HKUST BUSINESS SCHOOL 香港科大商學院
WORLD CLASS IN ASIA
Center for Economic Policy 經濟政策研究中心

Search Algorithm and Sales on Online Platforms: Evidence from Food Delivery Platforms

Yangguang Huang

HKUST CEP Working Paper No. 2021-01

January 2021

HKUST Center for Economic Policy
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
T: 3469 3007 E:cep@ust.hk W: cep.hkust.edu.hk

Search Algorithm and Sales on Online Platforms: Evidence from Food Delivery Platforms

Yangguang Huang

HKUST CEP Working Paper No. 2021-01
January 2021

Abstract

One prominent feature of online sales is that buyers rely on the search tools offered by platforms to process information when searching for products. We develop a model that captures how the search algorithm affects buyers' search process, which influences the market equilibrium and welfare. The development of online platforms can reduce buyers' search costs and promote competition among sellers, but a platform may design a search algorithm that is too “selective” from the social welfare perspective, which causes consumers to consider fewer options and suppresses competition. By using data from food delivery platforms, we provide empirical evidence that search algorithms deeply affect restaurant revenues. Markets with more chain restaurants with established brands tend to have more concentrated sales. This is partly caused by search algorithms being biased towards large restaurant chains.

Author's contact information

Yangguang Huang

Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong
Email: huangyg@ust.hk

Search Algorithm and Sales on Online Platforms: Evidence from Food Delivery Platforms*

Yangguang Huang[†]

January 2021

Abstract

One prominent feature of online sales is that buyers rely on the search tools offered by platforms to process information when searching for products. We develop a model that captures how the search algorithm affects buyers' search process, which influences the market equilibrium and welfare. The development of online platforms can reduce buyers' search costs and promote competition among sellers, but a platform may design a search algorithm that is too "selective" from the social welfare perspective, which causes consumers to consider fewer options and suppresses competition. By using data from food delivery platforms, we provide empirical evidence that search algorithms deeply affect restaurant revenues. Markets with more chain restaurants with established brands tend to have more concentrated sales. This is partly caused by search algorithms being biased towards large restaurant chains.

Keywords: online platform, search algorithm, consideration set, food-delivery platform

JEL codes: D83, L11, L13, L42

*I appreciate all the valuable comments and suggestions by Mark Armstrong, John Asker, Chiara Farronato, Ginger Jin, Julian Wright, and Daniel Xu. I would like to thank seminar participants at HIT-SZ, HKUST IO workshop, SUFE IO conference, University of Tokyo, Hitostubashi University, BU platform symposium, EARIE conference, Asia-pacific IO conference, NBER summer institute, and Econometric Society World Congress. Yang Liu serves as an excellent research assistant for this project. An earlier version of the paper is circulated with the title "A Model of Sales on Online Platforms."

†Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. Email: huangyg@ust.hk.

1 Introduction

Online platforms have reshaped many industries and profoundly affected their market structures. In this paper, we construct a model that provides an anatomy of how online platforms influence the market structure and welfare through search algorithms. By using data from food delivery platforms, we show empirical evidence that search algorithms deeply affect restaurant revenues and sales distribution.

The development of online platforms plays a dual role in affecting market structures and welfare. The literature has shown that platforms are pro-competitive and welfare-improving in many aspects. For example, online platforms reduce search costs (Goldmanis et al., 2010) and enable buyers to better learn about products and to compare more options. Brynjolfsson et al. (2011) find that the internet channel exhibits a significantly less concentrated sales distribution than the traditional channel when the two channels have exactly the same catalog and prices. Moreover, most online platforms operate reputation systems that gather information from platform data and user-contributed content such as ratings and reviews. Well-functioning reputation systems can provide credible information about quality and other characteristics and thus alleviate asymmetric information (Akerlof, 1970; Klein et al., 2016), facilitate trust (Tadelis, 2016), and induce sellers to behave honestly and maintain good quality (Cabral and Hortacsu, 2010). By providing more quality information, online platforms increase search targetability (Yang, 2013) and allow buyers to make more informed decisions in seeking the best-suited products.

However, online platforms can also entail anti-competitive impacts especially when serving as information gatekeepers (Baye and Morgan, 2001). Platforms usually host a large number of sellers (firms) that offer heterogeneous products with a tremendous amount of information. It is often impossible for buyers (consumers) to thoroughly search and study all of the available products due to information overload (Anderson and De Palma, 2009). As a result, one of the most important features of online sales is that buyers rely on the search tools provided by the platforms to search for and learn about products.

Search tools such as search engines, recommender systems and price-comparison shopbots use information technology to assist buyers in searching for product information and in learning about product characteristics on the platforms. Because the search tools are designed and operated by for-profit online platforms, they might not be designed in a pro-competitive manner. Regulators and antitrust authorities have recognized the power that these platforms possess to influence the market via search tools. In June 2017, the European Commission (2017) fined Google €2.42 billion for "abusing dominance as search engine by giving illegal advantage to own comparison shopping service." Figure 1 illustrates Google's ability to control users' attention (clicks). The Federal Trade Commission (2013) requires search engines to "distinguish between advertisements and search results," because sponsored searches can profoundly influence buyers' decisions and cause potential efficiency losses (Ghose and Yang, 2009).

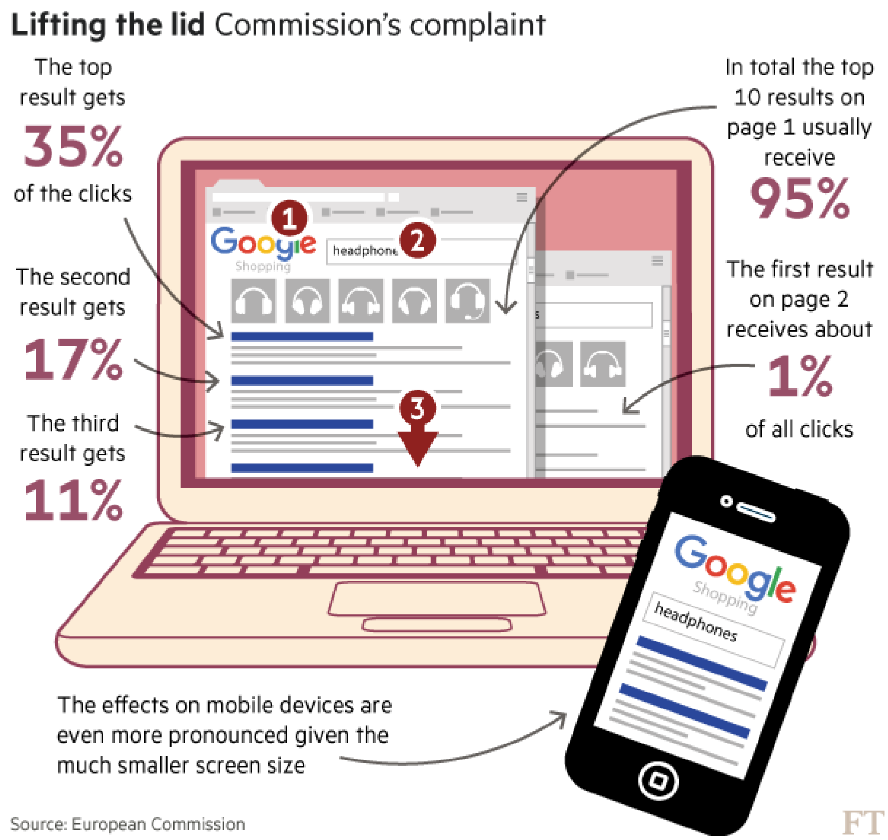


Figure 1: How Search Algorithms Affect Clicks

Online platforms have the private incentive to adopt an unequal or even highly selective algorithm that suppresses competition. Hagiu and Jullien (2011) note that platforms have two main motives to divert searches: reaping higher revenue from participating users and affecting sellers' choice in pricing or other strategic variables. A platform has a clear incentive to favor sellers who are vertically integrated with the platform (De Corniere and Taylor, 2019). In designing the search algorithm, platforms might bias the search results towards their own content or sponsored sellers (De Corniere and Taylor, 2014, 2019). Search advertising also deeply affects consumers' search behaviors and social welfare (Athey and Ellison, 2011; Chen and He, 2011; Eliaz and Spiegler, 2011b; Blake et al., 2015).[1] De Corniere (2016) shows that even with the neck-and-neck competition among search engines, sub-optimal sponsored links remain, and welfare can be worsened.

We construct a model that characterizes the dual role of online platforms. In the model, buyers search for products on the platform with the search tool offered by the platform. After the search process, each buyer obtains a consideration set that contains a number of options. The buyers then choose their favorite option within the consideration set (Eliaz and Spiegler, 2011a). The key component of the model is the formation of the consideration set under the influence of the search

[1] We do not explicitly distinguish between organic and sponsored search results in this paper. Buyers will naturally discount sponsored links and generally find sponsored content to be less relevant (Jansen and Resnick, 2006).

algorithm.[2] The search algorithm determines the ranking and composition of product information in the search results, which further determines the probability of each product being sampled. The algorithm usually does not treat sellers equally; therefore, products appear in the consideration set with different probabilities. Backus et al. (2014) shows that because of the eBay search algorithm, identical items can have vastly different visibility, which leads to dispersion in prices and the number of bidders. Chen and Tsai (2020) find empirical evidence that Amazon's own products are more likely to be recommended in the frequently-bought-together lists than the same products carried by third-party sellers.

In addition to the composition of product information, we emphasize that the search algorithm also affects the size of the consideration set since buyers spend only a limited amount of effort in searching. Dukes and Liu (2015) find that platforms can strategically increase search costs to discourage buyers from evaluating too many sellers. Although there might be many sellers in the market, the intensity of competition depends on the number of options in the consideration set. By using results from the theory of majorization, we show that the expected size of the consideration set increases as the algorithm becomes more "equal" in the sense of Lorenz ordering.

By affecting the composition and size of the consideration set, the search algorithm influences the sales distribution and welfare. We show that buyer-side surplus and total welfare improve if the platform adopts a more equal search algorithm. However, both the platform and sellers have the private incentive to make the search algorithm "selective," which is against the public interest.

Next, we use the data from food delivery platforms to provide some empirical evidence for the model. We show that a restaurant's revenue is critically determined by its default position in search results. The search algorithm tends to rank restaurants with high ratings and good reviews at the top. Ratings and reviews affect sales not only directly since they provide quality information (Cabral and Hortacsu, 2010) but also indirectly through the search algorithm.

Based on market-level regressions, we empirically demonstrate the dual role of online platforms. On the pro-competitive aspect, markets with more reviews tend to have a smoother sales concentration, which suggests that consumers explore more stores when they have more quality information. However, online platforms also impose strong anti-competitive impacts on the market. The search algorithm offers advantages to chain restaurants with established brands. As a result, markets with more large chain restaurants exhibit less competitive pressure, higher average prices, and a higher sales concentration.

2 Baseline Model

The model intends to capture how the platform affects market equilibrium through the search algorithm. There are I buyers (consumers, him) and J heterogeneous sellers (firms, her) that interact on a platform with buyer-side search frictions and seller-side market power.

[2] Our use of consideration set in this paper is similar to the concept of search pools in Eliaz and Spiegler (2016).

Sellers are indexed by $j = 1, 2, ..., J$. A generic seller draws an efficiency parameter θ as her private type independently from a distribution F with a compact support $[\underline{\theta}, \overline{\theta}]$ and continuous density f. Each seller offers a product as a price-quality combination $(p, q) \in \mathbb{R}_+^2$. For a seller with efficiency parameter θ, the cost of producing y units of products at quality q is $y \cdot c(q, \theta)$, where $c(\cdot, \cdot)$ is continuous and twice differentiable with $c_q > 0$, $c_{qq} > 0$, $c_\theta < 0$, and $c_{q\theta} < 0$.

For the I buyers in the market, each buyer purchases one unit of a product. Buyers have quasilinear preferences $u(p, q) = v(q) - p$, where $v(\cdot)$ is continuous and twice differentiable with $v_q > 0$ and $v_{qq} \leq 0$. Buyers face search frictions, and the platform can influence the search process through the search algorithm. After the search process, each buyer obtains a consideration set $\mathcal{C}$ that contains a number of products and chooses his favorite one within the consideration set. Therefore, the quantity of demand for seller j's product is $y_j = I \cdot \Pr(j \in \mathcal{C}) \cdot \Pr(j \text{ is chosen}|j \in \mathcal{C})$. The seller chooses the product (p, q) by maximizing her expected profit

$$\max_{p,q} [p - c(q, \theta_j)] \, y_j = [p - c(q, \theta_j)] \times I \times \Pr(j \in \mathcal{C}) \times \Pr(j \text{ is chosen}|j \in \mathcal{C}), \tag{1}$$

where the last two quantities are determined by the search process under the influence of the search algorithm.

In the baseline model, the search algorithm is fixed at the beginning. Sellers first draw their efficiency parameters as private information. Sellers choose p and q simultaneously. Then, buyers search for a product by using the search algorithm and make a purchasing decision in the way described as follows.

2.1 Search algorithm and search process

The output of a search algorithm is a ordered list of options. The ranking of the options determines the probability of each option being sampled by buyers. The click-through rate of a seller decreases substantially in its position in the search result (De los Santos and Koulayev, 2017; Ursu, 2018). This fact motivates us to model the search algorithm as a vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_J)$, where σ_j denotes the probability of seller j being sampled and $\sum_{j=1}^{J} \sigma_j = 1$. Buyers conduct fixed-sample-size searches as in Burdett and Judd (1983).[3] Each buyer draw K samples, where K is a positive integer that measures the search intensity.[4] By varying the search intensity K, the model nests two extreme cases of uninformed buyers ($K = 1$) and fully informed buyers ($K \to \infty$) in Varian (1980).

After taking K samples, a buyer obtains a search result $\boldsymbol{x} = (x_1, x_2, ..., x_J)$, where x_j is the number of samples taken from seller j. $\boldsymbol{x}$ follows a multinomial distribution with a probability mass

[3] The fixed-sample-size search can be optimal when searching involves a fixed cost so that the average search cost decreases in the number of searches (Hong and Shum, 2006; Morgan and Manning, 1985). De los Santos et al. (2012) provide empirical evidence of consumers adopting the fixed-sample-size search.

