Do Birds of A Feather Flock Together?
Platform's Quality Screening and End-Users' Choices
Theory and Empirical Study of Online Trading Platforms

Jin Wang*

Job Market Paper

This version: January 10, 2015

Abstract

In two-sided markets, consumers care not only about the number of sellers with which they can interact, but also about the quality of products or services these sellers provide. Previous work on two-sided markets has mainly focused on the quantity externality; Little attention has been paid to the quality externality. This paper proposes a model in which both quality and quantity play a role in the end-users' interactions, and analyzes how platforms can use quality screening to alleviate the problem of asymmetric information of quality and motivate end-users' participations. I address the question from theoretical and empirical perspectives. In the theory, I build a model in which two platforms compete but only one of them screens sellers' products. I show that the quality screening influences consumers' expectations of product quality and their choice of sellers and platforms. The resulting screening effect, together with the network and competition effects, further drives sellers to enter different platforms. Comparative static analysis indicates that sellers' incentives to join the platform that screens products follow an inverted-U curve with respect to the observable quality of products. In the empirical study, I transfer the theory to a simultaneous entry game with incomplete information, and carry out the estimation using the Nested Pseudo Likelihood (NPL) estimator with the data from Tmall and Taobao, two online trading platforms operated by Alibaba. The estimation results are consistent with the theory. The counterfactual analysis suggests that quality screening increases the consumers' utility and also improves Alibaba's market share.

Keywords: two-sided markets, asymmetric information, quality screening, duopoly platform competition, structural estimation, Alibaba

JEL: D43, L13, L14, L15

---

* Email: jinw@usc.edu, address: Department of Economics, University of Southern California, 3620 S. Vermont Ave., KAP 300, Los Angeles, CA 90089. I am very grateful to Guofu Tan and Geert Ridder for their enormously valuable advice and guidance. I also thank Harrison Cheng, Cheng Chou, Anthony Dukes, Yu-Wei Hsieh, Jeffrey Nugent, John Strauss, Simon Wilkie, Haojun Yu and participants in Econometrics Workshop at USC and Western Economic Association International Conference for useful discussions and comments. All the errors are mine.

1 Introduction

Many firms like eBay do their business by acting as an intermediary, or platform, which enables interactions between two groups of users: sellers and buyers. Through this or other platforms, end-users on the two sides trade products or services and realize their cross-group externality: the utility of users on one side depends on the performance of users on the other side. The platform operator profits by collecting usage fees from end-users onboard. In this economy, both end-users' benefits and the platform's profits depend heavily on how well the platform can help users to exploit their cross-group externality. This externality can involve end-users' interdependence in two dimensions: quality and quantity. For instance, when a consumer visits an online trading platform, she cares not only about the number of sellers available to choose from but also about the quality of products these sellers provide.

Previous studies of two-sided markets have mainly focused on the quantity externality: a user's utility is purely determined by the number of users on the other side. Less attention has been paid to end-users' quality concerns. When users evaluate the benefit of participating in a platform, they often take into account the quality of service and product exchanged in the interaction. If users know they may have interactions of poor quality, they will likely reduce the usage of the platform. The situation worsens if there is asymmetric information in quality between users on two sides. Following the classical economic theory since Akerlof (1970), asymmetric information can result in end-users' adverse selections and impede interactions which could have benefited users on both sides. Given the emphasis users place on quality, platform providers have a strong incentive to manage end-users' quality.

In this paper, I propose a model in which both quality and quantity play a role in the end-users' interactions, and analyze how platforms can use quality screening to alleviate the asymmetric information problem and motivate end-users' participation. Quality screening has a popular application in two-sided platforms: Prestigious shopping malls carefully select the brands that are going to enter the mall. Many academic journals maintain good reputations and large readership because they have a rigorous refereeing process. In spite of the broad use of quality screening in two-sided markets, there have been few studies to investigate the economic implications of screening for players' choices and welfare. For instance, does screening benefit platform providers? How does quality screening affect end-users' interactions? If there are two platforms with different quality screening standards, which platform will users choose? My study will address these questions.

I study quality screening in the context of online trading platforms, serving two types of end-users: sellers and buyers. Within a platform, sellers compete in selling a particular product to consumers. The quality of products is heterogeneous and composed of two parts: the quality observable to all players and the private quality signal known only by the seller. To overcome the information asymmetry, the platform can implement quality screening by charging sellers different usage fees depending on their products' quality. More specifically, the platform collects each seller's quality information by randomly sampling products or through consumer feedback. The quality signal which the platform obtains mixes the true product quality plus a noise term. If the value of the signal lies above a threshold, the platform offers the seller a discount on the usage fee. Otherwise, the platform charges the seller a high usage fee and reduces the seller's profit to zero. Under this rule, the higher the product quality is, the more likely the seller will enjoy a lower platform usage fee.

To investigate the effect of quality screening on buyers' and sellers' entry decisions, I assume two platforms in the model. One platform employs quality screening as described above and I denote it as Platform S. The other platform is a free-entry market and refers to Platform NS. Both buyers and sellers are single-homing. The model illustrates that the number of consumers on a platform depends on the total utility provided by the entire seller group on a platform. That is, the platform which gathers more sellers of better quality can seize a larger market share on the consumer side. On the seller side, a merchant's location decision is governed by three effects: screening, network and competition. First, screening effect means that sellers can convince consumers that their products are of good quality, by showing their willingness to pay a usage fee correlated with quality. Entering Platform S enables a seller to signal the unobserved quality of the product and increase the chances of sales. The network effect and competition effect characterize two ways that a seller's participation decision is influenced by the number of consumers on the platform. On one hand, a merchant wants to enter the platform with a larger number of consumers. By interacting with more consumers, the merchant can sell more products. On the other hand, the number of consumers on a platform is positively correlated with the utility provided by the entire seller group. So for a seller, joining a platform which has more consumers means that it has to face fiercer competition. This may reduce this seller's market share. The relative magnitude of two effects determines how a seller responds to the consumers' participation. If network effect dominates, a seller is willing to join the platform that attracts more consumers. Otherwise, the seller would rather avoid peer competition by choosing the platform which has fewer consumers.

Under the influence of the screening effect, the proportion of sellers choosing Platform S first increases with the observable quality. Once the observable quality reaches a certain point, the network and competition effects take over and may reduce the proportion of sellers on Platform S. The intuition works as follows: For a merchant that sells a product of low or medium observable quality, entering Platform S can signal the product's unobserved quality and increase sales. But this may also bring the risk of a high usage fee if the product fails Platform S's quality screening. When the observable quality increases, the product is more likely to meet the standards of Platform S and therefore the sellers will be more willing to enter this platform. If the product is of sufficiently high observable quality, a seller will not worry about passing the quality screening. The unobserved quality of the product can not be inferred by consumers regardless of the platform the product is located on. So the screening effect vanishes and only the competition and network effects matter in the sellers' decisions. The seller chooses its desired platform by weighing the comparative magnitude of both effects.

To test the theory, I apply the two-sided market model to the empirical study. By doing this, I am able to estimate the end-users' responses to quality screening and quantify the effect of screening on the profits of the platform provider. I select Alibaba because of both its size and business strategy. Firstly, Alibaba is a monster E-commerce firm. It has so far captured over 90% of the online market in China. In 2013, its total sales were $248 billion dollars – more than eBay and Amazon combined. Second, Alibaba operates two online-trading platforms: Taobao and Tmall. Taobao is a free-entry market where any product can be posted for sale, while Tmall sets a quality standard and charges sellers usage fees depending on their quality. Therefore, Taobao and Tmall are exactly the counterparts of Platform NS and Platform S in the theory, and their business models fit the setup of my theory well.

The data is collected from these two websites. It contains sellers' information including each seller's rating score, price, monthly sales and location (Tmall or Tabao). The variable rating score is treated as the proxy of the seller's observable quality. In the data, as the rating score increases, the proportion of sellers that choose Tmall first increases then decreases, which is consistent with the prediction of the theory.

To proceed with the estimation, I show that the two-sided market model can be equally formalized as a game in which sellers simultaneously choose their desired platform with incomplete information about other sellers' quality. Unlike previous entry games which mostly assume that sellers have abstract profit functions, my two-sided market model provides a well-defined profit function which incorporates consumers' belief of a seller's quality conditional on the seller's location choice, rival sellers' location choices probabilities, and the seller's entry cost. Sellers' equilibrium choices in the two-sided market model are readily transformed into a set of Bayesian Nash equilibrium choice probabilities. These probabilities are a fixed point, which is determined by the mapping from a seller's conjecture of competitors' choices to its competitors' conjectures of this seller's choice. And they are to be estimated together with the seller's profit function.

The estimation is carried out using the Nested Pseudo Likelihood (NPL) estimator proposed by Aguirregabiria and Mira (2007). In the demand function, the total utility provided by the entire seller group has a parameter of positive sign. It suggests the network effect dominates the competition effect and sellers benefit from pooling with rivals of good quality. The impact of the expected unobservable quality on a seller's demand varies by seller types. For sellers with low or medium rating scores, the estimator is positive and significant, while for sellers with high rating scores, the estimator is negative and close to zero. This supports my theory that in equilibrium the screening effect only takes effect on sellers of low and medium observable quality. These types of sellers can convince consumers that their products are of good quality if they submit to the quality screening of Tmall. In the entry cost function, the estimation results indicate that a seller of a higher rating score or higher unobservable quality pays a lower usage fee on Tmall, which coincides with the spirit of quality screening.

