). Thus the first part should be true. To be rigorous, \( W^s(\hat{\gamma}^{e_1}, \hat{\gamma}^s) \) is well-defined according to the information updating process. Since the exposure at the sponsored slot is \( s(\hat{\gamma}^s) \), the sales number follows a Poisson distribution

\[
Q^{e_1} \sim NB \left( (\beta_0 + \pi_0) \bar{\gamma}^{e_1}, \frac{\Pi^s(\hat{\gamma}^s)}{\Pi^s(\hat{\gamma}^s) + (\beta_0 + \pi_0)} \right)
\]

Note that this negative binomial distribution can be related to the number of successes until \( (\beta_0 + \pi_0) \bar{\gamma}^{e_1} \) failures happens in an infinite sequence of independent Bernoulli trials with success probability \( \frac{\Pi^s(\hat{\gamma}^s)}{\Pi^s(\hat{\gamma}^s) + (\beta_0 + \pi_0)} \). Hence, when \( (\beta_0 + \pi_0) \bar{\gamma}^{e_1} \) is higher, for any realization of the Bernoulli sequence, the realization of \( Q^{e_1} \) is higher, hence the probability for \( Q^{e_1} \) to exceed a given threshold is higher.

Van de Ven and Weber (1993) proves that the median of a negative binomial distribution \( NB (r, p) \) is bounded from below by \( \frac{[(r-1) p / (1-p)] + 1}{\Pi^s(\hat{\gamma}^s)} \). Hence, if \( \gamma^{e_1} = \bar{\gamma}_0 + \frac{1}{\beta_0 + \pi_0} \), the median of the distribution of \( Q^{e_1} \) is bounded from below by \( Q(\hat{\gamma}^s) \), which gives the second part of the lemma.

Furthermore, an exact expression of \( W^s(\hat{\gamma}^{e_1}, \hat{\gamma}^s) \) can be derived. Denote the CDF of \( Q^{e_1} \) as \( F_{Q^{e_1}} \), then

\[
W^s(\hat{\gamma}^{e_1}, \hat{\gamma}^s) = 1 - F_{Q^{e_1}} ((\Pi^s(\hat{\gamma}^s) + \pi_0) \bar{\gamma}_0 - q_0)
\]

Because the CDF of a negative binomial distribution \( NB (r, p) \) can be defined by the regularized incomplete beta function \( 1 - I_p (k + 1, r) \),

\[
W^s(\hat{\gamma}^{e_1}, \hat{\gamma}^s) = I \left( (\Pi^s(\hat{\gamma}^s) + \pi_0) \bar{\gamma}_0 - q_0 + 1, (\beta_0 + \pi_0) \bar{\gamma}^{e_1} \right)
\]

Proof of Proposition 1. I consider an entrant \( e_1 \) with posterior expectation \( \hat{\gamma}^{e_1} \), after
observing his private signal. Denote his equilibrium bid as \( b^{e1} = b^E (\tilde{\gamma}^{e1}) \). Denote \( U (b, \tilde{\gamma}^{e1}) \) as the entrant’s payoff for bidding \( b \), given the entrant’s expected quality \( \tilde{\gamma}^{e1} \) and other players’ playing equilibrium strategy. Because the equilibrium bid \( b^{e1} \) maximizes \( U (b, \tilde{\gamma}^{e1}) \), for any \( b' < b^{e1}, \) \( U (b^{e1}, \tilde{\gamma}^{e1}) - U (b', \tilde{\gamma}^{e1}) \geq 0 \). I argue that \( U (b^{e1}, \tilde{\gamma}^{e1}) - U (b', \tilde{\gamma}^{e1}) \) increases in \( \tilde{\gamma}^{e1} \), which would implies that for small \( d\gamma > 0 \), \( U (b^{e1}, \tilde{\gamma}^{e1} + d\gamma) - U (b', \tilde{\gamma}^{e1} + d\gamma) \) increases in \( \tilde{\gamma}^{e1} \), which would implies that for small \( d\gamma > 0 \), \( U (b^{e1}, \tilde{\gamma}^{e1}) - U (b', \tilde{\gamma}^{e1}) > 0 \), so \( b^E (\tilde{\gamma}^{e1} + d\gamma) \) is at least as high as \( b^E (\tilde{\gamma}^{e1}) \).