[4] The model can incorporate heterogeneous search costs by having buyers with different K's.

function

$$h(\boldsymbol{x}) = \frac{K!}{x_1!x_2!\cdots x_J!}\sigma_1^{x_1}\sigma_2^{x_2}\cdots\sigma_J^{x_J}, \quad \text{for } \sum_{j=1}^{J} x_j = K.$$

Seller j is included in the consideration set if at least one sample is taken from her, so that $\mathcal{C} \equiv \{j : x_j > 0, j = 1, 2, ..., J\}$. The probability of having product j in the consideration set is

$$\Pr(j \in \mathcal{C}) = 1 - \Pr(j \notin \mathcal{C}) = 1 - (1 - \sigma_j)^K. \tag{2}$$

$\Pr(j \in \mathcal{C})$ has the following properties:

Lemma 1. *Ceteris paribus, for all* $j = 1, 2, ..., J$,
(a) $\Pr(j \in \mathcal{C})$ *increases in* K *with* $\lim_{K\to\infty} \Pr(j \in \mathcal{C}) = 1$;
(b) $\Pr(j \in \mathcal{C})$ *increases in* σ_j *with* $\lim_{\sigma_j\to 0} \Pr(j \in \mathcal{C}) = 0$ *and* $\lim_{\sigma_j\to 1} \Pr(j \in \mathcal{C}) = 1$.

Next, let $N \equiv |\mathcal{C}|$ denote the cardinality of the consideration set. N is a random variable that can take the value of $n = 1, 2, ..., J$. The probability that the consideration set has exactly n elements is

$$P_n \equiv \Pr(N = n) = \sum_{\boldsymbol{x}\in\mathcal{X}(n)} h(\boldsymbol{x}),$$

where $\mathcal{X}(n) = \left\{\boldsymbol{x} : \sum_{j=1}^{J} x_j = K \text{ and } \sum_{j=1}^{J} \mathbb{I}(x_j > 0) = n\right\}$.[5]

Given a consideration set, a buyer chooses the product that yields the highest utility, $u_j = v(q_j) - p_j$. Provided that product j is in the consideration set, the probability that a buyer chooses j depends on the total number of options, that is,

$$\Pr\left(j \text{ is chosen}|j \in \mathcal{C}\right) = \sum_{n=1}^{J} \left\{P_n \times \Pr\left(j = \arg\max_{j'\in\mathcal{C}}\{u_{j'}\}\right)\right\}, \tag{3}$$

which is increasing in u_j.[6] From the expression of (3), we can observe that the stochastic order of N determines the intensity of competition. The search intensity K and search algorithm $\boldsymbol{\sigma}$ affect N in a systematic way as follows.

Lemma 2. *Let* N_1 *and* N_2 *denote the consideration set size under search intensity* K_1 *and* K_2, *respectively. If* $K_1 > K_2$, *then* N_1 *first-order stochastically dominates (FSD)* N_2, *which is denoted as* $N_1 \succcurlyeq_1 N_2$. *It follows that* $E[N_1] > E[N_2]$.

Therefore, as the search intensity increases, the consideration set size N increases in the sense of FSD; thus, the market becomes more competitive.

Next, we use the results from the theory of majorization (Marshall et al., 2011) to study the impact of the search algorithm $\boldsymbol{\sigma}$. Without a loss of generality, let $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_J)$ be indexed such

[5] $\mathbb{I}(\cdot)$ is the indicator function throughout this paper.

[6] The last equality is based on the assumption that the search process is not affected by the utilities of products. It is possible to relax this assumption as long as $\Pr\left(j \text{ is chosen}|j \in \mathcal{C}(K)\right)$ is increasing in u_j.

that $\sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_J$. The concept of majorization is developed to formally define a vector $\boldsymbol{\sigma}_1 = (\sigma_{11}, \sigma_{12}, ..., \sigma_{1J})$ that is "less spread out" or "more equal" than $\boldsymbol{\sigma}_2 = (\sigma_{21}, \sigma_{22}, ..., \sigma_{2J})$.

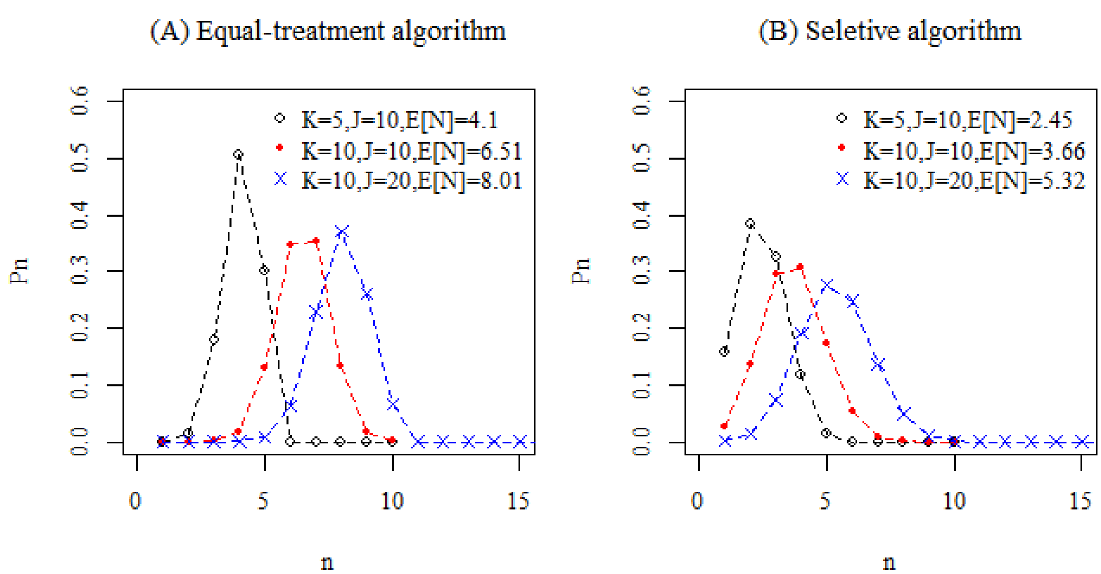
Definition. For $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \Delta^{J-1} \subset \mathbb{R}^J$, if $\sum_{j=1}^{k} \sigma_{1j} \geq \sum_{j=1}^{k} \sigma_{2j}$ for $k = 1, 2, ..., J-1$, we say that $\boldsymbol{\sigma}_1$ *is majorized by* $\boldsymbol{\sigma}_2$, which is denoted as $\boldsymbol{\sigma}_1 \prec \boldsymbol{\sigma}_2$.[7]

Intuitively, having $\boldsymbol{\sigma}_1 \prec \boldsymbol{\sigma}_2$ is equivalent to the Lorenz curve of $\boldsymbol{\sigma}_1$ being above the Lorenz curve of $\boldsymbol{\sigma}_2$. Therefore, majorization is also called *Lorenz ordering*. The theory of majorization helps us to establish the relationship between the search algorithm and the size of the consideration set.

Lemma 3. *Let N_1 and N_2 denote the consideration set sizes under search algorithm $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, respectively. If $\boldsymbol{\sigma}_1 \prec \boldsymbol{\sigma}_2$, then $N_1 \succcurlyeq_1 N_2$.*

That is, when the search algorithm $\boldsymbol{\sigma}_1$ is majorized by $\boldsymbol{\sigma}_2$, then the size of the consideration set decreases in the sense of FSD. Note that, the equal-treatment algorithm,[8] $\boldsymbol{\sigma} = (\frac{1}{J}, \frac{1}{J}, ..., \frac{1}{J})$, is majorized by all algorithms, i.e., $(\frac{1}{J}, ..., \frac{1}{J}) \prec (\sigma_1, ..., \sigma_J) \prec (0, ..., 0, 1)$. Therefore, we have the following result.

Lemma 4. *The equal-treatment algorithm maximizes N in the sense of FSD.*



Note: We generate Figures 2, 3, and 4 by letting $\theta \sim U[0,1]$, $c(q,\theta) = q^2/(2\theta)$, $v(q) = q$. For the selective algorithm, we set $\alpha = 10$.

Figure 2: Illustration of the Random Consideration Set Size

[7]The general definition of majorization does not require the vectors belong to the $J-1$ simplex.

[8]The name "equal-treatment" is taken from European Commission (2017) as it charges Google to "comply with the simple principle of giving equal treatment."

Lemmas 3 and 4 imply that if the search algorithm emphasizes some option(s), then it causes the buyers to obtain a smaller consideration set. Figure 2 illustrates Lemmas 2-4. Compare the equal-treatment algorithm with a *selective algorithm*, $\boldsymbol{\sigma} = \left(\frac{1}{\alpha-1+J}, \frac{1}{\alpha-1+J} \cdots, \frac{\alpha}{\alpha-1+J}\right)$, which makes seller J α-times more likely to be sampled by buyers.[9] Under the selective algorithm, the consideration set size is smaller than the counterparts under the equal-treatment algorithm.

The search intensity (K), the number of products (J), and the search algorithm ($\boldsymbol{\sigma}$) jointly determine the probability that product j is in the consideration set ($\Pr(j \in \mathcal{C})$) and the distribution of the consideration set size (P_n). When a buyer searches for product information, it is possible to sample the same seller multiple times. In practice, if advertising is persuasive (Bloch and Manceau, 1999), sellers may intentionally use sponsored links to reach the buyer multiple times. Even if advertising is purely informative (Grossman and Shapiro, 1984), sellers may want to flood the market with sponsored advertisements so that consumers cannot easily find information about other sellers. As a result, after spending a limited amount of effort in searching, the buyer might not end up with many options.

2.2 Market equilibrium and welfare

Given the search algorithm, a generic seller j of type θ chooses the price and quality through the maximization problem (1). Because I and $\Pr(j \in \mathcal{C})$ are not affected by the seller's choice of price and quality, the maximization problem above is equivalent to

$$\max_{p,q} \left[p - c(q, \theta)\right] \Pr\left(j = \arg\max\nolimits_{j' \in \mathcal{C}} \{u_{j'}\}\right). \tag{4}$$

The surplus of a product with quality q is the difference between its value to the buyer and the cost of the product, i.e., $v(q) - c(q, \theta)$. Given that v is concave and c is strictly convex in q, there is a unique solution of $\max_q \{v(q) - c(q, \theta)\}$ characterized by the first-order condition $v_q(q^*) - c_q(q^*, \theta) = 0$. Define this solution as

$$q^*(\theta) = \arg\max_q \{v(q) - c(q, \theta)\}. \tag{5}$$

For a seller with type θ, by producing at quality $q^*(\theta)$, the product generates the largest social surplus, $w^*(\theta) = v(q^*(\theta)) - c(q^*(\theta), \theta)$. Both $q^*(\theta)$ and $w^*(\theta)$ are increasing in the efficiency parameter θ.[10] $w^*(\cdot)$ is a one-to-one mapping from type θ to the surplus measure w. Given the distribution function F, there is a unique distribution function G for w. G has a compact support $[\underline{w}, \overline{w}]$, where $\underline{w} = w^*(\underline{\theta})$ and $\overline{w} = w^*(\overline{\theta})$.

We can then rewrite the maximization problem (4) in a way that the seller first chooses a utility level u of her product based on the social surplus w and seeks a price-quality combination to fulfill

[9] We can generalize this selective algorithm to let it emphasize L favored sellers: $\boldsymbol{\sigma} = (\frac{1}{(\alpha-1)L+J}, \ldots, \frac{1}{(\alpha-1)L+J}, \frac{\alpha}{(\alpha-1)L+J} \cdots, \frac{\alpha}{(\alpha-1)L+J})$. When fixing L, increasing α increases the selectiveness or inequality of the algorithm. Formally, the degree of inequality can be measured by the order of majorization. See the supplemental materials.

[10] Because $c_{q\theta} < 0$, according to the implicit function theorem, $\frac{dq^*}{d\theta} = -\frac{-c_{q\theta}}{v_{qq} - C_{qq}} = \frac{c_{q\theta}}{v_{qq} - c_{qq}} > 0$.

this utility level.

$$
\begin{aligned}
(4) \Leftrightarrow & \max_u \left\{ \max_{(p,q) \text{ s.t. } v(q)-p=u} [p - c(q,\theta)] \Pr\left(u \geq \max_{j' \in \mathcal{C}}\{u_{j'}\}\right) \right\} \\
\Leftrightarrow & \max_u \left\{ \max_q [v(q) - c(q,\theta) - u] \Pr\left(u \geq \max_{j' \in \mathcal{C}}\{u_{j'}\}\right) \right\} \\
\Leftrightarrow & \max_u \left\{ [v(q^*(\theta)) - c(q^*(\theta),\theta) - u] \Pr\left(u \geq \max_{j' \in \mathcal{C}}\{u_{j'}\}\right) \right\} \\
\Leftrightarrow & \max_u \left\{ [w^*(\theta) - u] \Pr\left(u \geq \max_{j' \in \mathcal{C}}\{u_{j'}\}\right) \right\}. \qquad (6)
\end{aligned}
$$

Ultimately, the utility of the product can determine whether j is chosen among the available options. Therefore, the seller will choose the efficient quality $q^*(\theta)$ in equilibrium because it maximizes the first term, $[v(q) - c(q,\theta) - u]$, without affecting the second term, $\Pr\left(u \geq \max_{j' \in \mathcal{C}}\{u_{j'}\}\right)$.

Given the social surplus w, the firm will choose the utility level of the product that solves (6). We focus on the symmetric and monotone Bayesian Nash equilibrium (BNE), $u = \mu(w)$. Let $W_{(1:N)}$ denote the greatest order statistic among N independent draws from G. $G_{(1:N)}(w) = \Pr(W_{(1:N)} \leq w)$ is the distribution function of $W_{(1:N)}$, where $G_{(1:N)}(w) = [G(w)]^N$. Given that $\mu(\cdot)$ increases in w, the *choice probability* of a seller with social surplus w is

$$
\mathcal{P}(w) = \sum_{n=1}^{J} \left\{ P_n \times G_{(1:n-1)}(w) \right\} = E_N \left[G_{(1:N-1)}(w) \right],
$$

which indicates the probability that a product has the highest social surplus among N options in the consideration set. The solution to (4) is characterized in the following theorem.