Last, I conduct a counterfactual analysis to investigate whether the strategy of market separation and quality screening can improve consumers' welfare and also help Alibaba to achieve more advantageous position when it competes with rival firms. I construct a variable which compares the consumer's expected utility from all sellers located on Alibaba before and after Alibaba would remove quality screening. I find that the total utility provided by the entire seller group gets increased with the presence of quality screening. This also means, by employing quality screening, Alibaba to some extent improves its market position.

Literature Review. This paper contributes to the literature of two-sided markets and network externalities (Katz and Shapiro (1985, 1986); Laffont et al. (1998a,b); Rochet and Tirole (2002, 2006); Caillaud and Jullien (2003); Armstrong (2006); Weyl (2010)). The two-sided markets are featured by the interactions between two groups of agents with cross-group externalities. To make profits, platforms need to solve the so-called "Chicken-and-Egg" problem and "get both sides on board" (Caillaud and Jullien (2003); Rochet and Tirole (2003); Rysman (2009)). That is, when platforms design their business strategies, they should consider the responses of agents on both sides. Existing literature studies the mechanism design on two-sided markets from various perspectives, which include the pricing schemes and price allocation on both sides (Rochet and Tirole (2003, 2006); Armstrong (2006)), platforms' price commitments (Hagiu (2006)), price discrimination and bundling (Damiano and Li (2007); Chao and Derdenger (2013)), exclusive contracts (Armstrong and Wright (2007)), platform or end-users coordination (Rochet and Tirole (2002); Ambrus and Argenziano (2004)) and so on. These studies, although addressing different questions on two sided markets, usually focus on the quantity externality: the utility of a user on one side depends only on the number of users on the other side. There are papers which allow end-users have heterogeneous benefits from interactions, but this heterogeneity in their models is assumed pre-determined and does not change with platform's decision. The importance of quality in end-users' interactions is rarely mentioned. My study proposes a model which captures end-users' interdependence on both quality and quantity. In this framework, I study how a platform can use quality screening to alleviate the problem of asymmetric information between end-users and prompt end-users' interactions and the platform's profit

This paper also joins the empirical study of two-sided markets. Some previous work has identified the network effect between end-users in different two-sided markets (Rysman (2004, 2007); Ackerberg and Gowrisankaran (2006); Argentesi and Filistrucchi (2007)). Lee (2013) develops a dynamic model to study consumers' purchases of hardware and software, and software providers' adoption on hardware, and the welfare implication if the integration and exclusive contract between hardware and software is prohibited. Zhou (2013) proposes a new method to estimate consumers' and software providers' decisions and points out the hardware firms' pricing leverages on the two sides are important for the platforms' launch success. In the second part of my paper, I build a structural model to estimate consumers' choices and sellers' entry decisions. When they make decision, both consumers and sellers take into account the impact of quality screening on their utility and profits. My work identifies the network, competition and screening effects in the interactions between sellers and buyers.

This paper is also closely related to the literature of the estimation of discrete choice games and its applications. This stream of literature starts from the seminal work by Bresnahan and Reiss (1990, 1991) and Berry (1992). They analyze firms' strategic entry decisions in the framework of a discrete choice game. Stavins (1995), Mazzeo (2002) and Toivanen and Waterson (2005) adopt the same framework and study firms' entry decisions in different scenarios. All these papers perform their work under the assumption that firms possess complete information of rivals' characteristics. During the estimation process, researchers have to check every firm's equilibrium conditions, which increases the computation burden, especially when the number of firms and their alternative choices are very large. In many applications, firms do not completely know other firms' decision variables. This makes the incomplete-information structure a more favorable choice. Under the structure of incomplete information, firms' equilibrium choices can be formalized as a set of Bayesian Nash equilibrium beliefs and the estimation gets easier (Rust (1996)). Estimators proposed by Hotz and Miller (1993), Aguirregabiria and Mira (2002, 2007), Pesendorfer and Schmidt-Dengler (2008) and Pakes et al. (2007) are easily implemented in games with a large number of players or alternative choices. The discrete game with imperfect information has many applications including the empirical study of firms' entry and special competition (Seim (2006); Zhu and Singh (2009); Vitorino (2012)) and social interactions (Brock and Durlauf (2001)). In my paper, I show that the theoretical model can be equally transferred to a discrete choice game where sellers simultaneously enter different platforms with imperfect information of rivals' quality. Unlike previous studies of entry games which assume abstract profit functions for entrants, I derive sellers' profit functions from the theoretical model which captures the interactions between sellers and buyers. Estimation is carried out by using NPL estimator proposed by Aguirregabiria and Mira (2007).

The rest of the paper is organized as follows. Section 2 presents the theoretical model and derives the equilibrium choices of game players. Section 3 transfers the theoretical model to a simultaneous entry game with incomplete information and discusses the estimation strategy. Section 4 firstly introduces the background of Alibaba and its two online trading platforms: Tmall and Taobao, discusses the data, and presents the results of estimation and counterfactual analysis. Section 5 concludes the paper.

2 Theoretical Model

2.1 Players and Payoffs

I study two-sided online trading markets. It is easy to generalize the results of my model to other two-sided platforms which use quality screening to regulate agents' entry and interaction. There are three types of players in the game: sellers, buyers and the platform provider. I will discuss their characteristics in this subsection:

2.1.1 Product Quality

There are N ($N >> 0$) sellers in the model. They compete for selling the same product to consumers. Sellers are heterogeneous in their product's quality. Denote the quality of products sold by a seller, j, as q_j. It is composed of two parts:

$$q_j = \mu_j + \theta_j \tag{1}$$

where μ_j is the observable quality of the product. Think about a merchant that sells a camera, say Nikon D3300, μ_j can be the camera's specs, the seller's reputation, or other observable characteristics. u_j is fixed and perfectly observed by all players. I do not consider the case in which the seller can manipulate or hide the information of μ_j to cheat consumers and platforms. μ_j satisfies Assumption 1.

Assumption 1 (Observable Quality)

The observable quality μ satisfies:

(i) $\mu \in (-\infty, +\infty)$.

(ii) Among the N sellers, the number of sellers with $\mu = x$ is N_x.

The second component, θ_j, is the private information about the quality of the product. θ_j is known only by seller j. Back to the previous example, the camera could be a fake or refurbished product, or its flash button does not function very well. This information is concealed by the seller for its own interest. Platforms and consumers have a prior on the value of θ_i, which is drawn from a distribution whose properties are common knowledge and satisfy Assumption 2.

Assumption 2 (Independent Symmetric Private Signal)

The distribution of θ satisfies:

(i). θ is independent with μ.

(ii). $\theta \sim f_\theta$ where f_θ is a log-concave continuous function on the closed interval $[\underline{\theta}, \overline{\theta}]$.

(iii). θ has a mean of zero: $\mathrm{E}(\theta) = 0$.

μ and θ are assumed to be independent with each other. It means that platforms and consumers are not able to completely infer the value of the private quality signal from the observable characteristic. My model can allow a random correlation between μ and θ, which does not change the main results. The assumption of log-concave distribution is made for the comparative static analysis. The family of log-concave probability distributions has wide applications in economics (Heckman and Honore (1990); An (1996); Bagnoli and Bergstrom (2005)). It includes many commonly-used distributions such as normal distribution and exponential distribution. The zero-mean assumption is made for normalization.

2.1.2 Platforms and Quality Screening

In the model, there are two platforms, which are denoted as $\{S, NS\}$ perceptively. My model assumes that two platforms are operated by the same company, but differentiate in their usage fees and quality screening policies.

Platform S and quality screening. Platform S screens sellers' products for quality and charges sellers different usage fees based on the screening results. Specifically, the platform employs quality screening through monitoring the quality of all products sold on the platform. It collects the quality information from three resources: consumers' feedbacks, the quality inspection by the official department, or the random sampling test by Platform S itself. First, Platform S can encourage consumers to provide feedbacks about the products they bought from the platform and learn the quality. Second, when the Commercial Administration Department of the government inspects sellers' products, the inspection results can also inform Platform S of the quality information. Last, Platform S can ask sellers to submit sample products and check their quality.

Through above three channels, Platform S obtains a noisy quality signal $\hat{q}_j$ for the product sold by seller j:

$$\begin{aligned}\hat{q}_j &= q_j + \varepsilon_j \\ &= \mu_j + \theta_j + \varepsilon_j \end{aligned} \tag{2}$$

in which q_j is seller j's true quality and ε_j represents a random shock to the quality of the product. For instance, back to the previous example of cameras, ε_j can be understood as the uncontrollable factors taking place during a camera's production, storage or delivery, which influences the quality judgment or evaluation by consumers, the administration department or the platform. In the model, ε_j is assumed to be a realization of a random variable ε, the distribution of which is common knowledge to all players.

Assumption 3 (Independent Quality Shock)

The distribution of ε satisfies

(i) ε is independent of μ and θ;

(ii) $\varepsilon \sim f_\varepsilon$ and f_ε is a continuous function on $[\underline{\varepsilon}, \bar{\varepsilon}]$.

Based on the quality signal $\hat{q}_j$, Platform S charges sellers different usage fees. The usage fee is assumed to be proportional to the seller's total transaction value, and the proportion $\hat{t}_S$ is determined by

$$\hat{t}_S = \begin{cases} t_S & \text{if } \hat{q}_j \geq k_S \\ 1 & \text{if } \hat{q}_j < k_S \end{cases}$$

where $t_S \in (0,1)$ and k_S represents a quality standard which is set by Platform S and announced to all other players. This formula illustrates that the proportional fee imposed on seller j depends on the quality signal $\hat{q}_j$. If $\hat{q}_j$ exceeds the quality standard k_S, Platform S only takes away t_S percent of the revenue of seller j. If $\hat{q}_j$ lies below k_S, seller j's product fails in the quality screening, and as a punishment, the seller loses all the transaction revenue.