The two bids, \( b^{e1} \) and \( b' \), only generate a difference in payoff when \( b^{e1} \) wins the auction while \( b' \) not. Denote \( e_h \) as the entrant other than \( e_1 \) with the highest posterior expectation. It suffices to compare payoffs when entrant when \( b^E (\tilde{\gamma}^{eh}) \) lies right between \( b' \) and \( b^{e1} \).

\( U (b^{e1}, \tilde{\gamma}^{e1}) - U (b', \tilde{\gamma}^{e1}) \) contains three components. The first component is the difference in profit in the first period, which is \( \Pi^s \tilde{\gamma}^{e1} \). The second component, is related to the change in the profit in the next period. Now focus on this case. If entrant \( e_1 \) bids \( b^{e1} \), he wins the sponsored slot and the probability for him to get the top organic slot in the next period is \( W^s (\tilde{\gamma}^{e1}, \tilde{\gamma}^{s}) \). Because earning the top organic slot means his sales in this period is higher than the threshold, so his posterior expectation conditional on earning the top organic slot, denoted as \( \tilde{\gamma}^{e1,w} \), is higher than \( \tilde{\gamma}^{e1} \). If entrant \( e_1 \) bids \( b' \), he loses the sponsored slot to entrant \( e_h \), then entrant \( e_1 \) will earn the top organic slot in the next period with probability \( (1 - W^s (\tilde{\gamma}^{eh}, \tilde{\gamma}^{s})) / (N - 1) \). Hence, the difference in the total profit from both periods can be expressed as

\[
\int_{b' < b^E (\tilde{\gamma}^{eh}) < b^{e1}} \left[ \Pi^s \tilde{\gamma}^{e1} + W^s (\tilde{\gamma}^{e1}, \tilde{\gamma}^{s}) \Pi^o \tilde{\gamma}^{e1,w} - \frac{(1 - W^s (\tilde{\gamma}^{eh}, \tilde{\gamma}^{s})) \Pi^o \tilde{\gamma}^{e1}}{N - 1} \right] f (\tilde{\gamma}^{eh}) \, d\tilde{\gamma}^{eh} \\
= \int_{b' < b^E (\tilde{\gamma}^{eh}) < b^{e1}} \left[ W^s (\tilde{\gamma}^{e1}, \tilde{\gamma}^{s}) \Pi^o \tilde{\gamma}^{e1,w} + \left( \Pi^s - \frac{(1 - W^s (\tilde{\gamma}^{eh}, \tilde{\gamma}^{s})) \Pi^o}{N - 1} \right) \tilde{\gamma}^{e1} \right] f (\tilde{\gamma}^{eh}) \, d\tilde{\gamma}^{eh}
\]

Because \( W^s (\tilde{\gamma}^{e1}, \tilde{\gamma}^{s}) \) and \( \tilde{\gamma}^{e1,w} \) increases in \( \tilde{\gamma}^{e1} \), the first term increases in \( \tilde{\gamma}^{e1} \) rapidly. If \( N \) is greater than \( \Pi^o / \Pi^s + 1 \), the second term also increases in \( \tilde{\gamma}^{e1} \), and this difference in total profits would increase in \( \tilde{\gamma}^{e1} \).

The last element of \( U (b^{e1}, \tilde{\gamma}^{e1}) - U (b', \tilde{\gamma}^{e1}) \) is the difference in auction payment. Since this is a second price auction, this difference only has to do with bids and does not change with \( \tilde{\gamma}^{e1} \). Hence, overall, \( U (b^{e1}, \tilde{\gamma}^{e1}) - U (b', \tilde{\gamma}^{e1}) \) has all three components increase or hold constant in \( \tilde{\gamma}^{e1} \). ■