Theorem 1. *A seller with efficiency parameter θ chooses the product with quality $q^*(\theta)$ given in (5) and $p^*(\theta) = v(q^*(\theta)) - \mu(w^*(\theta))$, where the utility of the product is*

$$
\mu(w^*(\theta)) = w^*(\theta) - \frac{\int_{\underline{w}}^{w^*(\theta)} \mathcal{P}(\omega)\, d\omega}{\mathcal{P}(w^*(\theta))}.
$$

Note that $\mu(w^*(\theta)) = v(q^*(\theta)) - p^*(\theta)$ is the utility or buyer's surplus of the product offered by the seller with type θ in equilibrium. The equilibrium price divides the social surplus of a transaction into the seller's and the buyer's surplus:

$$
\underbrace{w^*(\theta)}_{\text{social surplus}} = \underbrace{[p^*(\theta) - c(q^*(\theta),\theta)]}_{\text{seller's profit}} + \underbrace{[v(q^*(\theta)) - p^*(\theta)]}_{\text{buyer's surplus}}.
$$

Let $\mu(w; N)$ and $p^*(\theta; N)$ denote the equilibrium utility and price, respectively, when the random consideration set size is N.

Corollary 1. *Let N_1 and N_2 be two consideration set sizes, and $N_1 \succcurlyeq_1 N_2$. Then, $\mu(w; N_1) \geq$*

$\mu(w; N_2)$ *and* $p^*(\theta; N_1) \leq p^*(\theta; N_2)$.

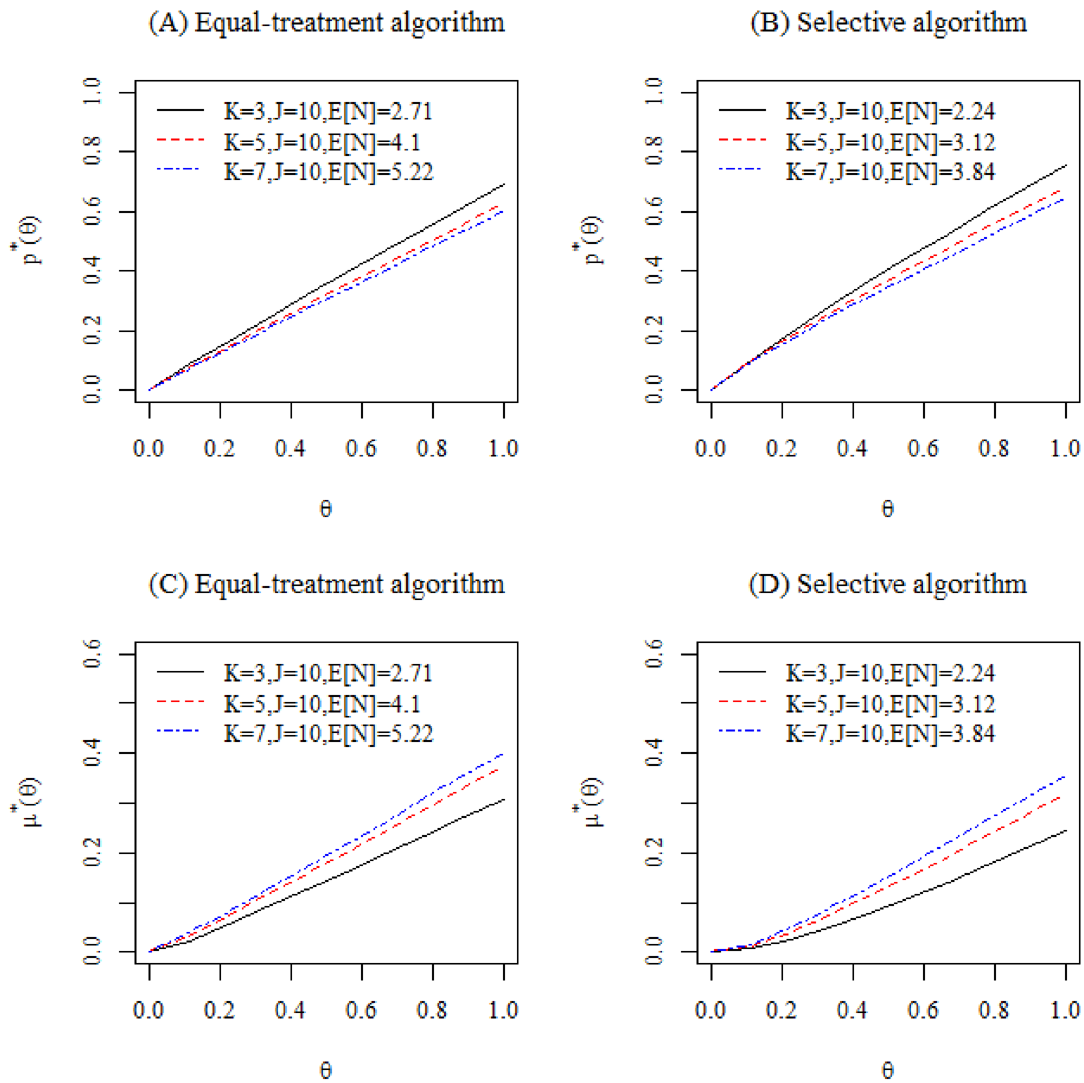


Figure 3: Illustration of the Equilibrium

Figure 3 illustrates Corollary 1, which indicates that if the expected level of competition (the number of rival sellers) increases, then a seller offers a more attractive product by setting lower prices. Therefore, the consideration set size critically determines how the seller and the buyer divide the social surplus. A larger consideration set (in the sense of FSD) implies that the seller gets less and the buyer gets more.

In equilibrium, the quantity demanded of the seller with type θ is

$$y^*(\theta) = I \times \Pr(j \in \mathcal{C}) \times \mathcal{P}\left(w^*(\theta)\right).$$

The first probability, $\Pr\left(j \in \mathcal{C}\right)$, is determined by the search algorithm ($\boldsymbol{\sigma}$) and the search intensity

(K). Seller j cannot affect this probability through her choice of price and quality.[11] Obviously, if $\Pr(j \in \mathcal{C})$ is large, seller j's sales and revenue will increase. In Section 4, we empirically show that a seller with a higher position earns a higher revenue. The second probability, $\mathcal{P}\left(w^*(\theta)\right)$, is the choice probability determined by the efficiency parameter of the seller (θ) and the consideration set size (N).

Given n options in the consideration set, a buyer purchases the product with the highest utility, which is offered by the seller with the highest social surplus among the n sellers. Considering the randomness of N, we can construct *ex ante* welfare measures. The total social welfare is

$$SW = I \cdot \sum_{n=1}^{J} P_n \left[\int_{\underline{w}}^{\overline{w}} w dG_{(1:N)}(w)\right] = I \cdot E_N \left[E\left[W_{(1:N)}\right]\right]. \tag{7}$$

Social welfare can be decomposed into the buyer-side surplus[12]

$$U = I \cdot \sum_{n=1}^{J} P_n \left[\int_{\underline{w}}^{\overline{w}} \mu(w) dG_{(1:N)}(w)\right] = I \cdot E_N \left[\mu(W_{(1:N)})\right] \tag{8}$$

and the seller-side profit

$$\mathit{\Pi} = I \cdot \sum_{n=1}^{J} P_n \left[\int_{\underline{w}}^{\overline{w}} \left[w - \mu(w)\right] dG_{(1:n)}(w)\right] = I \cdot E_N \left[E\left[W_{(1:N)} - \mu(W_{(1:N)})\right]\right]. \tag{9}$$

These three welfare measures critically depend on the distribution of N:

Theorem 2. *If N increases in the sense of FSD, then SW increases, U increases, and $\mathit{\Pi}$ decreases.*

Together with Lemmas 2, 3, and 4, we have the following corollaries.

Corollary 2. *When K increases, SW increases, U increases, and $\mathit{\Pi}$ decreases.*

Corollary 3. *For two search algorithms with $\boldsymbol{\sigma}_1 \prec \boldsymbol{\sigma}_2$, $SW(\boldsymbol{\sigma}_1) > SW(\boldsymbol{\sigma}_2)$, $U(\boldsymbol{\sigma}_1) > U(\boldsymbol{\sigma}_2)$, and $\mathit{\Pi}(\boldsymbol{\sigma}_1) < \mathit{\Pi}(\boldsymbol{\sigma}_2)$.*

Corollary 4. *The equal-treatment algorithm maximizes SW and U.*

Figure 4 illustrates the main results above and demonstrates the dual role of the platform. The platform can reduce search costs and facilitates a greater search intensity (K),[13] so each seller is

[11] If there are no search frictions, the consideration set includes products from all sellers. Without the random utility term in discrete choice models, the most efficient seller will offer the best product (in terms of utility) and capture the entire market. The source of market power is rooted in buyers' inability to search all of the products. Therefore, the most efficient seller cannot sell to all buyers due to search frictions.

[12] By assuming a fixed-sample-size search, we do not consider the saving of search costs from using the search tools in the welfare computation. With buyers searching sequentially and strategically, the search design will affect the search costs and buyer welfare in a complicated and ambiguous way. See Chen and Zhang (2017, 2018).

[13] The reduction of search costs might be due to convenient search tools, recommender systems, friendly user interfaces, and more quality information.

more likely to appear in the consideration set (Lemma 1), and the consideration set size expands (Lemma 2). A larger consideration set implies more intense competition, lower prices, and buyers purchasing better products (Corollary 1). The overall efficiency of the market and buyer-side welfare both improve because the market becomes more competitive (Theorem 2 and Corollary 2). This is the pro-competitive aspect.

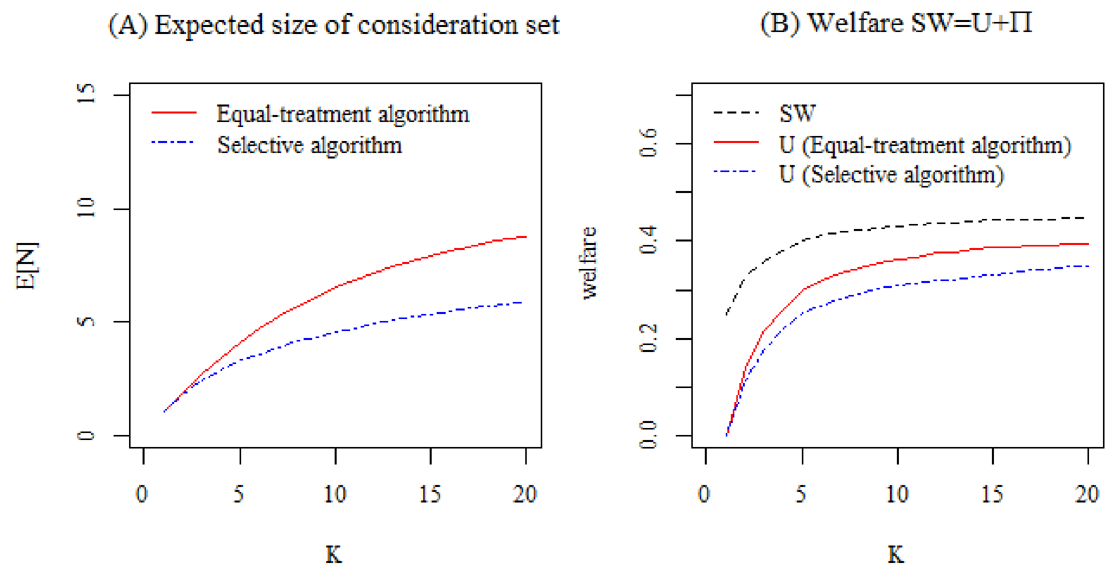


Figure 4: Welfare Implications of the Search Algorithm

However, the platform may adopt a highly selective search algorithm. As the algorithm becomes "less equal," the consideration set size shrinks. It softens the competition among sellers (Lemma 3) and raises prices (Corollary 1). As a result, the platform, by adopting a selective algorithm, can harm social welfare and the buyer-side surplus (Corollary 3). De Corniere and Taylor (2019) reaches a similar result that a biased intermediary harms consumers because the favored firm offers a product with lower utility.

Because the equal-treatment algorithm maximizes the consideration set, it maximizes social welfare at all levels of search intensity (Corollary 4). Allowing the search algorithm to emphasize some option(s) causes the buyers to obtain a smaller consideration set and softens the competition among sellers. These results support the ruling by the European Commission (2017). Moreover, the requirement by the Federal Trade Commission (2013) on distinguishing sponsored and organic search results can help buyers to search more effectively. As the search intensity increases, social welfare and the buyer-side surplus will improve.

3 Extension and Discussion

Our baseline model can be extended in various directions, and we discuss two of them.

3.1 Informative search algorithm

The results above are obtained under the setting that the platform commits to a search algorithm at the beginning of the game. Because all sellers are *ex ante* symmetric, the search algorithm is *uninformative* because it does not reflect sellers' types or other characteristics (such as price and quality). The uninformative setting establishes a benchmark on how search algorithms affect market equilibrium and welfare.

In reality, the platform usually has the information to distinguish sellers and the incentive to strategically design the search algorithm based on seller attributes. Online platforms usually possess data about the sellers and use the data as inputs into their search algorithms.[14] Therefore, the search algorithm can be *informative* in the sense that it promotes sellers with certain characteristics. The promotion can be personalized for different buyers based on buyer-side data such as queries and purchase histories.

If the platform has information about product quality, it can allow the search algorithm to rank an efficient seller up front.[15] However, a self-interest platform might not have the incentive to do this. It is possible that highly efficient sellers are disfavored by the algorithm because a low-quality seller pays the platform for search advertisements.[16] In this case, the platform is likely to cause welfare loss because high-quality sellers appear less frequently in the consideration sets, and low-quality sellers that are favored obtain greater market power. Moreover, even if the highest-quality seller is ranked in first place in the search outcome, the platform still has an incentive to reduce competition by suppressing other sellers. Under a proportional fee scheme, by giving the promoted sellers more market power, the platform collects more fees from higher prices.