To sum up, Platform S screens sellers' products and charges its sellers transaction fees which depend on whether the sellers can pass the screening.

Platform NS: Platform NS charges sellers a uniform proportional fee $t_{NS} \in [0,1)$. This is equivalent to the case that Platform NS implements quality screening, but sets the standard $k_{NS} = -\infty$. And eventually every seller can pass the screening.

2.1.3 Consumer's Decisions

The measure of potential consumers is normalized to be one. Consumers are homogeneous and single-homing. They make independent entry and purchase decisions which can be summarized as two steps. Firstly, the consumer decides which marketplace it visits among three options: Platform S, Platform NS, and the outside option which is denoted as 0. After choosing the market, the consumer makes the second-step decision: it evaluates the utility obtained from the sellers on that marketplace and purchases one unit of the product through the seller which offers the consumer the largest utility. Next, I model the consumer's two-step decision in a backward sequence: I first discuss the purchase decision and then analyze the entry.

2.1.4 Consumer's Purchase Decision

Let $\mathcal{J}(m)$ denote the set of sellers on the platform m where $m \in \{S,\ NS\}$, consumer i obtains a random utility $\widetilde{u}_{i,j,m}$ through purchasing the product from the seller $j \in \mathcal{J}(m)$, where $\widetilde{u}_{i,j,m}$ takes a linear form:

$$\widetilde{u}_{i,j,m} = \mathbb{E}(q_j|\mu_j,\ k_m) - price_j + \epsilon_{i,j,m} \tag{3}$$

where $\mathbb{E}(q_j|\mu_j,\ k_m)$ represents the consumer's belief about the seller's product quality conditional on the product's observable quality μ_j and the platform's quality standard, k_m. And $price_j$ is the price of the product. $\{\epsilon_{i,j,m} : \text{all } j \in \mathcal{J}(m)\}$ are i.i.d seller-specific random utility shocks following Type-I extreme value distribution $f_\epsilon(0,1,0)$, which are independent of μ and θ .

Define $u_{j,m} = \mathbb{E}(q_j|\mu_j,\ k_m) - price_j$ as the consumer's expected utility of shopping from the seller $j \in \mathcal{J}(m)$. The probability that the consumer i buys from seller j can be expressed as:

$$\begin{aligned} d_{j|j\in\mathcal{J}(m)} &= \Pr(\widetilde{u}_{j,m} \geq \widetilde{u}_{j',m},\ \text{all } j' \in \mathcal{J}(m)) \\ &= \frac{\exp\left[u_{j,m}\right]}{\sum_{\text{all } j' \in \mathcal{J}(m)} \exp\left[u_{j',m}\right]} \end{aligned} \tag{4}$$

. It shows that within a platform, a merchant which provides higher utility can attract more consumers. The competition among sellers becomes intensified as the total utility that consumers obtain from the entire seller group increases.

2.1.5 Consumer's Entry Decision

Consumer i's utility of patronizing platform $m: \ m \in \{S, \ NS\}$ equals to:

$$\widetilde{\nu}_{i,m} = \lambda EU(m) + \eta_{i,m}$$

where $EU(m)$ stands for the expected maximum utility offered by sellers located on platform m. Since $\epsilon_{i,j,m}$ follows Type-I extreme value distribution, according to Rust (1987), the expected maximum utility has a closed form:

$$\begin{aligned} EU(m) &= \mathbb{E}\left[\max \mu_{j',m} \middle| \text{all } j' \in \mathcal{J}(m)\right] \\ &= \ln\left(\sum_{j' \in \mathcal{J}(m)} \exp\left[u_{j',m}\right]\right) \end{aligned} \tag{5}$$

As for the expected maximum utility from the outside option, without loss of generality, I normalize it to be 0, i.e., $EU(0) = 0$. Therefore the consumer's utility of choosing the outside option equals to

$$\tilde{\nu}_0 = 0 + \eta_0$$

. $\{\eta_S, \eta_{NS}, \eta_0\}$ represent the consumer's idiosyncratic preference for the three markets. $\{\eta_S, \eta_{NS}, \eta_0\}$ are randomly drawn from Type-I extreme value distribution $f_\eta(0, \frac{1}{\lambda}, 0)$ where $\frac{1}{\lambda} > 0$ is the scale of the distribution. η is assumed to be independent of $EU(m)$.

The probability that consumer i goes to platform $m: \ m \in \{S, NS\}$ is

$$d_m = \frac{\exp[\lambda EU_m]}{1+\sum_{m' \in \{S,NS\}} \exp[\lambda EU_{m'}]} \tag{6}$$

Notice that d_m is the market share of platform m in the whole market composed by Platform S, Platform NS and the outside option. And $d_{j|j \in \mathcal{J}(m)}$ is the probability that the consumer purchases from seller j provided that the consumer has decided to patronize platform m. Using these two probabilities, I can derive the unconditional market share of seller j in the whole market:

$$\begin{aligned} d_{j,m} &= d_{j|j \in \mathcal{J}(m)} d_m \\ &= \frac{\exp[u_{j,m}]}{\sum_{all \ j' \in \mathcal{J}(m)} \exp[u_{j',m}]} \frac{\exp\left[\lambda EU_m\right]}{1 + \sum_{m' \in \{S,NS\}} \exp\left[\lambda EU_{m'}\right]} \\ &= \exp[u_{j,m}] \frac{\exp\left[(\lambda - 1) EU_m\right]}{1 + \sum_{m' \in \{S,NS\}} \exp\left[\lambda EU_{m'}\right]} \end{aligned} \tag{7}$$

. The consumer's expected maximum utility on platform m, $EU(m)$, influences seller j's market share on platform m, $d_{j,m}$, in two opposing ways: First, since d_m is positively correlated with $EU(m)$. When $EU(m)$ increases, there are more consumers to patronize platform m. By interacting with these consumers, sellers can increase the sales. I call this effect the network effect. On the other wide, as shown by (5), $EU(m)$ incorporates the utility provided by all sellers on platform m. A bigger $EU(m)$ means the seller faces fiercer competition from rivals which can decrease the seller's market share. I call this effect the competition effect. In (7), the magnitudes of the network and competition effect are measured by λ and 1 respectively. If $\lambda > 1$, the network effect dominates the competition effect and the seller prefers to participate in a platform which hosts more consumers. Otherwise, the seller would rather avoid peer competition by attending a platform with a smaller number of consumers.

2.1.6 Seller's Profit and Entry Decision

The profit of seller j on platform m can be expressed as

$$\pi_{j,m}(\mu_j, \theta_j;\, k_m, t_m) = (\bar{p} * d_{j,m}) * (1 - t_m) * \Pr(\varepsilon_j \geq k_m - \mu_j - \theta_j) \tag{8}$$

where$(\bar{p} * d_{j,m})$ represents the seller's revenue on platform m and $(1 - t_m)$ is the percent of revenue the seller keeps after paying the proportional fee to the platform. The last term $\Pr(\varepsilon_j \geq k_m - \mu_j - \theta_j)$ measures the effect of quality screening on the seller's profit. The seller makes zero profit if its product does not pass the quality screening.

The seller is assumed to be single-homing and joins the platform in which it can earn a larger profit. The total number of sellers on the platform m equals to

$$N_m = \sum_{j=1\ldots N} 1[\pi_{jm}(\mu_j, \theta_j;\, k_m, t_m) \geq \pi_{jm'}(\mu_j, \theta_j;\, k_{m'}, t_m)]$$

2.2 Sequence of the Game

The game proceeds in the following sequence:

Period 0: The nature determines the quality of the product sold by seller j, $q_j = u_j + \theta_j$, where μ_j is observable to all players. θ_j is the private information known only by the seller j.

Period 1: Platform $m : m \in \{S, NS\}$ announces the proportional fee, t_m, and the quality standard, k_m.

Period 2: Sellers simultaneously choose which platform to participate in.

Period 3: Consumers enter their desired platforms and make purchasing decisions.

2.3 Equilibrium Analysis

This section analyzes the equilibrium strategies of sellers and consumers. Following the sequence of the game, the seller makes decision at Period 2. Its strategy is to choose Platform S or Platform NS, in anticipation of the consumer's belief of its product quality and the product's market share on each platform. The seller's decision depends on several factors: the observable quality, μ, the private quality signal, θ, two platforms' price menu $\{t_S, t_{NS}\}$ and the screening standards $\{k_S, k_{NS}\}$, and the expected number of consumers on two platforms $\{d_S, d_{NS}\}$. At Period 3, consumers take actions. They patronize their favorite platforms and sellers, based on the distribution of sellers on the two platforms and the expected quality of products. In equilibrium, consumers' belief of the private quality signal θ should be consistent with the seller's choice.

Proposition 1. (Equilibrium)

Given two platforms' policy variables $\{t_m,\ k_m | m \in \{S,\ NS\}\}$, there always exists an equilibrium such that

Consumers. Consumers participate in Platform S and Platform NS with probabilities$\{d_S^*, d_{NS}^*\}$. And the consumers' belief of θ can be expressed as

(i) When $\mu < k_S - \underline{\theta} - \underline{\varepsilon}$,

$$E(\theta|\mu, k_S) = E(\theta|\theta \geq \theta^*(\mu))$$

and

$$E(\theta|\mu, k_{NS}) = E(\theta|\theta \leq \theta^*(\mu))$$

(ii) When $\mu \geq k_S - \underline{\theta} - \underline{\varepsilon}$,

$$E(\theta|\mu, k_S) = E(\theta|\mu, k_{NS}) = 0$$

Sellers. Sellers take the strategy as follows:

(i) When $\mu < k_S - \underline{\theta} - \underline{\varepsilon}$, sellers enter Platform S if $\theta \geq \theta^*(\mu)$ and participate Platform NS if $\theta < \theta^*(\mu)$, where $\theta^*(\mu)$ is determined by:

$$0 = \ln \frac{1 - t_S}{1 - t_{NS}} + \frac{(\lambda - 1)}{\lambda} \ln \frac{d_S^*}{d_{NS}^*} + \mathbb{E}(\theta|\theta \geq \theta^*(\mu)) - \mathbb{E}(\theta|\theta \leq \theta^*(\mu)) + \ln(1 - F_\varepsilon(k_S - \mu - \theta^*(\mu)))$$

(ii) When $\mu \geq k_S - \underline{\theta} - \underline{\varepsilon}$, sellers always enter Platform S if

$$\ln \frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda} \ln \frac{d_S^*}{d_{NS}^*} > 0$$

, and they choose Platform NS otherwise.