Consider a simple scenario that the search algorithm depends on the realized efficiency parameters. Because sellers still solve the profit maximization problem (4), the equilibrium in Theorem 1 still holds.[17] Let the most efficient seller be promoted by the selective algorithm with $\alpha \geq 1$. Figure 5 demonstrates the welfare implications. As shown by the red-dashed curve, social welfare increases in α. However, if the least efficient seller is promoted, social welfare decreases in α, as illustrated by the blue-dashed curve. Note that if the selective algorithm is uninformative, then the social welfare still decreases in α (Corollary 3).

[14] The data can come from various sources including self-reported information from sellers, past sales histories, ratings and reviews from consumers, and data brokers.

[15] If products are horizontally differentiated, then the search algorithm can promote products that fit the tastes of different buyers. Queries and buyer-side data can be used to infer the preferences of buyers.

[16] For example, in 2016, a Chinese college student died after receiving an experimental treatment for synovial sarcoma promoted by Baidu. Because of this case, the Cyberspace Administration of China imposed new restrictions on search advertisements. See `en.wikipedia.org/wiki/Death_of_Wei_Zexi`.

[17] Note that, if the search algorithm is not based on θ but on q or p, then the profit maximization problem (1) is no longer equivalent to (4). The equilibrium choice of p and q in Theorem 1 will not hold because sellers will strategically choose prices and quantity levels in response to the design of the search algorithm.

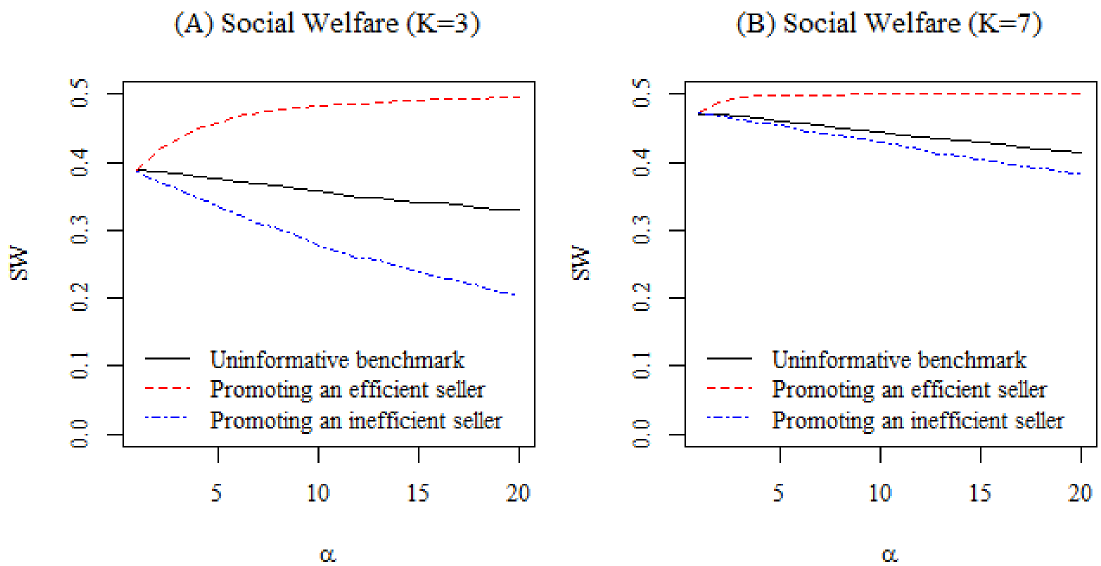


Figure 5: Informative Algorithm and Welfare

Sellers can also influence the search algorithm through search advertisements or other approaches. Showing sponsored searches in the results of search algorithms can substantially affect the search processes of buyers (Ricci et al., 2011). By paying for sponsored searches, advertising sellers not only increase their chances of being in the consideration sets but also indirectly suppress competition since other products are being considered less often. The platform also has a strong incentive to promote sponsored results and render the search algorithm highly selective. By filling buyer consideration sets with few sponsored options, the platform obtains not only more advertisement revenue but also higher commissions from higher prices under the popular proportional fee scheme. In addition, the brand equity of a seller can greatly influence her position in the organic search results.[18] Because buyers are less likely to purchase niche products, the platform might be inclined to promote already popular mass-market products (Bar-Isaac et al., 2012). In Section 4, we find that food delivery platforms indeed promote chain restaurants with established brands.

Moreover, the platform has a natural incentive to over-emphasize sellers that are vertically integrated with it or paying higher commissions Inderst and Ottaviani (2012); Teh and Wright (2020). For example, the European Commission (2017) find that "Google systematically gave prominent placement to its own in-house service and demoted rival comparison shopping services in search results, so even the most highly ranked rival service appears on average only on **page four** of Google's search results."

In summary, although the search algorithm can provide buyer information about sellers, it nevertheless can be strongly anti-competitive if the search algorithm is selective and biased. The overall welfare implication of online platforms is ambiguous.

[18]In the marketing literature, brand equity is a major factor that affects choice probability in addition to product attributes (Srinivasan et al., 2005).

3.2 Endogenous participation and the platform's choice

Essentially, the platform has the power to manipulate the entire consideration set by selecting which items to show to each consumer. By adjusting the algorithm, it can ensure that one option always appears at the top of the search results or cause another option to never appear. To some extent, the platform will design the algorithm to keep with social welfare because it must attract users to participate. The endogenous participation and network effect restrict the extent of the platform to use the search algorithm for its private interest.

Consider the endogenous participation of buyers and sellers in the two-sided market hosted by the platform. The indirect network externalities arise naturally from the model because seller-side profit and buyer-side surplus are both increasing functions of the other side's participation. We can extend the model to incorporate endogenous participation and the platform's optimization problem based on the following timeline:

(i) The platform chooses the fee and establishes the search algorithm.

(ii) Buyers and sellers decide whether to participate in the platform.

(iii) Sellers draw their efficiency parameters and choose products as price-quality combinations.

(iv) Buyers search via the search algorithm, form consideration sets, and buy products.

Under this setting, the model in Section 2 is a subgame at stage (iii) with I buyers and J sellers that participate in the market. At stage (ii), all sellers are symmetric, and the expected profit of a seller is $\pi(I, J, \boldsymbol{\sigma}) = \frac{1}{J}\mathit{\Pi} = \frac{1}{J} \cdot I \cdot E_N\left[E\left[W_{(1:N)} - \mu(W_{(1:N)})\right]\right]$, which increases in I and decreases in J. Suppose that the platform charges a uniform membership fee $\tau \geq 0$ for each seller as in Armstrong (2006).[19] Seller participation is determined by the zero profit condition

$$\pi(I, J, \boldsymbol{\sigma}) = \tau \Leftrightarrow J = \mathcal{J}(I, \tau, \boldsymbol{\sigma}), \tag{10}$$

where $\mathcal{J}$ is an increasing function of I and a decreasing function of τ.

Suppose that buyers have heterogeneous reservation utility r, which represents the payoff of not participating in the platform or purchasing from another competing platform. At the beginning of the participation stage, each buyer draws an r from a commonly known distribution $\mathit{\Gamma}(\cdot)$. $\mathit{\Gamma}$ has continuous support and is strictly increasing. From (8), the expected payoff of a generic buyer i is $u(J, \tau, \boldsymbol{\sigma}) = E_N\left[\mu(W_{(1:N)})\right]$,which increases in J. A buyer participates if $u(J, \tau, \boldsymbol{\sigma}) \geq r$. Given $\mathbb{I}$ potential buyers, the number of buyers that will participate in the platform is

$$I = \mathbb{I} \times \Pr\left(u(J, \tau, \boldsymbol{\sigma}) \geq r\right) = \mathbb{I} \times \mathit{\Gamma}\left(u(J, \boldsymbol{\sigma}, \tau)\right) = \mathcal{I}(J, \boldsymbol{\sigma}, \tau), \tag{11}$$

where $\mathcal{I}$ is an increasing function of J and a decreasing function of τ.

Equations (10) and (11) constitute a typical two-sided market with indirect network effects as in Rochet and Tirole (2006). Given a membership fee τ and a search algorithm $\boldsymbol{\sigma}$, the equilibrium

[19] If the platform charges a proportional or usage fee as in Rochet and Tirole (2003), the fee will distort the quality provision, but the result will be qualitatively similar.

numbers of sellers and buyers are determined by

$$\begin{cases} I = \mathcal{I}(J, \tau, \boldsymbol{\sigma}) \\ J = \mathcal{J}(I, \tau, \boldsymbol{\sigma}) \end{cases} \Rightarrow \begin{cases} J^*(\tau, \boldsymbol{\sigma}) \\ I^*(\tau, \boldsymbol{\sigma}). \end{cases}$$

By extending the model in this way, we can compare the platform's profit-maximizing result and the socially optimal result. We can show that the membership fee chosen based on the platform's private interest is too high from a social welfare perspective (Theorem 3 in the Appendix).[20] The discrepancy between the platform's interest and social welfare also lies in the choice of the search algorithm. Because most online platforms mainly collect fees from the seller side, they will bias the search algorithm towards the seller side. Under the proportional fee scheme, platforms aim to maximize the transaction volume, not the social welfare or consumer surplus.

In light of the model, the general goal of regulating search algorithms is to induce the algorithm to not only present the most relevant options on top but also offer a large number and variety of alternative options. Having alternative options available is an important competitive force in protecting buyers' interests. Regulators can require a search algorithm to present the alternative options immediately below sponsored links. Regulators can also prohibit a search algorithm from presenting the same product or seller multiple times in the search results. The modeling framework developed in this paper could potentially be used to evaluate the effect of regulatory proposals.

4 Empirics

We use the data from Chinese food delivery platforms to explore how online platforms affect sales, especially through their search algorithms.

4.1 Industry background

Online food delivery platforms allow buyers to search for food from nearby restaurants and to place orders through mobile applications (apps) or websites. Much like other e-commerce platforms, as buyers open the app, a list of restaurants appears. Buyers can scroll down for more options and refine the search results by queries or filters based on the food category, delivery time, style, and price. By clicking a restaurant's picture on the list, the buyer goes to the restaurant's page. The page contains a food menu with images, prices, estimated delivery times, ratings, reviews, and other textual descriptions. Buyers then decide what to order and pay electronically.

The catering industry in China has experienced drastic changes since food delivery platforms became popular in 2013. The market size of takeout food in China grew from less than US$10 billion in 2013 to more than US$37 billion in 2017.[21] In 2018, there were more than 256 million

[20]There is also a misalignment of private and public interest in the choice of contract forms. Examples include exclusive dealings (Armstrong and Wright, 2007) and price-parity clauses (Wang and Wright, 2020).

[21]Sources: `www.iresearchchina.com`. In the United States, the market size of takeout food is approximately US$10

active users of online food delivery services in China, which covers more than 1,300 cities. There are three major food delivery platforms, namely, Meituan, Ele.me, and Baidu.[22] Meituan and Ele.me together account for more than 80% of all takeout food transactions.[23] The three platforms work with restaurants in a similar revenue-sharing fee scheme. The platform usually collects 5% to 20% of the revenue of restaurants. Some restaurants join the platform and use the delivery service offered by the platform, while other restaurants continue to make their own deliveries. The restaurants that operate their delivery independently pay a lower share of revenue than those using the platforms' delivery services. Many restaurants choose to operate on multiple platforms (multi-homing), while most buyers use only one platform (single-homing).[24]

With food delivery platforms, small independent restaurants used to serving only nearby communities are empowered by inexpensive cheap delivery services, advertising channels, and ways to build their reputation. Large restaurant chains also benefit from convenient ordering and payment systems. Restaurants have access to more buyers but face more intense competition. Buyers incur lower search costs and can utilize the quality information on the platforms to find the restaurants best suited to their tastes. However, the search process is now guided by the search algorithms provided by the platforms. On the one hand, small restaurants have the opportunity to develop their reputations and attract buyers who search carefully. Given sufficient taste heterogeneity, reducing the search costs empowers small and high-quality sellers, which leads to less concentrated sales. On the other hand, large sellers with established brands and chain stores could gain a greater advantage due to their dominant positions on online platforms.[25] The overall effect of online platforms on sales concentration is ambiguous.

Food delivery platforms serve as an excellent data source for studying online platforms' influence on sales distribution. The industry is naturally divided into many relatively separate markets due to geographic restrictions on food delivery services.[26] Therefore, there are many market-level observations with variations in their seller composition and the amount of quality information. More importantly, buyers' search processes are heavily influenced by the restaurant list provided by the food delivery platform. Once buyers turn on the mobile app, the search algorithm will

billion. The largest platform, GrubHub/Seamless, has a similar market share to Domino's Pizza. Hence, food delivery platforms in the US are much less influential than in China.

[22]Meituan (`waimai.meituan.com`) is backed by Tencent. Alibaba is the major investor in Ele.me (`www.ele.me`). The food delivery service offered by Baidu (`waimai.baidu.com`) is facilitated by the major internet search engine in China. See `www.scmp.com/business/companies/article/2111163/dinner-your-door-inside-chinas-us37-billion-online-food-delivery`. In August 2017, Ele.me acquired the food delivery business of Baidu. In October 2018, the Baidu platform was rebranded as `star.ele.me`.

[23]Sources: `www.itrustdata.cn`.

[24]Almost 60% of the restaurants on Ele.me also operate on Meituan, but only 7.6% of buyers actively use both Ele.me and Meituan. This two-sided market structure fits the case of the competitive bottleneck in Armstrong (2006), in which platforms compete for a larger installed base of single-homing buyers, and grant them market power in addressing multi-homing sellers.

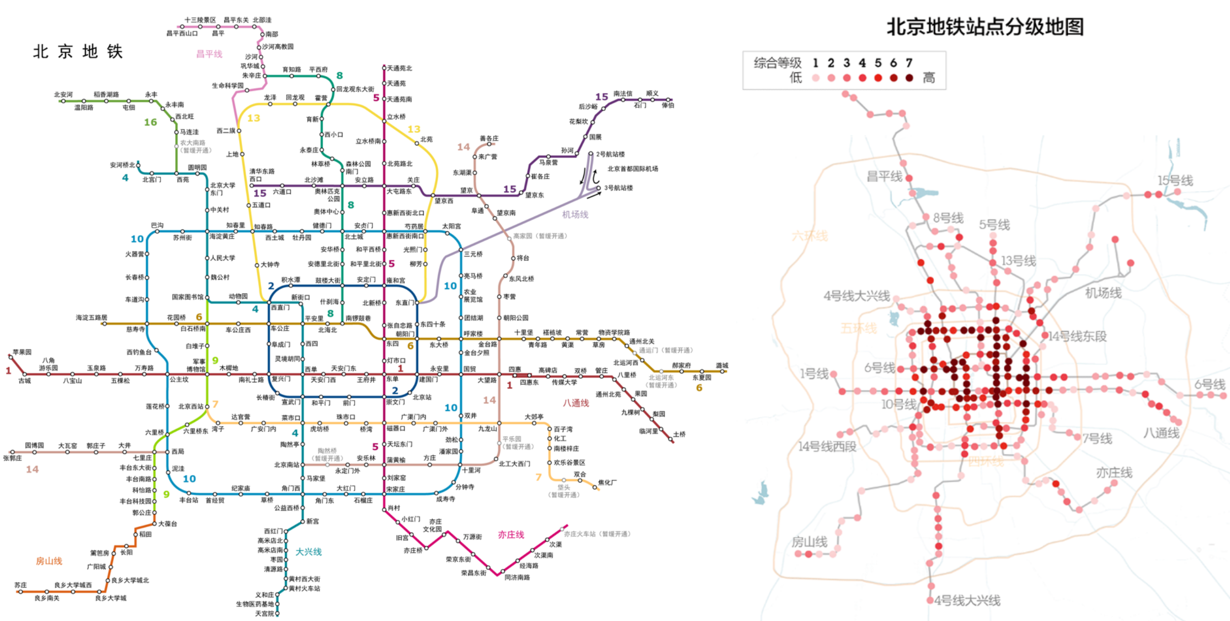
[25]In 2018, the top five sellers on Ele.me were Kentucky Fried Chicken, McDonald's, Pizza Hut, Burger King, and Yoshinoya (Sources: `www.cbndata.com`).

[26]Food delivery services are restricted by the physical locations of restaurants because the food must be in good condition after delivery. The retention of heat critically depends on the delivery distance.

provide a list of nearby restaurants without taking any query. This restaurant list is mainly based on buyer locations.[27]

4.2 Data

The data cover restaurants that operate on three food delivery platforms in Beijing in August 2018. Beijing has a large and highly active takeout food market with more than 100 thousand active restaurants and more than 10 million active users on these platforms. The data consist of all restaurants that provide services in the vicinity of all 296 subway stations in Beijing. Figure 6 shows a map of Beijing's subway stations. Table 1 summarizes the platform-level aggregate information. The largest platform, Meituan, has a monthly revenue more than CNY 700 million. For the rest of this paper, all monetary units are in CNY. Meituan hosts a relatively small number of brands, but its market share is the largest (66.96%). It also has the largest number of ratings and reviews, which is approximately 10 times that of Ele.me. On average, the markets on Meituan have the least concentrated sales, while the markets on Baidu are the most concentrated.



Note: The right panel shows the economic activeness of subway stations. Sources: DTCJ (`www.dtcj.com`).

Figure 6: Beijing Subway Stations and Their Activeness Levels

[27]It can also depend on other data including purchase histories and other personal characteristics. However, for takeout food, location is the most important factor.

Table 1: Summary Information of the Three Platforms

	Baidu	Ele.me	Meituan	Total
No. of stores	25,321	36,735	39,981	102,037
No. of brands	14,071	20,820	13,605	34,823
Average store revenue	3,776	6,840	17,582	
Platform aggregate revenue	95,599,690	251,265,384	702,961,273	
Platform revenue share	9.11%	23.93%	66.96%	
Aggregate *N.rating*	90,353	3,023,859	29,383,773	
Average *Gini.revenue*	0.688	0.639	0.535	
Average *top-20 share*	0.742	0.675	0.630	

Note: There are 3,222 brands operating on all three platforms.

Table 2: Summary Statistics of Store-level Data

Variable	Obs	Overall Mean	St. Dev.	Min	Max
Ele.me	102,037	0.360	0.480	0	1
Meituan	102,037	0.392	0.488	0	1
y	102,037	436.635	916.256	1	16,712
p	102,037	32.049	97.133	0.010	9,999
revenue	102,037	10,288.680	37,780.600	0.010	2,225,805
rating	91,340	4.587	0.309	1.800	5
N.rating	98,297	338.883	1,407.747	0	61,644
deliv.min.p	92,129	24.362	22.567	0	2,000
deliv.fee	98,016	6.690	6.184	0	205
deliv.time	96,188	38.956	12.851	0	565
J	102,037	154.821	160.569	1	761
activeness	102,037	3.581	2.055	1	7
weight.position	102,037	199.814	168.093	1	1,707.083
lag.weight.position	102,037	267.102	172.475	1	1,836.143
position.other.market	70,165	224.731	145.906	1	1,707.083
N.brand.stores	102,037	45.651	104.396	1	640
N.brand.stores.plat	102,037	19.617	44.296	1	328

Table 2 summarizes the store-level data. A store is defined as a restaurant page on a food delivery platform. One physical restaurant might operate on multiple platforms, and each of them is treated as a different store. For each store, we observe its monthly average number of orders, which is treated as the quantity y. Moreover, for each item on the menu, we observe its price and monthly sales, from which we compute the weighted average price p and *revenue*.[28] On average, a store earns a monthly revenue of CNY 10,289 by operating on one food delivery platform. We observe the overall *rating* of the store on a scale from 0 to 5 and the total number of ratings

[28]For example, there are three items on the menu of a store with prices 2, 3, and 4. The monthly average sales of the three items are 15, 12, and 5, respectively. The revenue is $2*15+3*12+4*5=86$. The weighted average price is $p=86/(15+12+5)=2.6875$.

(*N.rating*) given by consumers.[29] For most stores, we also observe their characteristics of delivery including the minimum price for delivery (*deliv.min.p*), delivery fee (*deliv.fee*) and average time of delivery (*deliv.time*) in minutes.

We define each market as a unique combination of location, category, and platform. Let J denote the number of stores in each market, which measure the competitiveness of the market. The variable *activeness* is a big-data based index measuring the activeness of economy in the region provided by DTCJ (`www.dtcj.com`), which is illustrated in Figure 6. We use *activeness* to control for the consumer size of different markets.

To investigate the effect of the search algorithm, we record the positions of all stores as appeared in the search result. In July 2018, we chose 30 random times between 10:00 AM to 2:00 AM and used a computer program to record the position of each store in the default search result by location (subway station), category, and platform. In August 2018, we conducted the same exercise. These position records are used to construct the variables, *lag.weight.position* and *weight.position* for each store in Table 2. Specifically, the store that appears at the top of the search result has *position*= 1; the store that appears in second place has *position*= 2, and so forth. Because many stores appear in several markets, we compute the weighted average position of each store by using the *activeness* of different markets as weights and obtain *lag.weight.position* and *weight.position*. The *weight.position* is the proxy for how difficult it is for a consumer to find this store given the algorithm. The variable *position.other.market* is the average *weight.position* of stores with the same brands that operate in at other markets. We will use *lag.weight.position* and *position.other.market* as instrumental variable (IV) for *weight.position* in regression of store *revenue*.

Note that the ranking of stores is likely to be personalized to consumers based on their transaction history and demographic information. We do not have administrative data that contains click-through records and personalized ranking of individual consumers, so we are unable to study the effect of personalized search results and targeted advertisement in this paper. Using the default search results match the baseline model in Section 2 in which all consumers face the same algorithm.

Table 3: Number of Stores and Revenue Share by Brand

N.brand.stores	Obs	Percentage	Total revenue	Revenue share
≥1	34,823	100%	1,049,826,347	100%
≥2	14,235	40.88%	974,365,741	92.81%
≥3	7,739	22.22%	901,015,003	85.83%
≥5	3,660	10.51%	801,825,095	76.38%
≥10	1,211	3.48%	590,139,462	56.21%
≥50	170	0.49%	282,892,568	26.95%

[29] The rating system is similar to Yelp. Consumers leave a rating from 0 to 5 stars. Then the system uses a formula to compute an overall rating. A majority of buyers give 5 stars; thus the average is 4.106. This phenomenon is normal in most online reputation systems. See Dellarocas and Wood (2008).

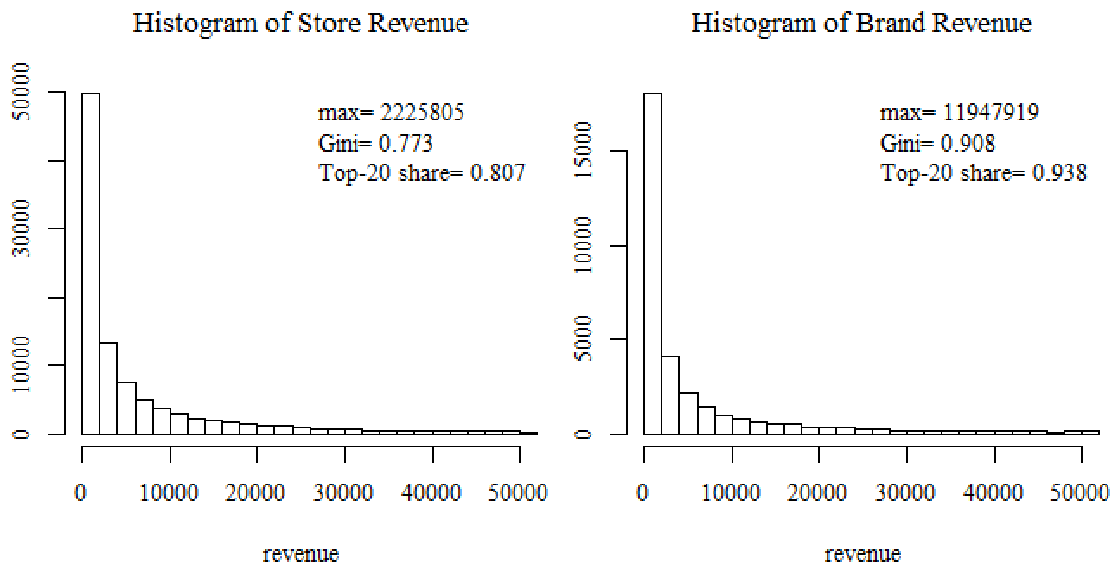


Figure 7: Measures of Market Concentration

By textual analyses, we find that the stores belong to 34,823 brands (restaurant chains that share the same brand). Based on brand identities, *N.brand.stores* counts the number of stores of each restaurant chain on all three platforms. For each store, *N.brand.stores.plat* counts the number of stores with the same brand on its platform. In the data, the largest restaurant chain is Shaxian Xiaochi with 640 stores in total on the three platforms. Table 3 shows that approximately 10% of the brands have more than 5 stores, but they comprise 76% of the total revenue in the entire market. Figure 7 shows that sales are substantially more concentrated by treating a brand as the unit instead of a store. These data patterns indicate that a branded store earns a much higher revenue than independent restaurants.

Table 4: Summary Information by Category of Food

Category	Code	Obs	Average	Total	Share
Formal meal	1	31,225	8,686	271,229,794	25.836%
Fast food	2	29,694	10,414	309,241,027	29.456%
Regional dishes	3	5,682	16,716	94,981,469	9.047%
Foreign food	4	1,941	14,800	28,726,306	2.736%
Afternoon tea	5	8,678	5,194	45,073,466	4.293%
Fruit and vegetable	6	3,670	19,899	73,028,113	6.956%
Supermarket	7	3,222	5,704	18,379,149	1.751%
Cake and flower	8	397	4,894	1,942,844	0.185%
Dessert and drink	9	5,511	19,769	108,948,352	10.378%
Snack	10	1,117	3,531	3,943,934	0.376%
Medicine	11	6,424	3,882	24,936,504	2.375%
Other	12	4,476	15,504	69,395,391	6.610%

Table 5: Summary Statistics of Market-level Data

Variable	Obs	Mean	St. Dev.	Min	Max
J	7,442	210.1	244.3	1	1,683
market.revenue	7,442	1,874,951	2,429,538	11.32	31,292,403
market.y	7,442	72,770	128,242	1	1,596,397
market.N.rating	7,442	123,532	232,539	0	1,533,184
market.ave.p	7,442	61.82	65.29	1.945	1,708
Gini.revenue	7,442	0.640	0.152	0	0.937
top-20 share	7,442	0.693	0.151	0	0.998
Gini.position	7,442	0.155	0.084	0	0.578
N.brand.stores≥ 10	7,442	69.24	82.943	0	584
N.brand.stores.plat≥ 10	7,442	41.67	52.48	0	462
activeness	7,442	3.382	1.970	1	7
Ele.me	7,442	0.435	0.496	0	1
Meituan	7,442	0.174	0.379	0	1

There are 12 major categories of items listed in Table 4. According to the location-category-platform combinations, there are 7,442 markets. Table 5 provides the summary statistics of the market-level data. For each market, we compute the market aggregate revenue (*market.revenue*), aggregate number of monthly orders (*market.y*), total numbers of ratings and reviews (*market.N.rating*), and the average price of delivery orders (*market.ave.p*). We measure the sales concentration of a market by the Gini coefficient of store revenues (*Gini.revenue*)[30] and the market share of the top 20% of sellers (*top-20 share*).[31] For each market, we count the number of brands with more than 10 stores on this platform (*N.brand.stores.plat*≥ 10) and across the three platforms (*N.brand.stores*≥ 10) based on brand identities. These two variables measure the presence of large chain restaurants in the market.