Proof: see in the appendix.

I sketch the equilibrium here, leaving the complete proof in the appendix. Following previous analysis, the seller makes the entry decision by comparing the profits on the two platforms, which can be expressed as :

$$\begin{aligned} \ln \pi_S - \ln \pi_{NS} &= \ln \frac{1-t_S}{1-t_{NS}} \qquad (\triangle\text{proportional fee}) \\ &+ \frac{(\lambda-1)}{\lambda} \ln \frac{d_S}{d_{NS}} \qquad (\triangle\text{network v.s. competition effects}) \\ &+ \mathbb{E}(\theta|\mu,\, k_S) - \mathbb{E}(\theta|\mu,\, k_{NS}) \qquad (\triangle\text{screening effect}) \\ &+ \ln(1 - F_\varepsilon(k_S - \mu - \theta)) \qquad (\triangle\text{screening cost}) \end{aligned} \tag{9}$$

. This formula shows that the seller evaluates the payoffs on two platforms from four aspects: the proportional fees, the network versus competition effects, the screening effects and screening costs. First, the seller compares the proportional fees charged by the two platforms, because they determine the seller's net revenue. Second, sellers are also concerned about the network and competition effects. As discussed in Section 2.1.5, these two effects characterize the impact of the number of consumers on a seller's profit. This impact is realized through the coordination and competition among sellers on the same platform. When a platform has more sellers of good quality, it influences the profit of individual seller in two directions: On one hand, this attracts more consumers to the platform and prompts each seller's sales. On the other hand, it intensifies the competition between sellers and deteriorates individual seller's profit. If $(\lambda - 1)$ is positive, the network effect becomes dominant and the seller wants to join the platform which have more consumers. Otherwise, the seller favors the platform with fewer consumers. The screening effect arises from the fact that consumers update their belief of sellers' unobservable quality according to sellers' participation decisions. Such an expectation should be higher for sellers on Platform S which adopts a higher quality standard. Although a seller can enter Platform S to signal its quality, it has to bear the risk that it may fail in the quality screening and earn zero profit. This screening cost can impede the seller's incentive to join Platform S.

In the four factors, the proportional fees and the network and competition effects are homogeneous among all types of sellers, while the magnitude of the screening effect and the screening cost vary with the quality of the product. To be specific, the screening effect depends on the observable quality μ and the platform the seller is located on. The screening cost depends on both the observable μ and the private quality signals θ. Sellers that hold the private quality information weigh the benefits and costs of screening and are self selected onto the different platforms. Anticipating the correlation between quality and seller's entry decision, the consumer rationalizes the belief of θ. That's how the quality screening mechanism works.

When the observable quality μ is less than $\mu < k_S - \underline{\theta} - \underline{\varepsilon}$, there is positive probability that the product fails in quality screening. In this case, by submitting to the quality screening of platform S, the seller can convince consumers the product is of good quality. Since the screening cost decreases with θ, the seller chooses Platform S if its private signal θ is large enough. And the threshold of θ, in equilibrium, is correctly inferred by the consumer.

However, the effect of screening vanishes when the observable quality μ is sufficiently high such that $\mu \geq k_S - \underline{\theta} - \underline{\varepsilon}$. In this case, the seller always passes the quality screening, which suggests θ is no longer a factor that plays a role in the seller's entry decision. Consumers are not able to update their belief of θ according to the seller's participation decision. Correspondingly, on both platforms, consumers adjust the belief of θ to the population mean:

$$\mathbb{E}(\theta|\mu, k_S) = \mathbb{E}(\theta|\mu, k_{NS}) = 0$$

. Therefore the seller's entry decision only relies on the proportional fees and the network effects on the two platforms.

Proposition 2. (Comparative Statics)

Suppose $\theta \sim f_\theta$ where f_θ is log-concave, and $\varepsilon \sim Exp(l)$ where $l \geq 1$.

(i) When $\mu < k_S - \underline{\theta}$, $\theta^*(\mu)$ is a decreasing function of μ.

(ii) When $\mu \geq k_S - \underline{\theta}$,

$$\theta^*(\mu) = \begin{cases} \underline{\theta} & \text{if} \ln \frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda} \ln \frac{d_S^*}{d_{NS}^*} \geq 0 \\ \bar{\theta} & \text{if} \ln \frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda} \ln \frac{d_S^*}{d_{NS}^*} < 0 \end{cases}$$

Proof: see in the appendix.

This proposition illustrates there may exist a non-monotonic relationship between the observable quality and the seller's entry decision. When the observable quality is less than $k_S - \underline{\theta}$, a seller wants to join Platform S to take the advantage of the screening effect. At the same time, it needs to pay the cost associated with the risk of failing the screening test. Under the distribution assumptions in Proposition 2, the benefit of taking quality screening outpaces the cost such that the net benefit of taking quality screening increases with μ. Sellers that have higher observable quality are more likely to choose Platform S. Once the observable quality reaches $k_S - \underline{\theta}$, both the benefit and cost of screening go to zero. The network and competition effects take over and play a key role in determining a seller's choice of platform. The usage fees on the two platforms also matters in the seller's decision. The seller's own characteristics, either the observable quality or the private signal, do not have any impact on the seller's entry. Sellers now become identical and make symmetric entry decisions. They may choose to enter Platform NS, if this platform provides a larger network effect or a smaller competition effect. Therefore, under the influences of the screening, network and competition effects, a seller's incentive to participate in Platform S may follow an inverted-U curve.

So far I have characterized the interactions between sellers and consumers on the two platforms and analyzed their equilibrium choices. But there is still a question which remains to be answered: Does the quality screening benefit consumers and the platform operator? Answering this question requires us to make a series of assumptions on the variables in the model, which include: the number of potential sellers, each seller's observable quality and the private signal, and the prior of the private signal and the screening noise. In stead of making these assumptions and giving an abstract answer, I choose to investigate the effect of screening on consumers' welfare and platforms' profits using the data from the real business. In the next section, I will discuss how to apply the model to the data and estimate it.

3 Structural Model and Estimation Strategy

In this section, I utilize previous theory to build a structural model and discuss the estimation strategy. I first show that the two-sided market model specified in Section 2 is equivalent to a discrete choice game in which sellers simultaneously decide whether to enter platform m, holding incomplete information of other sellers' quality. Then I present the estimation procedure of this incomplete information game.

3.1 Model Specification

In order to apply the model in Section 2 to the data, I rewrite players' payoff functions with the deterministic characteristics and associated parameters. Suppose there are T independent products. For each product, there are a particular number of sellers and consumers that play the game as described in Section 2.

Consumers. I start from the consumer's preference and utility function. A representative consumer believes that the quality of product t sold by seller j takes following form:

$$q_j^t = X_j^t \alpha + \theta_j^t \beta, \tag{10}$$

where X_j^t is a vector of observable attributes of the seller, e.g. reputation, the dummy variable for free-shipping, return policies, et. al.; X_j^t is pre-determined and does not change with the seller's choice of platform. θ_j^t represents the private quality signal possessed by the seller j, for instance, the long-term reliability or the authenticity of the product. θ_j^t is assumed to be a single random variable whose distribution satisfies Assumption 2. The parameter α and β measure the weights of X_j^t and θ_j^t in the composition of the product's overall quality.

Given the quality specification, the consumer i's utility of purchasing from seller j located on platform m can be expressed as

$$\tilde{u}_{i,j,m}^t = u_{i,j,m}^t + \epsilon_{i,j,m}^t$$

in which the expected utility equals to

$$u_{j,m}^t = X_j^t \alpha + \mathbb{E}\left[\theta_j^t | X_j^t,\, m\right] \beta + price_j^t \gamma$$

,where γ reflects the influence of product price on the expected quality.

Following the analysis in Section 2, the choice probability of seller j in the whole market of product t in (7) can be explicitly written as

$$d_{j,m}^t \quad = \quad \exp[u_{j,m}] \frac{\exp\left[(\lambda - 1)\, EU_m\right]}{\sum_{\text{all } m' \in \Omega} \exp\left[\lambda EU_{m'}\right]} \tag{11}$$

with

$$\begin{aligned} EU_m &= \ln\left(\sum_{j'\in\mathcal{J}^t(m)} \exp\left[{u^t}_{j',m}\right]\right) \\ &= \ln\left(\sum_{j'\in\mathcal{J}^t(m)} \exp\left[X^t_{j'}\alpha + \mathbb{E}\left[\theta^t_{j'}|X^t_{j'},\, m\right]\beta + price^t_{j'}\gamma\right]\right) \end{aligned}$$

Sellers. The seller's profit of participating platform m in equation (8) can be equally summarized as

$$\pi^t_{j,m} = [d^t_{j,m} \times price_j] * \Gamma(X^t_j,\, \theta^t_j,\, m)]$$

The first term in bracket represents the seller's sales revenue which is a product of the seller's market share, $d^t_{j,m}$, defined in (11) and the seller's price, $price_j$. $\Gamma(X^t_j, \theta^t_j, m) \in (0,1)$ stands for the fraction of revenue the seller earns net of the screening cost and the platform usage fee . According to Section 2, it is a function of the seller's quality $(X^t_j,\, \theta^t_j)$ and the index of the platform. For sellers located on Platform S, $\Gamma(X^t_j,\, \theta^t_j,\, m)$ is positively correlated with the seller's quality $(X^t_j,\, \theta^t_j)$, while for sellers on Platform NS, $\Gamma(X^t_j,\, \theta^t_j,\, m)$ is independent of the seller's quality type as this platform has no quality regulation on its entrants.