In Table 1, we observe that the total number of ratings and reviews on Meituan is significantly larger than those of the other two platforms. The substantial difference in the volume of ratings and reviews across these three platforms is rooted in the difference in their reputation systems. Meituan has the best and most informative reputation system because it is accompanied by Dianping, the counterpart of Yelp in China. On Meituan, there are enormous volumes of lengthy and detailed user-contributed reviews. Ele.me has a less informative reputation system where most information comes in the form of scores (stars) instead of textual reviews. Baidu has the weakest reputation system that has nearly no textual reviews. Consumers who use Meituan receive more quality information than if they use the other two platforms. Figure 8 depicts the *Gini.revenue* and *top-20 share* of different markets across the three platforms. We find that Meituan exhibits the smoothest sales, while the sales on Baidu are more concentrated. Therefore, having a better reputation system, and

[30]The Gini coefficient is widely used as a measure of sales concentration (Fleder and Hosanagar, 2009, 2007; Brynjolfsson et al., 2011). We compute it using the method in Gaswirth (1972).

[31]We select 20% because the Pareto principle (`en.wikipedia.org/wiki/Pareto_principle`) suggests that the top 20% of sellers can capture 80% of the market (Brynjolfsson et al., 2011).

more quality information is correlated with less concentrated sales. This phenomenon is consistent with the classic result in the literature that reputation systems alleviate asymmetric information (e.g., Klein et al. 2016).

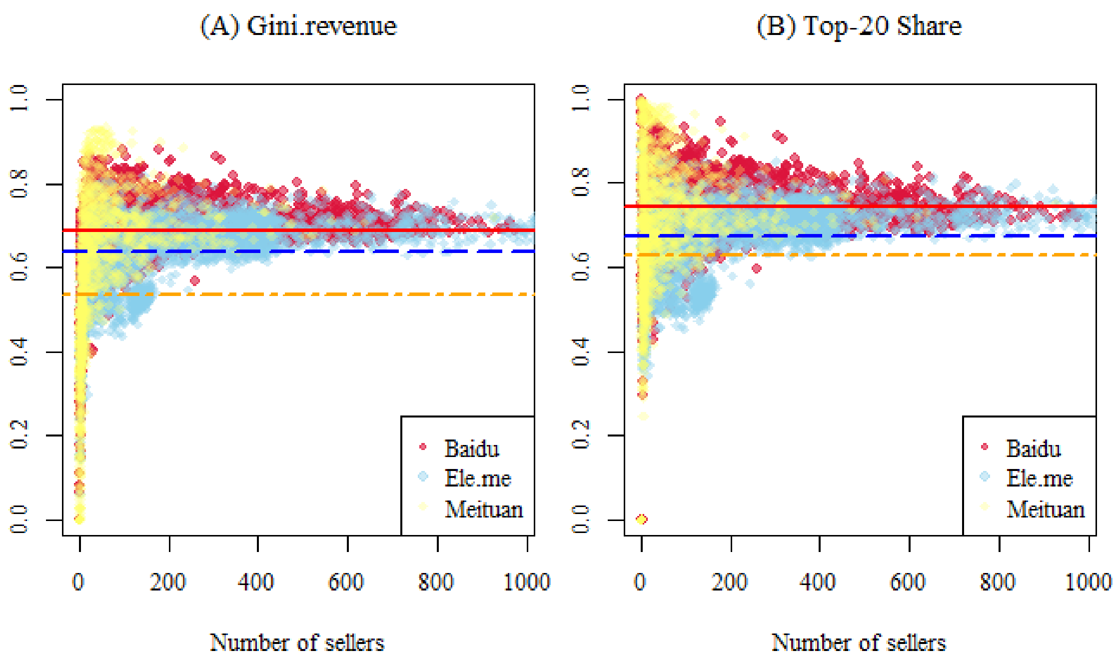


Figure 8: Measures of Market Concentration

For each market, we compute the Gini coefficient (*Gini.position*) of *weight.position* for all stores in the market and use it to measure how selective is the search algorithm. Under the equal-treatment search algorithm, the ranking of the stores are random. As a result, the *weight.position* of all stores will be close to one another, and the *Gini.position* will be close to zero. By contrast, if the search algorithm is highly selective, some stores will always appear on top of the search results, while some other stores will always appear at the bottom. The large difference of the *weight.position* across stores result in a large *Gini.position* for the market.

4.3 Empirical results

We perform several sets of linear regressions that demonstrate the role of the search algorithm in determining the sales of individual stores and the sales distributions of the market.

First, in Table 6, we use *weight.position* as the dependent variable and explore which factors affect store rankings. A store with a smaller *weight.position* indicates that the search algorithm ranks this store higher. From the negative coefficients of *N.brand.stores* and *N.brand.stores.plat*, we observe that large chain restaurants receive substantial advantages from the search algorithm. Stores with established brands appear more often at the top of search results. Stores with a higher *rating*, more ratings (*N.rating*), and more sales (y) or *revenue* tend to be ranked higher. In general, the search algorithms promote stores with better consumer reviews and greater historical sales.

Table 6: Regressions of Weighted Position in Search Results

	Dependent variable: *weight.position*					
	(1)	(2)	(3)	(4)	(5)	(6)
Platform	Baidu	Ele.me	Meituan	Baidu	Ele.me	Meituan
N.brand.stores.plat	−0.253***	−0.072***	−0.032**	−0.278***	−0.200***	0.494***
	(0.017)	(0.009)	(0.015)	(0.043)	(0.021)	(0.042)
N.brand.stores				0.009	0.061***	−0.259***
				(0.014)	(0.009)	(0.019)
rating	16.543***	−1.579	−39.851***	16.601***	−1.561	−41.549***
	(3.073)	(1.286)	(2.253)	(3.075)	(1.285)	(2.250)
N.rating	−0.004***	−0.283***	−0.093***	−0.004***	−0.288***	−0.086***
	(0.0003)	(0.006)	(0.015)	(0.0003)	(0.006)	(0.015)
J	0.134***	0.058***	0.449***	0.134***	0.059***	0.446***
	(0.010)	(0.004)	(0.015)	(0.010)	(0.004)	(0.015)
revenue	−0.002***	−0.0004***	−0.0001***	−0.002***	−0.0004***	−0.0001***
	(0.0001)	(0.00003)	(0.00001)	(0.0001)	(0.00003)	(0.00001)
category FE	Y	Y	Y	Y	Y	Y
station FE	Y	Y	Y	Y	Y	Y
Observations	25,321	32,897	30,663	25,321	32,897	30,663
R^2	0.446	0.434	0.640	0.447	0.434	0.642

Note: For all regressions results reported in this paper, * indicates significance at 10%; ** indicates significance at 5%; and *** indicates significance at 1%. The dependent variable *weight.position* is the weighted average position of the store. *N.brand.stores.plat* and *N.brand.stores* are number of stores share the same brand on the platform and across three platforms, respectively.

Next, we conduct store-level regressions by using *revenue* as the dependent variable. Table 7 reports the results. The main coefficients of interest are those for *weight.position*. Because the *weight.position* of a store endogenously depends on its *revenue* (as shown in Table 6, we need to introduce IVs. The first IV is *lag.weight.position*, which is the weighted position of the store from last month. This IV is valid given that the position of the store in the previous month does not directly affect the *revenue*. The second IV is *position.other.market*, which is the position of the store in search results at other markets. It is unlikely that the ranking of stores in one market depends on their positions at the other markets on different platforms. However, *position.other.market* does not cover independent restaurants with only one store on one platform. Table 10 shows that both instruments are sufficiently strong.

From Table 7, we have the following findings. First, revenue decreases drastically in the store position in the default search result. The estimated coefficients for *weight.position* in regressions (4) and (8) suggest that being ranked one position lower leads to more than a CNY20 monthly revenue reduction.[32] The scale is similar to that of Ursu (2018), who finds that the average position effect is USD1.92 on Expedia. Consumers heavily rely on the default ranking when making purchases on

[32]We implicitly assume that the effect is linear. In reality, the difference between position 1 and 10 can be much larger than the difference between position 101 and 110.

Table 7: Store-level Regression Results

	Dependent variable: *revenue*							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	TSLS	TSLS	OLS	OLS	TSLS	TSLS
weight.position		−28.841***	−10.155***	−34.111***		−25.489***	−4.404***	−22.052***
		(1.141)	(2.777)	(3.749)		(0.620)	(1.495)	(1.935)
rating	1,576.349***	960.159**	1,359.398***	328.122	5,275.698***	4,851.036***	5,202.327***	6,589.669***
	(456.713)	(455.731)	(459.611)	(672.921)	(257.317)	(254.751)	(257.653)	(360.037)
N.rating	1.819***	1.565***	1.730***	1.468***	1.746***	1.539***	1.710***	1.468***
	(0.096)	(0.096)	(0.099)	(0.126)	(0.049)	(0.048)	(0.050)	(0.060)
N.brand.stores	2.096	−0.774	1.086	−3.401	−19.655***	−21.107***	−19.906***	−23.063***
	(3.026)	(3.017)	(3.032)	(3.630)	(1.991)	(1.970)	(1.986)	(2.232)
N.brand.stores.plat	−8.444	−4.454	−7.039	−10.347	58.331***	59.170***	58.476***	51.957***
	(7.178)	(7.154)	(7.173)	(8.534)	(5.218)	(5.161)	(5.200)	(5.823)
J	6.849***	11.118***	8.352***	12.835***	3.231***	6.483***	3.793***	6.263***
	(1.366)	(1.372)	(1.424)	(2.093)	(0.747)	(0.743)	(0.768)	(1.059)
activeness	−52.485	474.911***	133.203	593.333**	−52.941	415.104***	27.926	410.141***
	(159.709)	(160.500)	(167.270)	(236.624)	(88.428)	(88.212)	(92.306)	(123.984)
Ele.me	3,919.699***	5,291.589***	4,402.722***	7,077.393***	6,359.224***	7,496.141***	6,555.655***	9,180.537***
	(541.290)	(542.078)	(556.081)	(825.718)	(298.112)	(296.191)	(304.496)	(422.766)
Meituan	19,791.070***	23,067.740***	20,944.740***	24,846.380***	15,768.380***	18,599.400***	16,257.510***	19,302.360***
	(594.901)	(606.788)	(672.274)	(982.181)	(323.552)	(327.390)	(362.712)	(498.445)
deliv.min.p					−10.943**	−9.930**	−10.768**	−17.078***
					(4.444)	(4.397)	(4.430)	(6.180)
deliv.fee					−103.004***	−83.006***	−99.549***	−97.073***
					(17.048)	(16.871)	(17.031)	(22.135)
deliv.time					86.643***	106.416***	90.059***	97.263***
					(7.849)	(7.779)	(7.907)	(10.217)
category FE	Y	Y	Y	Y	Y	Y	Y	Y
station FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	88,881	88,881	88,881	61,277	77,310	77,310	77,310	52,793
R^2	0.059	0.066	0.063	0.059	0.121	0.140	0.127	0.132

Note: Columns (3), (4), (7), and (8) report results from two-stage-least-square (TSLS) regressions. Regressions (3) and (7) use one instrumental variable *lag.weight.position*. Regressions (4) and (8) use two instrumental variables *lag.weight.position* and *position.other.markets*.

online platforms. Hence, the search algorithm significantly influences individual stores' revenues, which is consistent with the setting of the model.

Second, a store with a better *rating* and more reviews tends to obtain more sales, which suggests that the reputation system is effective. This result is intuitive and consistent with many existing studies. For example, Fang (2020) finds that online reviews help consumers learn about restaurants' quality and redistribute revenues from low-quality restaurants to high-quality ones. Cabral and Hortacsu (2010) show that negative reviews lead to a significant and persistent reduction in sales on eBay. We emphasize that good ratings and reviews not only directly attract consumers but also indirectly raise sales on the platform through the search algorithm. In Table 7, regression (1) and (5) are similar specifications as those in Cabral and Hortacsu (2010) and Fang (2020). After adding *weight.position*, the coefficients of *rating* decrease in most specifications. From Table 6, we learn that *rating* affects sales through the search algorithm. Accordingly, if we ignore that the channel of search algorithms, the direct effect of *rating* on revenue would be over-estimated.

Third, in regressions (5)-(8) of Table 7, the coefficients of *N.brand.stores.plat* are positive and have larger magnitudes than those of *N.brand.stores*. This indicates that having more of the same-branded stores that operate on the same platform significantly increases the average store revenue. As shown in Table 6, large chain stores earn more revenue partly because they are promoted by the search algorithm. According to Corollaries 3 and 4, adopting such selective algorithms may suppress competition, drive up prices, and hurt consumer welfare.

We continue with some market-level regressions. In Table 8, regressions (1) and (2) use the *Gini.revenue* in the market as the dependent variable. In regressions (3) and (4), the dependent variable is the market share of the top 20% stores. Table 9 uses the market average price (*market.ave.p*) as the dependent variable. It is reasonable to assume that restaurants select their locations without considering the impact of the food delivery platforms. Most restaurants started their businesses before food-delivery platforms became popular in 2013, and location choice is based primarily on offline buyers. With this assumption, the number and composition of stores in each market is exogenously formed. We can then explore whether having more stores with established brands causes sales to be more concentrated.

From the results in Tables 8 and 9, we can observe the dual role of online platforms characterized in the model. The coefficients of *Gini.position* are all significantly positive. This indicates that markets with less equal search results tend to have more skewed sales distributions and higher average prices. This suggests that the skewed-distributed positions of stores reduce competitive pressure faced by stores. The coefficients of J are positive in all cases; thus, having more stores in a market cannot smooth the sales distribution or lower the average price. One explanation is that consumers' consideration sets do not expand when having more stores available in the market. The results are consistent with Corollary 1.