The profit function can be transferred to a linear function by taking logarithm on both sides :

$$\ln \pi^t_{j,m} = \ln d^t_{j,m} + \ln price_j + \ln \Gamma(X^t_j,\, \theta^t_j,\, m)$$

. When the screening noise ε in (2) follows exponential distribution, $\ln \Gamma(X^t_j,\, \theta^t_j,\, m)$ also has a linear form:

$$\ln \Gamma(X^t_j,\, \theta^t_j,\, m) = \rho^m_0 + \rho^m_1 X^t_j + \rho^m_2 \theta^t_j$$

. Therefore, the difference of the seller's log-profits on the two platforms is:

$$\begin{aligned} \triangle\pi^t_j &\equiv \ln \pi^t_{j,S} - \ln \pi^t_{j,NS} \\ &= \ln d^t_{j,S} - \ln d^t_{j,NS} + \rho_0 + \rho_1 X^t_j + \rho_2 \theta^t_j \end{aligned} \tag{12}$$

where $\rho_k = \rho^S_k - \rho^{NS}_k$, $k \in \{0, 1, 2\}$. And according to the rule of quality screening, it can be expected that $\rho_1, \rho_2 > 0$.

With the specification of $d^t_{j,m}$ in (11), $\triangle\pi^t_j$ can be further decomposed as following:

$$\begin{aligned} \triangle\pi^t_j &= \ln\left(\sum_{j'\in\mathcal{J}^t(S)} \exp\left[{u^t}_{j',S}\right]\right) - \ln\left(\sum_{j'\in\mathcal{J}^t(NS)} \exp\left[{u^t}_{j',NS}\right]\right)(\lambda - 1) \\ &+ \left(\mathbb{E}\left[\theta^t_j \middle| X^t_j,\, m = S\right] - \mathbb{E}\left[\theta^t_j \middle| X^t_j,\, m = NS\right]\right)\beta \\ &+ \rho_0 + \rho_1 X^t_j + \rho_2 \theta^t_j \end{aligned} \tag{13}$$

. It indicates a seller's choice of platform is determined by three terms. The first is the number sellers located on each platform and the total utility consumers obtain from these sellers. The way this term affects the seller's profitability depends on the magnitude of network effect and competitive effect, which are correspondingly measured by λ and 1. When $\lambda > 1$, the network effect dominates and the seller favors the platform where sellers deliver higher total utility. When $\lambda < 1$, the competition effects dominates and the seller wants to soften the competition by joining the platform which gathers sellers with lower quality. The second term which influences the seller's decision is the consumer's conjectures of the seller's unobservable quality, conditional on the seller's observable attributes and participation of platform: $\mathbb{E}\left[\theta_j^t \middle| X_j^t,\, m\right]$. In equilibrium, these conjectures should coincide with the seller's optimal choice. The last term that plays a role is the costs of participating in two platforms, which depends on the observable quality, X, the private quality information, θ, and two platforms' fees which are summarized in a constant term, ρ_0.

3.2 Simultaneous Entry Game

3.2.1 Sellers' Expected Payoffs

Combining (11) and (12) which respectively characterize the consumer's and seller's strategies, I show that the seller's entry decision can be summarized as (13). It is easy to find that the game presented in Section 2 is equivalent to an entry game in which sellers simultaneously decide which platform they want to participate in. A seller makes decision according to (13), which is ultimately determined by the seller's own characteristics and other sellers' observable characteristics and their entry decisions. The fact that θ in (13) is private information implies there is imperfect information among sellers. When a seller takes action, he can not perfectly know competitor's choices. So the seller can only form the expectation of profits on two platforms based on a prior of competitor's entry decisions. Seller j's expected profits on the two platforms equals to

$$\begin{aligned}\mathbb{E}\left[\triangle\pi_j^t \middle| X_j^t,\, \Omega^t,\, P^t\right] &= \mathbb{E}\left[\ln d_{j,S}^t - \ln d_{j,NS}^t \middle| X_j^t,\, \Omega^t,\, P^t\right] \\ &+ \rho_0 + \rho_1 X_j^t + \rho_2\theta \end{aligned} \tag{14}$$

where

$$\begin{aligned}\mathbb{E}\left[\ln d_{j,S}^t - \ln d_{j,NS}^t \middle| X_j^t,\, \Omega^t,\, P^t\right] &= \mathbb{E}\left[\ln\left(\sum_{j'\in\mathcal{J}^t(S)} \exp\left[u^t{}_{j',S}\right]\right) - \ln\left(\sum_{j'\in\mathcal{J}^t(NS)} \exp\left[u^t{}_{j',NS}\right]\right) |X_j^t,\, \Omega^t,\, P^t\right](\lambda-1) \\ &+ \left(\mathbb{E}\left[\theta_j^t \middle| X_j^t,\, m=S\right] - \mathbb{E}\left[\theta_j^t \middle| X_j^t,\, m=NS\right]\right)\beta \end{aligned} \tag{15}$$

Here I define the vector P^t as

$$P^t \equiv \left\{p_g^t : \text{any seller } g \text{ of product } t\right\}$$

where p_g^t is the seller j's conjecture about the probability that seller g chooses Platform S. Since the private quality $\left\{\theta_g^t\right\}$ are independently and identically distributed, any two pairs of sellers, say seller j and seller j', have the same perception of a third seller g's private signal and its entry strategy. So the conjecture vector P^t is symmetric among all sellers of product t.

Denote a seller's information set as

$$\Omega^t \equiv \left\{X_{j'}^t,\ price_{j'}^t : \text{any seller } j' \text{ of product } t\right\}$$

which includes all sellers' observable characteristics.

Seller j chooses platform S iff

$$\mathbb{E}\left[\triangle\pi_j^t | X_j^t,\ \Omega^t,\ P^t\right] \geq 0$$

. From the perspective of competitors and we researchers, the probability that the seller participates in Platform S is given by:

$$\begin{aligned} p_j^t &\equiv \Pr(m = S|\, X_j^t,\ \Omega^t,\ P^t) \\ &= \Pr(\mathbb{E}\left[\triangle\pi_j^t | X_j^t,\ \Omega^t,\ P^t\right] \geq 0) \\ &= \Pr(\mathbb{E}\left[\ln d_{j,S}^t - \ln d_{j,NS}^t \middle|\, X_j^t,\ \Omega^t,\ P^t\right] + \rho_0 + \rho_1 X_j^t + \rho_2\theta \geq 0) \\ &= \Phi(\frac{1}{\rho_2}\mathbb{E}\left[\ln d_{j,S}^t - \ln d_{j,NS}^t \middle|\, X_j^t,\ \Omega^t,\ P^t\right] + \frac{\rho_0}{\rho_2} + \frac{\rho_1}{\rho_2} X_j^t) \end{aligned} \quad (16)$$

, which is a function of the seller's observable characteristics X_j^t.

Since the perception about seller j's choices varies with its observables X_j^t , the vector P^t can be of a large dimension when the competition in product t involves a lot of sellers with different observable characteristics. To simplify the following analysis, I assume that X is a discrete variable taking K possible values: $\{x_k : k = 1, ..., K\}$. Therefore, conjectures of sellers' entry to platform S can be reduced to a K by 1 vector $P^t = \{p_k^t : k \in \{1, ...K\}\}$.

To construct the seller's expected profit function, I still need to know the consumer's expectation about the private quality: $\mathbb{E}\left[\theta_j^t \middle|\, X_j^t,\ m\right]$. I assume the private quality θ satisfies Assumption 2 and follows standard truncated normal distribution on $[\underline{\theta},\ \bar{\theta}]$. The truncated normal distribution is appealing here because it offers a closed form expression of the conditional expectation According to Proposition 1, in equilibrium, the consumer's belief of θ is consistent with the seller's choice.

$$\begin{aligned}\mathbb{E}\left[\theta^t \middle| X_k^t, S\right] &= \mathbb{E}\left[\theta^t \middle| \theta^t \geq \theta^{t*}(X_k^t, S)\right] \\ &= \frac{\phi(\theta_j^{t*}) - \phi(\bar{\theta})}{\Phi(\bar{\theta}) - \Phi(\theta_j^{t*})}\end{aligned} \tag{17}$$

where θ_j^{t*} is determined by

$$p_k^t = \Pr(\theta^t \geq \theta^{t*}(X_k^t, S)) = \frac{\Phi(\bar{\theta}) - \Phi(\theta_j^{t*})}{\Phi(\bar{\theta}) - \Phi(\underline{\theta})}$$

. Based this equations, θ_j^{t*} can be expressed as a function of p_k^t :

$$\theta_j^{t*} = \Phi^{-1}\left[\Phi(\bar{\theta}) - p_k^t[\Phi(\bar{\theta}) - \Phi(\underline{\theta})]\right]$$

and so does $\mathbb{E}\left[\theta^t \middle| X_k^t, S\right]$. Similarly, the consumer's expectation of θ on Platform NS can be written as:

$$\begin{aligned}\mathbb{E}\left[\theta^t \middle| X_k^t, NS\right] &= \mathbb{E}\left[\theta^t \middle| \theta^t \leq \theta^{t*}(X_k^t, S)\right] \\ &= \frac{\phi(\underline{\theta}) - \phi(\theta_j^{t*})}{\Phi(\theta_j^{t*}) - \Phi(\underline{\theta})}\end{aligned} \tag{18}$$

3.2.2 Equilibrium

As the two-sided market theory in Section 2 is equivalent to a seller simultaneous entry game with incomplete information, the equilibrium established in Proposition 1 of Section 2 can be rephrased as Bayesian Nash equilibrium conjectures $\{p_k^{t*} : k \in \{1, ...K\}\}$ such that

1. Consumers form a correct expectation of θ conditional on the seller's choice of platform and the observable characteristics, as specified in equations (17) and (18). Observing the number of sellers on each platform and products' observable quality, consumers evaluate the payoffs from sellers and platforms shown in equation (4) and equation (6), and make choices according to (11) .