Table 8: Market-level Regression on Inequality of Sales Distribution

Dependent variable	*Gini.revenue*×100		*top-20 share*×100	
	(1)	(2)	(3)	(4)
Gini.position×100	0.493***	0.490***	0.445***	0.442***
	(0.043)	(0.043)	(0.048)	(0.048)
J	0.016***	0.017***	0.013***	0.018***
	(0.003)	(0.002)	(0.003)	(0.002)
activeness	−10.588***	−10.137***	−6.612**	−6.366**
	(2.365)	(2.370)	(2.650)	(2.656)
N.brand.stores≥10	0.048***		0.046***	
	(0.009)		(0.009)	
N.brand.stores.plat≥10		0.080***		0.052***
		(0.010)		(0.010)
market.N.rating	−0.00001***	−0.00001***	−0.00001***	−0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
market.y	−0.00002***	−0.00002***	−0.00002***	−0.00002***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
market.ave.p	0.005	0.006	0.004	0.004
	(0.007)	(0.007)	(0.008)	(0.008)
Ele.me	−7.952***	−7.646***	−10.946***	−10.838***
	(0.463)	(0.472)	(0.499)	(0.509)
Meituan	−12.313***	−12.462***	−10.300***	−10.380***
	(0.846)	(0.847)	(0.904)	(0.906)
category FE	Y	Y	Y	Y
station FE	Y	Y	Y	Y
Observations	7,442	7,442	7,442	7,442
R^2	0.361	0.363	0.265	0.265

Note: The dependent variable *Gini.revenue* is the Gini coefficient of revenue of stores operating in the market; *top-20 share* is the share of total revenue of top 20% largest stores in the market. *N.brand.stores≥10* and *N.brand.stores.plat≥10* count the number of large chain stores (with more than 10 stores overall and on this platform, respectively) operating in the market.

Moreover, the coefficient estimates of *N.brand.stores≥10* and *N.brand.stores.plat≥10* are significantly positive in all four regressions in Table 8, which suggests that having more large chain restaurants in the market leads to a higher sales inequality. The results in Tables 6 and 7 suggest that large chain restaurants suppress competition partly through their advantageous position in the search results. As discussed in Section 3.1, online platforms have the incentive to promote restaurants with established brands. Therefore, compared to offline channels, online platforms may have magnified the advantage of large chain restaurants over small independent restaurants. Sellers with established brands might not always be the best options for consumers. Bronnenberg et al. (2015) find that the preference for large brands decreases as consumers accumulate more experiences and information, which implies that brands might cause consumers to purchase the "wrong" products given limited information.

Table 9: Market-level Regression Results on Market Average Price

	Dependent variable: *market.ave.p*			
	(1)	(2)	(3)	(4)
Gini.position×100	0.764***	0.619***	0.705***	0.716***
	(0.107)	(0.148)	(0.159)	(0.160)
J			0.038***	0.045***
			(0.008)	(0.008)
activeness			1.941	0.996
			(7.709)	(7.710)
N.brand.stores≥10			−0.069***	
			(0.023)	
N.brand.stores.plat≥10				−0.143***
				(0.026)
Ele.me		13.955***	10.946***	10.028***
		(0.620)	(0.880)	(0.980)
Meituan		42.825***	44.231***	44.527***
		(3.380)	(3.748)	(3.709)
category FE	N	Y	Y	Y
station FE	N	Y	Y	Y
Observations	7,442	7,442	7,442	7,442
R^2	0.010	0.396	0.397	0.397

Note: The dependent variable *market.ave.p* is the weighted average price of all delivery orders across all stores operating in the market.

In addition to the anti-competitive impacts rendered by online platforms, we also find evidence for the pro-competitive impact. Table 8 shows that as the market accumulates more sales (*market.y*) and more reviews (*market.N.rating*), the sales concentration decreases. One reason could be the quality information released through sales and reviews (Huang and Zuo, 2020), which helps consumers learn the quality of small independent restaurants and rely less on brands. As shown in Figure 8, the effectiveness of reputation systems varies across the platforms. Small restaurants perform better on the platform with more transparent quality information, which helps buyers search and discover products.

In summary, the position of a restaurant in search results significantly affects its revenue. Good reviews and established chain-store brands not only directly increase the revenue of a store but also indirectly increase it through the search algorithm as they help the store obtain a higher position. Because the search algorithms provide an advantage to chain restaurants, having more stores with established brands in the market substantially increases sales concentration. Moreover, having a better reputation system with more quality information reduces the sales concentration. These observations demonstrate the dual role of online platforms presented in the model.

5 Conclusion

Many industries have experienced drastic changes and have even been reshuffled by online platforms. As sales shift from offline to online, platforms obtain the power to influence the consumers' searching behavior through the design of search algorithms. We develop a novel model that captures how a search algorithm affects buyers' search processes, which further affect market equilibrium and welfare. Adopting a highly selective search algorithm can be anti-competitive because consumers consider fewer options, and sellers can charge higher prices. Consumers' interest can be further jeopardized if the search algorithm promotes low-quality products.

We study a novel dataset from food delivery platforms that contains information about seller positions in the ranking of search results. We find that the search algorithm can deeply influence store revenue and sales distribution. In general, the search algorithm places sellers with established brands on top of the search results, which leads to more concentrated sales distributions. These results are consistent with the model implications.

The modeling framework proposed in this paper provides a tractable analysis of search algorithms, which can potentially be used for welfare analysis, antitrust investigations, and the regulation of online platforms. For example, as the platform chooses the algorithm for its private purposes, how much will the outcome deviate from the welfare-maximizing outcome? If there is platform-level competition, how will the algorithm change compare with the case of a monopoly platform? Finally, knowing that the search algorithm has such a substantial impact on the market structure and welfare, how should regulators and antitrust authorities oversee search algorithms? We leave these questions for future research.

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics 84*(3), 488–500.

Anderson, S. P. and A. De Palma (2009). Information congestion. *RAND Journal of Economics 40*(4), 688–709.

Armstrong, M. (2006). Competition in two-sided markets. *RAND Journal of Economics 37*(3), 668–691.

Armstrong, M. and J. Wright (2007). Two-sided markets, competitive bottlenecks and exclusive contracts. *Economic Theory 32*(2), 353–380.

Athey, S. and G. Ellison (2011). Position auctions with consumer search. *Quarterly Journal of Economics 126*(3), 1213–1270.

Backus, M. R., J. U. Podwol, and H. S. Schneider (2014). Search costs and equilibrium price dispersion in auction markets. *European Economic Review 71*, 173–192.

Bar-Isaac, H., G. Caruana, and V. Cuñat (2012). Search, design, and market structure. *American Economic Review 102*(2), 1140–1160.

Baye, M. R. and J. Morgan (2001). Information gatekeepers on the internet and the competitiveness of homogeneous product markets. *American Economic Review 91*(3), 454–474.

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

Bloch, F. and D. Manceau (1999). Persuasive advertising in hotelling's model of product differentiation. *International Journal of Industrial Organization 17*(4), 557–574.

Bronnenberg, B. J., J.-P. Dubé, M. Gentzkow, and J. M. Shapiro (2015). Do pharmacists buy bayer? informed shoppers and the brand premium. *Quarterly Journal of Economics 130*(4), 1669–1726.

Brynjolfsson, E., Y. Hu, and D. Simester (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science 57*(8), 1373–1386.

Burdett, K. and K. L. Judd (1983). Equilibrium price dispersion. *Econometrica*, 955–969.

Cabral, L. and A. Hortacsu (2010). The dynamics of seller reputation: evidence from ebay. *The Journal of Industrial Economics 58*(1), 54–78.

Chen, N. and H.-T. Tsai (2020). Steering via algorithmic recommendations. *Working Paper*.

Chen, Y. and C. He (2011). Paid placement: Advertising and search on the internet. *Economic Journal 121*(556), F309–F328.

Chen, Y. and T. Zhang (2017). Entry and welfare in search markets. *Economic Journal 128*(608), 55–80.

Chen, Y. and T. Zhang (2018). Intermediaries and consumer search. *International Journal of Industrial Organization 57*, 255–277.

De Corniere, A. (2016). Search advertising. *American Economic Journal: Microeconomics 8*(3), 156–88.

De Corniere, A. and G. Taylor (2014). Integration and search engine bias. *RAND Journal of Economics 45*(3), 576–597.

De Corniere, A. and G. Taylor (2019). A model of biased intermediation. *RAND Journal of Economics 50*(4), 854–882.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review 102*(6), 2955–80.

De los Santos, B. and S. Koulayev (2017). Optimizing click-through in online rankings with endogenous search refinement. *Marketing Science 36*(4), 542–564.

Dellarocas, C. and C. A. Wood (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science 54*(3), 460–476.

Dukes, A. and L. Liu (2015). Online shopping intermediaries: The strategic design of search environments. *Management Science 62*(4), 1064–1077.

Eliaz, K. and R. Spiegler (2011a). Consideration sets and competitive marketing. *Review of Economic Studies 78*(1), 235–262.

Eliaz, K. and R. Spiegler (2011b). A simple model of search engine pricing. *Economic Journal 121*(556), F329–F339.

Eliaz, K. and R. Spiegler (2016). Search design and broad matching. *American Economic Review 106*(3), 563–86.

European Commission (2017). Antitrust: Commission fines Google EUR2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service. `ec.europa.eu/commission/presscorner/detail/en/IP_17_1784`.

Fang, L. (2020). The effects of online review platforms on restaurant revenue, survival rate, consumer learning and welfare. *Working Paper*.

Federal Trade Commission (2013). Ftc consumer protection staff updates agency's guidance to search engine industry on the need to distinguish between advertisements and search results. www.ftc.gov.

Fleder, D. and K. Hosanagar (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science 55*(5), 697–712.

Fleder, D. M. and K. Hosanagar (2007). Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce*, pp. 192–199. ACM.

Gaswirth, J. L. (1972). The estimation of the Lorenz curve and Gini index. *Review of Economics and Statistics 54*, 306–316.

Ghose, A. and S. Yang (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science 55*(10), 1605–1622.

Goldmanis, M., A. Hortaçsu, C. Syverson, and Ö. Emre (2010). E-commerce and the market structure of retail industries. *Economic Journal 120*(545), 651–682.

Grossman, G. M. and C. Shapiro (1984). Informative advertising with differentiated products. *Review of Economic Studies 51*(1), 63–81.

Hagiu, A. and B. Jullien (2011). Why do intermediaries divert search? *RAND Journal of Economics 42*(2), 337–362.

Harstad, R. M., J. H. Kagel, and D. Levin (1990). Equilibrium bid functions for auctions with an uncertain number of bidders. *Economics Letters 33*(1), 35–40.

Hong, H. and M. Shum (2006). Using price distributions to estimate search costs. *RAND Journal of Economics 37*(2), 257–275.

Huang, Y. and S. Zuo (2020). Building reputation on online platforms. *Working Paper*.

Inderst, R. and M. Ottaviani (2012). Competition through commissions and kickbacks. *American Economic Review 102*(2), 780–809.

Jansen, B. J. and M. Resnick (2006). An examination of searcher's perceptions of nonsponsored and sponsored links during ecommerce web searching. *Journal of the Association for Information Science and Technology 57*(14), 1949–1961.

Klein, T. J., C. Lambertz, and K. Stahl (2016). Market transparency, adverse selection, and moral hazard. *Journal of Political Economy 124*(6), 1677–1713.

Marshall, A. W., I. Olkin, and B. C. Arnold (2011). *Inequalities: Theory of Majorization and Its Applications*. Springer Science & Business Media.

McAfee, R. P. and J. McMillan (1987). Auctions with a stochastic number of bidders. *Journal of Economic Theory 43*(1), 1–19.

Morgan, P. and R. Manning (1985). Optimal search. *Econometrica*, 923–944.

Ricci, F., L. Rokach, and B. Shapira (2011). Introduction to recommender systems handbook. In *Recommender systems handbook*, pp. 1–35. Springer.

Rochet, J.-C. and J. Tirole (2003). Platform competition in two-sided markets. *Journal of the European Economic Association 1*(4), 990–1029.

Rochet, J.-C. and J. Tirole (2006). Two-sided markets: A progress report. *RAND Journal of Economics*, 645–667.

Schur, I. (1923). Uber eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie. *Sitzungsberichte der Berliner Mathematischen Gesellschaft 22*(9-20).

Srinivasan, V., C. S. Park, and D. R. Chang (2005). An approach to the measurement, analysis, and prediction of brand equity and its sources. *Management Science 51*(9), 1433–1448.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics 8*, 321–340.

Teh, T.-H. and J. Wright (2020). Intermediation and steering: Competition in prices and commissions. *American Economic Journal: Microeconomics*.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science 37*(4), 530–552.

Varian, H. (1980). A model of sales. *American Economic Review 70*(4), 651–59.

Wang, C. and J. Wright (2020). Search platforms: Showrooming and price parity clauses. *RAND Journal of Economics 51*(1), 32–58.

Wong, C.-K. and P. Yue (1973). A majorization theorem for the number of distinct outcomes in n independent trials. *Discrete Mathematics 6*(4), 391–398.

Yang, H. (2013). Targeted search and the long tail effect. *RAND Journal of Economics 44*(4), 733–756.

Appendix

Proof of Lemma 1. The proof is straightforward from formula (2).

Proof of Lemma 2. Let $\mathcal{C}(K)$ denote the consideration set with search intensity K. After taking K_1 samples, we have a probability distribution of N. Then, consider taking one more sample. In this sampling, the buyer obtains ball with kind j. If $j \in \mathcal{C}(K_1)$, then $\mathcal{C}(K_1+1) = \mathcal{C}(K_1)$ and $\Pr(N \leq n) = \Pr(|\mathcal{C}(K_1+1)| \leq n)$. If $j \notin \mathcal{C}(K_1)$, then $\mathcal{C}(K_1+1) = \mathcal{C}(K_1) \bigcup \{j\}$. In this case, $|\mathcal{C}(K_1+1)| = J+1$ and $|\mathcal{C}(K_1+1)| > |\mathcal{C}(K_1)|$.

Because $K_1 < J$, the latter event occurs with positive probability, i.e, $\Pr(j \notin \mathcal{C}(K_1)) > 0$,

$$\begin{aligned}\Pr(|\mathcal{C}(K_1+1)| \leq n) &= \Pr(|\mathcal{C}(K_1)| \leq n) - \Pr(|\mathcal{C}(K_1+1)| = n+1)\\ &= \Pr(|\mathcal{C}(K_1)| \leq n) - \Pr(j \notin \mathcal{C}(K_1)) < \Pr(|\mathcal{C}(K_1)| \leq n).\end{aligned}$$

With this inequality, we can easily see the FSD property in the lemma. □

Proof of Lemma 3. We first introduce the concept of Schur-convex function. A real-valued function $\phi : \mathcal{A} \subset \mathbb{R}^j \to \mathbb{R}$ is said to be *Schur-convex*[33] on $\mathcal{A}$ if $\boldsymbol{\sigma}_1 \prec \boldsymbol{\sigma}_2$ implies $\phi(\boldsymbol{\sigma}_1) \leq \phi(\boldsymbol{\sigma}_2)$.