2. A seller chooses between two platforms to maximize its expected profit, based on its conjecture about competitors' strategies. Given the observable quality type x_k^t , the probability that the seller enters Platform S is

$$p_k^{t*} = \widetilde{\Phi}\left(\frac{\rho_0}{\rho_2} + \frac{1}{\rho_2}\mathbb{E}\left[\ln \frac{d_{j,S}^t}{d_{j,NS}^t}\middle| x_k^t,\ \Omega^t,\ P^t\right] + \frac{\rho_1}{\rho_2}x_k^t\right) \quad \forall k = 1, ...K \tag{19}$$

where

$$\widetilde{\Phi}(x) \equiv \Pr(\theta \geq x) = \frac{\Phi(\bar{\theta}) - \Phi(x)}{\Phi(\bar{\theta}) - \Phi(\underline{\theta})}$$

. Therefore, the equilibrium conjectures about all sellers' entries can be summarized with the following equation

$$\begin{aligned} P^t &= \begin{bmatrix} \frac{\Phi(\bar{\theta})-\Phi(-\mathbb{E}\left(d^t_{1,S}-d^t_{1,NS}|x^t_1,\Omega^t,P^t\right)\frac{1}{\rho_2}-\frac{\rho_0}{\rho_2}-x^t_1\frac{\rho_1}{\rho_2})}{\Phi(\bar{\theta})-\Phi(\underline{\theta})} \\ \cdots \\ \frac{\Phi(\bar{\theta})-\Phi(-\mathbb{E}\left(d^t_{k,S}-d^t_{k,NS}|x^t_k,\Omega^t,P^t\right)\frac{1}{\rho_2}-\frac{\rho_0}{\rho_2}-x^t_k\frac{\rho_1}{\rho_2})}{\Phi(\bar{\theta})-\Phi(\underline{\theta})} \\ \cdots \\ \frac{\Phi(\bar{\theta})-\Phi(-\mathbb{E}\left(d^t_{K,S}-d^t_{K,NS}|x^t_K,\Omega^t,P^t\right)\frac{1}{\rho_2}-\frac{\rho_0}{\rho_2}-x^t_K\frac{\rho_1}{\rho_2})}{\Phi(\bar{\theta})-\Phi(\underline{\theta})} \end{bmatrix} \\ &\equiv \Psi(P^t;\Omega^t,[\alpha,\beta,\gamma,\lambda]',[\rho_0,\rho_1,\rho_2]') \end{aligned} \tag{20}$$

. The equilibrium conjecture P^{t*} is defined as a fixed point, which satisfies the mapping from a seller's belief of its competitors' entry decisions to its competitors' beliefs of the seller's decision. The existence of P^{t*} can be directly proved by the Brouwer Fixed Point Theorem.

3.3 Estimation Strategy

In this simultaneous entry game, the seller's location choice involves two sets of parameters

$$\Theta = \left\{[\alpha,\beta,\gamma,\lambda]',[\rho_0,\rho_1,\rho_2]'\right\}$$

. $[\alpha,\beta,\gamma,\lambda]'$ captures the consumer's preferences and determines the demand function. $[\rho_0,\rho_1,\rho_2]'$ describes the seller's responses to competitor's choices and platforms' screening policies and takes effect in the seller's entry cost function. The estimation of Θ is carried out using the Nested Pseudo Likelihood (NPL) estimator proposed by Aguirregabiria and Mira (2007). The estimation procedures are presented as follows.

Step 1. Estimation of $[\alpha,\beta,\gamma,\lambda]'$

I start the estimation by making an initial guess of sellers' strategies:

$$\hat{P}^{t,0} = \left\{p^{t,0}_k : k \in \{1,...K\}\right\}$$

. According to Aguirregabiria and Mira (2007), this guess needs not to be a consistent estimator of sellers' equilibrium strategies,

$$\hat{P}^{t*} = \left\{p^{t*}_k : k \in \{1,...K\}\right\}$$

. Using $\hat{P}^{t,0}$, I can construct the consumer's perceptions of unobservable quality $\mathbb{E}\left[\theta_j^t | x_j^t,\ m\right]$ according to (17) and (18), and write down the demand function of seller j on platform m: $d_{j,m}^t$ in (11). In order to simplify the estimation, I transfer $d_{j,m}^t$ to a linear function by taking the log-difference between $d_{j,m}$ and the market share of outside option.

$$\begin{aligned} \ln d_{j,m}^t - \ln d_0^t &= \alpha x_j^t + \mathbb{E}\left[\theta_j | \, x_j^t, m\right]\beta + price_j^t \gamma \\ &+ \ln\left(\sum_{j' \in \mathcal{J}^t(m)} \exp\left[x_{j'}^t \alpha + \mathbb{E}\left[\theta_{j'}^t | x_{j'}^t,\ m\right]\beta + price_{j'}^t \gamma\right]\right)(\lambda - 1) \end{aligned} \tag{21}$$

where $d_{j,0}^t$ is the number of consumers who choose the outside option. Calculating $d_{j,0}^t$ requires the information of total number of potential consumers in the whole market, which is usually not available in the data. To deal with this problem, I follow the previous literature and assume a fixed number of potential consumers for product t.

I estimate $[\alpha, \beta, \gamma, \lambda]'$ by minimizing the distance between market shares predicted by the model and those observed in the data.

$$[\hat{\alpha}^1, \hat{\beta}^1, \hat{\gamma}^1, \hat{\lambda}^1]' = \operatorname{argmin} \sum_t \sum_j \left(\overline{\ln d_{j,m}^t} - \ln d_{j,m}^t\right)^2$$

where $\overline{\ln d_{j,m}^t}$ is the observed market share in the data.

Step 2. Estimation of $[\rho_0, \rho_1, \rho_2]'$

Given $P^{t,0}$ and the estimator $[\hat{\alpha}^1, \hat{\beta}^1, \hat{\gamma}^1, \hat{\lambda}^1]'$, the difference of seller j's expected profits on the two platform can be expressed as

$$\begin{aligned} \mathbb{E}\left[\triangle\pi_j^{t,1} \middle| \, X_j^t,\ \Omega^t,\ P^{t,0}\right] &= \mathbb{E}\left[\ln \hat{d}_{j,S}^{t,1} - \ln \hat{d}_{j,NS}^{t,1} \middle| \, X_j^t,\ \Omega^t,\ P^{t,0}\right] \\ &+ \rho_0 + \rho_1 X_j^t + \rho_2 \theta \end{aligned}$$

where the difference of the expected demands are determined by

$$\begin{aligned} \mathbb{E}\left[\ln \hat{d}_{j,S}^t - \ln \hat{d}_{j,NS}^t \middle| X_j^t,\ \Omega^t,\ P^{t,0}\right] &= \mathbb{E}\left[\ln\left(\sum_{j' \in \mathcal{J}^t(S)} \exp\left[\hat{u}^t{}_{j',S}\right]\right) - \ln\left(\sum_{j' \in \mathcal{J}^t(NS)} \exp\left[\hat{u}^t{}_{j',NS}\right]\right) |X_j^t,\ \Omega^t,\ P^t\right](\hat{\lambda}^1 - 1) \\ &+ \left(\mathbb{E}\left[\theta_j^t | \, X_j^t,\ m = S\right] - \mathbb{E}\left[\theta_j^t | \, X_j^t,\ m = NS\right]\right)\hat{\beta}^1 \end{aligned}$$

and a consumer's utility from seller j' on platform m predicted by the estimator $[\hat{\alpha}^1, \hat{\beta}^1, \hat{\gamma}^1]'$ equals to

$$u^t_{j',m} = x^t_{j'}\hat{\alpha}^1 + \mathbb{E}\left[\theta^t_{j'}|x^t_{j'},\ m\right]\hat{\beta}^1 + price^t_{j'}\hat{\gamma}^1$$

.

Therefore, the probability that a seller of quality type x_k chooses Platform S equals to

$$p^t_k = \widetilde{\Phi}(\frac{\rho_0}{\rho_2} + \frac{1}{\rho_2}\mathbb{E}\left[\ln \frac{\hat{d}^{t,1}_{j,S}}{\hat{d}^{t,1}_{j,NS}}\middle| x^t_k,\ \Omega^t,\ P^t\right] + \frac{\rho_1}{\rho_2}x^t_k) \quad \forall k = 1, ...K$$

. The estimator $[\hat{\rho}^1_0, \hat{\rho}^1_1, \hat{\rho}^1_2]'$ are obtained by maximizing the likelihood of all sellers' choices which are observed in the data.

$$\left[\hat{\rho}^1_0, \hat{\rho}^1_1, \hat{\rho}^1_2\right]' \quad = \quad \text{argmax} \prod_{t=1}^{T}\prod_{k=1}^{K}\left(p^t_k\right)^{N^t_{k,S}}\left(1-p^t_k\right)^{N^t_k - N^t_{k,S}}$$

where $\{N^t_k : k = 1, ..., K\}$ is the number of sellers of type $\{x_k : k = 1, ..., K\}$. $\left\{N^t_{k,S} : k = 1, ..., K\right\}$ denotes the number of sellers that join Platform S . Both information can be observed in the data.

Step 3. Fixed-Point Algorithm

Using $\hat{P}^{t,0}$ and estimators $[\hat{\alpha}^1, \hat{\beta}^1, \hat{\gamma}^1, \hat{\lambda}^1]'$ and $[\hat{\rho}^1_0, \hat{\rho}^1_1, \hat{\rho}^1_2]'$, we can calculate sellers' strategies $\hat{P}^{t,1}$

$$\hat{P}^{t,1} = \Psi\left(\hat{P}^{t,0},\ \Omega^t,\ [\hat{\alpha}^1, \hat{\beta}^1, \hat{\gamma}^1, \hat{\lambda}^1]',\ \left[\hat{\rho}^1_0, \hat{\rho}^1_1, \hat{\rho}^1_2\right]'\right)$$

.

If the distance between $\hat{P}^{t,1}$ and $\hat{P}^{t,0}$ is very small such that

$$\parallel \hat{P}^{t,1} - \hat{P}^{t,0} \parallel \leq c_0$$

where c_0 is a small positive number near zero, it means that the seller's equilibrium strategy generated from $\hat{P}^{t,0}$ coincides with its competitors' conjecture of the seller's strategy. In other words, $\hat{P}^{t,0}$ is the fixed point that satisfies the mapping defined by (20) and the equilibrium solution of the game is obtained.

If $\parallel \hat{P}^{t,1} - \hat{P}^{t,0} \parallel \geq c_0$, the estimation continues. I replace $\hat{P}^{t,0}$ with $\hat{P}^{t,1}$ and repeat Step 1 and Step 2 to get $[\hat{\alpha}^2, \hat{\beta}^2, \hat{\gamma}^2, \hat{\lambda}^2]'$ and $\hat{P}^{t,2}$.... and keep the iteration until $\hat{P}^{t,R-1}$ converges, that is

$$\hat{P}^{t,R} = \Psi(\hat{P}^{t,R-1}; data, [\hat{\alpha}^{R-1}, \hat{\beta}^{R-1}, \hat{\gamma}^{R-1}, \hat{\lambda}^{R-1}]', \left[\hat{\rho}^{R-1}_0, \hat{\rho}^{R-1}_1, \hat{\rho}^{R-1}_2\right]')$$

and

$$\parallel \hat{P}^{t,R} - \hat{P}^{t,R-1} \parallel \leq c_0$$

. Then $\hat{P}^{t,R-1}$ is the vector of sellers' equilibrium strategies and the estimators $[\hat{\alpha}^{R-1}, \hat{\beta}^{R-1}, \hat{\gamma}^{R-1}, \hat{\lambda}^{R-1}]'$ and $\left[\hat{\rho}_0^{R-1}, \hat{\rho}_1^{R-1}, \hat{\rho}_2^{R-1}\right]'$ are the parameters which best describe the observed consumers' and sellers' decisions.

4 Data and Estimation Results

4.1 Background

In section 4, I apply the structural simultaneous entry game specified in Section 3 to the data of Alibaba. In this sub-section, I first introduce the background of Alibaba and its business strategy and then discuss the data.

The Alibaba Group, a Chinese firm which operates online trading platforms, has seized a leadership position in the worldwide fast-growing E-commerce market. According to the financial report from Alibaba, the gross merchandise value processed on Alibaba's platforms in 2013 is $248 billion, exceeding that of Amazon ($116.4 billion) and Ebay ($85.7 billion) combined. The company also outpaces their American counterparts in terms of the profit margin and the growth of annual revenue. As for the market size, Alibaba operates the largest online trading platforms in the world. As of 2013, Alibaba's platforms had hosted more than 7 million merchants selling various products. On the consumer side, there were 231 million active users across Alibaba's platforms and each active user made 49 purchases during that year, according to Alibaba.

To better serve sellers and consumers, Alibaba divides its marketplace into two shopping sites: Tmall and Taobao and implements differentiated quality management strategies on the two platforms. First, Taobao is a free-to-entry market which does not screen sellers' products. Sellers do not pay transaction fees to the platform unless they use the services provided by Taobao such as advertisements and the payment system. Compared to Taobao, Tmall has a rigorous control of the quality of products sold on the platform. Tmall targets sellers which are brand owners or the retailers that are officially registered with the Department of Administration for Industry and Commerce. A merchant must submit to Tmall the copies of the registration certificate and other related legal documents in order to qualify for entering Tmall. Tmall may ask a seller to provide sample products for examinations. More importantly, Tmall requires a deposit from every seller upon entry. If a seller is found to sell inferior products, Tmall will punish the seller by confiscating part of the deposit, the amount of which can be as much as five times of the transaction value of the product. The seller has to refill the deposit in order to continue its business on Tmall. Using this policy, Tmall makes a seller's profit positively correlated with the quality of product.

Tmall and Taobao, although controlling product quality in different ways, share some features on the platform settings: They both help sellers to establish their individual webpages on which sellers can upload product pictures and add descriptive introductions of products. Sellers' webpages on the two sites have the same format. Both platforms provide rating systems through which consumers can rate the products they have bought. The statistics of rating are observable on the seller's webpage. As for sellers' participations of platform, Alibaba requires sellers to be single-homing which means a seller cannot simultaneously do business on Tmall and Taobao. But Alibaba offers sellers the flexibility of transferring across the two markets. A merchant can initiate a transfer application anytime and move to the other platform as long as its application is approved by Alibaba. The seller's historical data including past sales, rating scores, payment records, product descriptions also gets re-posted on the new site.

According to above descriptions, it is easy to find that Tmall and Taobao are exactly counterparts of Platform S and Platform NS in my model. Both platforms are operated by Alibaba but have different quality screening standards. Tmall screens sellers' products for quality. While on Taobao, there is no quality regulation. Alibaba requires sellers to be single-homing but allows them to transfer across two markets without losing any historical records. These business strategies fairly match the setups of my theoretical model and provide a good context to empirically study the impact of platform screening on the choices of consumers and sellers.

The size and business strategy of Alibaba make it an attractive data source to perform the empirical study. In order to obtain the information on Alibaba's two platforms, I develop a web-crawling program using Python and Perl to collect data, which is about 89 camera models sold at Tmall and Taobao in July 2014. This seller-level cross-sectional data contains variables including camera ID, the platform the seller is located on (Tmall or Taobao), price, sales within the past 30 days, rating scores. The definitions of these variables are presented as following.

Camera ID: each camera model is treated as an independent product. For example, Nikon D 3300 is one product and Canon 500D is another. The variable of camera ID helps to identify all sellers of a particular camera.

Sales within the past 30 days: This variable records the number of a camera sold by a seller in the past month. Every time a seller receives an order, the system updates the seller's sales information and publishes it on the seller's webpage. This variable will be used to construct the seller's market share in the later estimation.

Rating score: Both Tmall and Taobao allow buyers to rate sellers and their products using three criteria: whether the item was as described, the shipping and return speed, and customer service. For each criterion, a buyer can submit a score ranging from 0 to 5. Using the scores from all previous buyers, the platform calculates the average score for each of above three criteria and posts them on the seller's webpage so that future buyers can view them. Several features of the rating scores need to be mentioned: First, the system does not distinguish the ratings about different products sold by the same seller. So the statistics not only represent consumers' feedbacks about a particular camera, but also reflect the average quality of all kinds of products sold by the seller. Second, compared to the sales data which gets updated immediately after orders are placed, the update of rating has a significant time-delay. This is because the system has to wait for the feedback from consumers. Suppose a consumer purchases a unit of Nikon D3300. The sales of this camera are immediately increased by 1. But it may take several days for the consumer to receive the product. When the consumer tries the product and posts the rating, additional one or two weeks may have passed. So the consumer's rating associated with this transaction could arrive to the platform's database three or four weeks later than the sales information.

The rating scores, although informative of the seller's reliability, only reflect partial information. First, as mentioned above, the rating score can be buyers' feedbacks on any products sold by the seller. Given a seller may carry many kinds of products which vary in quality, the rating score does not necessarily demonstrate the quality of a particular product. Second, according to the policy of Tmall and Taobao, a buyer has only 15 days to submit his score since receiving the product. It means that the rating score only reveals the consumer's experience in a very short time period. For durable products such as cameras, two-week is a too short time for consumers to tell the overall quality. Based on these reasons, the variable of rating score can serve as a proxy of the observable quality of the product. In the estimation, the score about "whether the item was as described" is treated as seller's observable characteristic.

Price: It represents the camera's listed price. Within a month, a camera may be traded in different prices. But given the fact that there is not holiday in the month when the data was collected, cameras' prices should not have a large variation.

Platform: This variable records the platform where the seller is located. It is a discrete variable and either equals to Tmall or Taobao.

In the context of Alibaba, the entry game in Section 3 can be briefly rephrased as following: At the beginning of a month, sellers located on Tmall and Taobao simultaneously decide whether to stay in the current platform or move to the other platform. Each seller makes decision based on the observable quality: rating scores, the private quality information and the expectation of other sellers' location choices. At the end of the month, the web-crawling program is run to collect the seller's choice of platform, monthly sales, prices and ratings.