By Proposition E.11.b. of Marshall et al. (2011),[34] suppose that an experiment with J possible outcome is repeated K times. The number N of distinct outcomes is a random variable representing the nonzero components of the multinomial random vector $\boldsymbol{x}$. Then, $\psi(\boldsymbol{\sigma}) = \Pr(N \leq n|\boldsymbol{\sigma})$ is a Schur-convex function of $\boldsymbol{\sigma}$ for all n. By the definition of a Schur-convex function,

$$\boldsymbol{\sigma}_1 \prec \boldsymbol{\sigma}_2 \Rightarrow \Pr(N \leq n|\boldsymbol{\sigma}_1) \leq \Pr(N \leq n|\boldsymbol{\sigma}_2) \Leftrightarrow N_1 \succcurlyeq_1 N_2.$$

□

Proof of Theorem 1. Consider a symmetric and monotone BNE $u = \mu(w)$. Let $\mu^{-1}(\cdot)$ denote the inverse of μ. Provided that all other sellers follow the BNE, the probability of the generic seller

[33] It is named after Schur (1923), who provides the first systematic study of order-preserving functions for majorization. The term "convex" comes from the property of a convex function. Given a random variable σ that can have different realizations, $\sigma_1, \sigma_2, .., \sigma_n$, for all convex functions $g : \mathbb{R} \to \mathbb{R}$, we have $\sum_{i=1}^n g(\bar{\sigma}) \leq \sum_{i=1}^n g(\sigma_i)$.

[34] This proposition is originally derived by Wong and Yue (1973)

j being chosen is

$$\begin{aligned}
\Pr\left(u \geq \max_{j' \in \mathcal{C}}\{\mu(w_{j'})\}\right) &= \Pr\left(\mu^{-1}(u) \geq \max_{j' \in \mathcal{C}}\{\mu^{-1}(\mu(w_{j'}))\}\right) \\
&= \Pr\left(\mu^{-1}(u) \geq \max_{j' \in \mathcal{C}}\{w_{j'}\}\right) \\
\text{By (3)} &= \sum_{n=1}^{J}\left\{P_n \times \Pr\left(\mu^{-1}(u) \geq W_{(1:n-1)}\right)\right\} \\
&= \sum_{n=1}^{J}\left\{P_n \times G_{(1:n-1)}\left(\mu^{-1}(u)\right)\right\} \equiv \mathcal{P}\left(\mu^{-1}(u)\right). \qquad (12)
\end{aligned}$$

In equilibrium, $\mathcal{P}\left(\mu^{-1}(u)\right) = \mathcal{P}(w)$ represents the choice probability of a seller with $w = w^*(\theta)$.

The generic seller with w chooses u by solving

$$\max_u (w-u)\mathcal{P}\left(\mu^{-1}(u)\right) = \max_u (w-u)\sum_{n=1}^{J}\left\{P_n \times G_{(1:n-1)}\left(\mu^{-1}(u)\right)\right\}.$$

This problem is equivalent to a first-price auction with an uncertain number of bidders. We derive its BNE following McAfee and McMillan (1987) and Harstad et al. (1990). The first-order condition yields

$$\frac{d\mathcal{P}\left(\mu^{-1}(u)\right)}{du}(w-u) - \mathcal{P}\left(\mu^{-1}(u)\right) = 0,$$

which is an ordinary differential equation. In equilibrium, $\mu^{-1}(u) = w$, $u = \mu(w)$, so

$$\frac{\mathcal{P}'(w)}{\mu'(w)}(w - \mu(w)) - \mathcal{P}(w) = 0.$$

which implies $\mathcal{P}(w)\,\mu'(w) + \mathcal{P}'(w)\,\mu(w) = w\mathcal{P}'(w)$, and thus,

$$\frac{d}{dw}\left[\mathcal{P}(w)\,\mu(w)\right] = w\mathcal{P}'(w).$$

Given the boundary condition $\mu(\underline{w}) = 0$ and integrating on both sides,

$$\mathcal{P}(w)\,\mu(w) = \int_{\underline{w}}^{w} \omega\mathcal{P}'(\omega)\,d\omega = w\mathcal{P}(w) - \int_{\underline{w}}^{w}\mathcal{P}(\omega)\,d\omega.$$

The solution is

$$\mu(w) = \frac{\int_{\underline{w}}^{w} \omega\mathcal{P}'(\omega)\,d\omega}{\mathcal{P}(w)} = w - \frac{\int_{\underline{w}}^{w}\mathcal{P}(\omega)\,d\omega}{\mathcal{P}(w)}, \quad w \in [\underline{w}, \overline{w}].$$

$\mu(w)$ is obviously increasing in w, so it is a proper monotone BNE.

With fixed quality at $q^*(\theta)$, to offer a product with utility $\mu(w^*(\theta))$, the seller sets prices at $p^*(\theta) = v(q^*(\theta)) - \mu(w^*(\theta))$. □

Proof of Corollary 1. We can express $\mu(w)$ as a conditional expectation taken upon the random consideration set size N:

$$\begin{aligned}\mu(w) &= \frac{\int_{\underline{w}}^{w} \omega \mathcal{P}'(\omega)\, d\omega}{\mathcal{P}(w)} = \frac{\int_{\underline{w}}^{w} \omega \frac{d\left[\sum_{n=1}^{J}\left\{P_n \times G_{(1:n-1)}(w)\right\}\right]}{d\omega} d\omega}{\mathcal{P}(w)} \\ &= \frac{\int_{\underline{w}}^{w} \omega \left[\sum_{n=1}^{J}\left\{P_n \times g_{(1:n-1)}(w)\right\}\right] d\omega}{\mathcal{P}(w)} = \sum_{n=1}^{J} P_n \times \frac{\int_{\underline{w}}^{w} \omega g_{(1:n-1)}(w)\, dw}{\mathcal{P}(w)} \\ &= \sum_{n=1}^{J} P_n \times E\left[W_{(1:N-1)} | W_{(1:N-1)} < w\right] = E_N\left\{E\left[W_{(1:N-1)} | W_{(1:N-1)} < w\right]\right\}.\end{aligned}$$

$E\left[W_{(1:N-1)} | W_{(1:N-1)} < w\right]$ is an conditional expectation of $W_{(1:N-1)}$. By the property of order statistics, the maximum among a larger sample must have a higher expectation, so $E\left[W_{(1:N-1)} | W_{(1:N-1)} < w\right]$ is an increasing function of N. By the property of stochastic order, given that , because $N_1 \succcurlyeq_1 N_2$, we have

$$\mu(w; N_1) = E_{N_1}\left\{E\left[W_{(1:N_1-1)} | W_{(1:N-1)} < w\right]\right\} \geq E_{N_2}\left\{E\left[W_{(1:N_2-1)} | W_{(1:N_2-1)} < w\right]\right\} = \mu(w; N_2).$$

Intuitively, it indicates that the seller "bids" more aggressively in an "auction" with more rival sellers.

It follows that $p^*(w; N_1) \leq p^*(w; N_2)$ because $p^*(\theta) = v(q^*(\theta)) - \mu(w^*(\theta))$ and $q^*(\theta)$ does not depend on N. □

Proof of Theorem 2. By the property of stochastic order, if $N_2 \succcurlyeq_1 N_1$, then $E[\varphi(N_2)] \geq E[\varphi((N_1)]$ for any (weakly) increasing function φ.

(1) Social welfare is, $SW_N = I \cdot E_N\left[E\left[W_{(1:N)}\right]\right]$. Define

$$\xi(N) = E\left[W_{(1:N)}\right] = \int_{\underline{w}}^{\overline{w}} w dG_{(1:N)}(w).$$

By the property of order statistics, the maximum among a larger sample must have a higher expectation, so $\xi(\cdot)$ is increasing in N. Because $N_1 \succcurlyeq_1 N_2$, we have $E_{N_1}[\xi(N_1)] \geq E_{N_2}[\xi(N_2)]$. Therefore,

$$SW_{N_1} = I \cdot E_{N_1}[\xi(N_1)] \geq I \cdot E_{N_2}[\xi(N_2)] = SW_{N_2}.$$

That is, when N increases in the sense of FSD, the social welfare SW increases.

(2) Buyer-side utility is $U_N = I \cdot E_N\left[E\left[\mu(W_{(1:N)})\right]\right]$. Define $\zeta(N) = E\left[\mu(W_{(1:N)})\right]$. Because both $\mu(\cdot)$ and $W_{(1:N)}$ are increasing in N, $\zeta(\cdot)$ is increasing in N. Hence, $N_1 \succcurlyeq_1 N_2$ implies $E_{N_1}[\zeta(N_1)] \geq E_{N_2}[\zeta(N_2)]$, which further implies

$$U_{N_1} = I \cdot E_{N_1}[\zeta(N_1)] \geq I \cdot E_{N_2}[\zeta(N_2)] = U_{N_2}.$$

(3) Seller-side profit is $\Pi_N = I \cdot E_N \left[E \left[W_{(1:N)} - \mu(W_{(1:N)}) \right] \right]$. Define $\varsigma(N) = E \left[W_{(1:N)} - \mu(W_{(1:N)}) \right]$. Recall that $\mu(w) = w - \int_{\underline{w}}^{w} \mathcal{P}(\omega)\, d\omega / \mathcal{P}(w)$, so $\varsigma(N) = E \left[\int_{\underline{w}}^{W_{(1:N)}} \mathcal{P}(\omega)\, d\omega / \mathcal{P}\left(W_{(1:N)} \right) \right]$ is the bid shading in a first-price auction with a uncertain number of bidders. By Theorems 1 and 2 from Harstad et al. (1990), $\varsigma(N)$ decreases in N.

Given that $N_1 \succcurlyeq_1 N_2$, $E_{N_1}[\varsigma(N_1)] \leq E_{N_2}[\varsigma(N_2)]$, which implies

$$\Pi_{N_1} = I \cdot E_{N_1}[\varsigma(N_1)] \leq I \cdot E_{N_2}[\varsigma(N_2)] = \Pi_{N_2}.$$

□

Theorem 3. *The platform's profit is maximized at a strictly positive membership fee $\tau_P > 0$. The social welfare is maximized at $\tau_S = 0$.*

Proof of Theorem 3. Fixing the equilibrium number of sellers and buyers can be solved as functions of the platform's choice variable τ, denoted as $J^*(\tau)$ and $I^*(\tau)$, respectively. Given $\boldsymbol{\sigma}$, when τ increases, both $J^*(\tau)$ and $I^*(\tau)$ decrease in τ. Suppose the platform does not incur any costs in operation. The platform chooses the membership fee τ to maximize its profit

$$T(\tau) = \tau \times J^*(\tau) = \tau \times \mathcal{J}(\mathcal{I}(J, \tau), \tau).$$

The identity $\eta(J, \tau) = J - \mathcal{J}(\mathcal{I}(J, \tau), \tau)$ determines the implicit function $J^*(\tau)$. The derivative,

$$\frac{dJ}{d\tau} = -\frac{\frac{\partial \eta}{\partial \tau}}{\frac{\partial \eta}{\partial J}} = \frac{\frac{\partial \mathcal{J}}{\partial I}\frac{\partial \mathcal{I}}{\partial \tau} + \frac{\partial \mathcal{J}}{\partial \tau}}{1 - \frac{\partial \mathcal{J}}{\partial I}\frac{\partial \mathcal{I}}{\partial J}},$$

is a finite quantity. The derivative of the platform's profit at $\tau = 0$ is

$$\lim_{\tau \to 0} \frac{dT}{d\tau} = \left[J^*(0) + \tau \times \frac{dJ^*}{d\tau} \right]_{\tau=0} = J^*(0) + 0 > 0.$$

Because $T(\tau)$ is a continuous function on the compact interval $\tau \in [0, \tilde{\tau}]$, $T(0) = 0$, $T(\tilde{\tau}) = 0$, and $\lim_{\tau \to 0} \frac{dT}{d\tau} > 0$, we can conclude that there exists some $\tau_P \in (0, \tilde{\tau})$ that maximize $T(\tau)$.

The social welfare $SW = I \cdot E_N \left[E \left[W_{(1:N)} \right] \right]$ increases in I and N. Because $J^*(\tau)$ and $I^*(\tau)$ decrease in τ, by Theorem 2, $\tau_S = 0$ maximizes social welfare. □

Table 10: First-stage Results of TSLS Regressions in Table 7

	Dependent variable: *weight.position*			
	(3)	(4)	(7)	(8)
lag.weight.position	0.433***	0.441***	0.389***	0.403***
	(0.003)	(0.003)	(0.004)	(0.004)
position.other.market			0.142***	0.123***
			(0.004)	(0.004)
rating	−18.862***	−15.833***	−21.814***	−20.095***
	(1.221)	(1.344)	(1.556)	(1.735)
N.rating	−0.006***	−0.006***	−0.005***	−0.004***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
N.brand.stores	−0.094***	−0.063***	−0.087***	−0.061***
	(0.008)	(0.010)	(0.008)	(0.011)
N.brand.stores.plat	0.105***	0.039	0.132***	0.081***
	(0.019)	(0.027)	(0.020)	(0.028)
J	0.104***	0.085***	0.122***	0.099***
	(0.004)	(0.004)	(0.005)	(0.005)
activeness	16.566***	17.104***	16.986***	17.213***
	(0.427)	(0.462)	(0.526)	(0.574)
Ele.me	17.353***	13.050***	24.653***	19.622***
	(1.464)	(1.576)	(1.891)	(2.020)
Meituan	34.840***	30.603***	47.064***	41.057***
	(1.695)	(1.804)	(2.136)	(2.274)
deliv.min.p		0.066***		0.087***
		(0.023)		(0.030)
deliv.fee		0.873***		0.678***
		(0.089)		(0.107)
deliv.time		0.691***		0.597***
		(0.041)		(0.049)
category FE	Y	Y	Y	Y
station FE	Y	Y	Y	Y
Observations	88,881	77,310	61,277	52,793
R^2	0.548	0.540	0.572	0.558