4.2 Data Description

The statistics of variables are presented in Table 1. The dataset includes 35384 observations of sellers for 89 camera models. The number of sellers for a camera varies from 101 to 1272. The average number of seller of a camera model is 559 with a standard deviation 285. Each merchant on average sells 2 units of cameras in a month and has a small market share: 0.125%. These evidences support my model's assumption: A lot of sellers compete in the market and each has a small market power. A dummy variable named positive sales is created to indicate whether a seller's monthly sales are non-zero. And about 12.84% of sellers have positive sales during the data period. The average price of cameras is 8405 RMB (about 1400 dollars). There are about 10% of sellers located on Tmall. The rating score about "whether the item was as described" has a mean 4.6 out of 5 and a standard variation 1.15.

In Table 2, I compare the characteristics of sellers located on Tmall and Taobao. Firstly, the average prices on Tmall and Taobao are very close, which equal to 8044 and 8444 RMB respectively. The price distributions on the two platforms are displayed in Figure 1. Most cameras sold on both platforms are priced less than $30,000$ RMB. Particularly, cameras priced between $5,000$ and $15,000$ RMB have a big population on the two markets. These cameras are usually low-end or medium-end models which target ordinary consumes instead of a small group of professional users like photographers. This guarantees the two-sided market model specified in Section 2 can be applied to the dataset. Then I look at the sales on two platforms. The average monthly sales of a Tmall seller is 4, twice more than that of a Taobao seller:2.08. About 28.6 percent of Tmall sellers are able to sell one or more cameras whiles this percentage falls to 12.1 on Taobao. As for individual market share, sellers on both platforms generally have very small market shares, although Tmall sellers have a slightly larger average market share .2572% compared to .1114% for Taobao sellers. At the end of Table 2, I compare the total number of sales realized on the two platforms. Although on individual level, a Tmall seller performs better on than its peers on Taobao, the total number sales processed on Tmall is 25 percent of those on Taobao.

Rating score and distribution. The variable of rating score serves as the seller's observable quality in the estimation. It is one of the most important explanatory variables and deserves further statistical analysis. For each platform, I calculate the proportion of sellers with particular rating scores and plot the results in Figure 2. The distributions of the rating score on Tmall and Taobao exhibit different trends. On Tmall, most rating scores fall into the range from 4.7 to 4.9. The distribution of rating reaches the peak at the value 4.8. There are only very small numbers of sellers located on the two tails of the line. While on Taobao, sellers' ratings are distributed much evener along the line.

Figure 3 displays the proportion of Tmall sellers in each rating score group. As the rating increases, the proportion of Tmall sellers first increases then decreases. This observation is consistent with the prediction of Proposition 2: when observable quality increases, the seller's inventive to participate in Platform S (Tmall in our case) may take an inverted-U curve.

To facilitate the analysis, I classify the rating scores into three groups and create a dummy variable to denote the group the seller belongs to. The rating group dummy is defined as follows: Group_1 stands for the group of sellers whose rating scores are less than 4.7; Group_2 sellers are those who have rating scores equal or larger than 4.7 but smaller than 4.9; The sellers obtaining rating scores equal or larger than 4.9 are classified to Group_3.

In Table 3, I present the number of sellers and the proportion of Tmall sellers in each rating group. Group_1 sellers are of a relatively small number:3785, about 11% of the total population. The number of Group_2 and Group_ 3 are 15,620 and 15,979 perceptively. In Group_1, about 12% of sellers participate in Tmall. This proportion increases to 18% in Group_2 and then drops to 1% in Group_3. Again, this non-monotonic relationship between the rating score and sellers' participation of Tmall coincides with the theory's prediction.

Table 4 compares the sellers' characteristics by rating groups and by platforms. The average price of cameras sold on Tmall increases with the rating, while on Taobao the sellers who have a medium rating post the lowest average price. Given the rating group, the average prices on the two platforms are slightly different. Sellers in Group_2 and Group_3 charge a higher average price on Tmall than that on Taobao. The reverse is true for the Group_1 sellers. The second sub-table compares the monthly sales. On Tmall, Group_2 sellers acquire the largest monthly sales. In contrast, on Taobao, Group_3 sellers obtain monthly sales of 2.5 cameras, better than that of Group_1 and Group_2:0.04 and 1.5. This implies that consumers on Tmall and Taobao have different responses to the rating when they make purchase decisions, because the quality screening by Tmall alters consumers' belief of the product quality.

4.3 Estimation Results

Table 5 presents the estimation results. There are two sets of parameters. The parameters in the demand function measure the consumer's responses to the seller's and the platform's characteristics. The second set of parameters account for the impacts of factors playing a role in the seller's profit function. These parameters are estimated under the assumption that there are $10,000$ potential consumers for each camera model. The dummy variable of seller rating is used as the observable quality. The unobserved quality, θ, is assumed to follow standard truncated normal distribution bounded on $[-50, 50]$.

All estimators are of theoretically anticipated signs and statistically significant. A seller that has a higher rating score acquires a larger market share. If a seller manages to raise his rating score from Group_1 (less than 4.7) to Group_2 (between 4.7 and 4.9), his (log) market share can increase by 3.05. Similarly, the sellers in Group_3 obtain a market share 2.46 bigger than that of Group 1. The price, as expected, has a negative effect on sales. An increase of $1,000$ RMB on a camera's price leads to a 0.051 loss of the market share.

According to the theory, consumers update their belief of the seller's private quality according to the seller's rating and choice of platform, and take the belief into account when they choose among sellers. This screening effect, in the empirical study, is captured by the parameters of $\mathbb{E}[\theta|X, m]$. In order to investigate the impacts of quality screening for sellers that belong to different rating groups, I construct an intersection terms using $\mathbb{E}[\theta|X, m]$ and the rating group dummy, and estimate their parameters. These estimators vary by the seller type. More specifically, for Group_1 sellers, the parameter of $\mathbb{E}[\theta|X, m]$ is 2.59, which suggests a strong screening effect. By joining Tmall, this type of sellers can improve consumers' expectation about the product quality and obtain a larger market share. As for the Group_2 sellers, this estimator slightly decreases by 0.77, but still remains positive and significant. It indicates the screening effect also plays a role here. For Group_3 sellers that have the highest rating score, the impact of $\mathbb{E}[\theta|X, m]$ is 2.59 minus 2.86, which is negative and almost zero. This means the screening effect disappears and joining Tmall does not help these sellers to prompt sales. These estimations of $\mathbb{E}[\theta|X, m]$ are consistent with Proposition 1: when the observable quality is lower than certain threshold, join Platform S can help sellers to signal their quality and increase the sales. This screening effect vanishes if the observable quality is sufficiently high such that sellers can always pass the screening.

Besides the screening effect, the network and competition effects also influence a seller's market share. The theory in Section 2 shows that, when a platform hosts more sellers of good quality and provides higher total expected utility to consumers, an individual seller can become better or worse, depending on whether the network effect dominates the competition effect. The relative magnitude of the network and competition effects is captured by $(\lambda - 1)$, the parameter in front of $\ln EU(m)$. Table 5 shows that this parameter is positive, suggesting that the network effect plays a dominant role on Tmall and Taobao. When the total expected utility offered by the entire group increases by 1, an individual seller's (log) market share enlarges by 1.71.

A seller's choice of platform is governed by the expected demands on two platforms and the seller own quality types. The parameters of these variables are reported in the second part of Table 5. The parameter of $\mathbb{E}[\ln \frac{d_{j,S}}{d_{j,NS}}]$ equals to 0.62, which is positive and significant. This implies that a seller is more willing to join Tmall if it expects to enjoy a larger market share on Tmall. Moreover, if we recall the seller's entry probabilities specified in (16), the parameter of $\mathbb{E}[\ln \frac{d_{j,S}}{d_{j,NS}}]$ also characterizes the impact of the unobservable quality on a seller's comparative profit on Tmall. The estimation result indicates this impact is positive, which is coherent with the theory. The coefficients of Group_2 and Group_3 are respectively 1.296 and 2.333, which suggests that a seller's relative profit on Tmall increases with the rating score. This is also consistent with the spirit of the quality screening: a seller of higher observable quality pays less cost of screening.

Using the estimators in Table 5, I obtain the sellers' choice probabilities predicted by the model. To test the model's prediction power, I compare the predicted probabilities with sellers' choices observed in the data, and measure the differences between the two with a variable: prediction error. Figure 4 displays the statistics and distribution of the prediction errors. The prediction errors have an almost zero average $-3.37 * 10^{-6}$ and a maximum 0.209 and minimum -0.376. This means that on average the predicted choice probabilities do not deviate too much from the observed data. About 33% prediction errors fall into a small interval of length 0.005 covering point zero. It suggests for these sellers, the model well predicts their choices. Over two thirds of prediction errors have absolute values less than 0.05. Some prediction errors are large but their densities are very small in the whole population. Overall, the sellers' choices of platform are well explained by the models in Section 2 and Section 3.

Counterfactual Analysis. The estimation results show that quality screening employed by Tmall plays an important role in determining the consumer's utility and seller's profits on the two platforms. It further influences the consumer's demand of platforms and sellers, and the seller's choices of platforms. This is important for Alibaba, the parent firm of Tmall and Taobao, which derives its profit mainly from the transactions processed in the two marketplaces. According to the theoretical model, the transaction volume on Alibaba depends on the number of consumers that patronize Tmall and Taobao. Given the utility provided by the outside option is fixed, we can expect that if Alibaba alters the screening policy of Tmall, the consumer's demand for Tmall and Taobao also get changed. To quantify the impact of screening on Alibaba's business, I conduct a counterfactual analysis in which I assume Alibaba cancels the screening policy and makes Tmall a free-entry market.

If Tmall becomes a free-to-enter market, it would have no difference from Taobao. So the case is the same as that if Alibaba only operates one platform: Taobao. Consumers choose either Taobao or the outside option depending on which provides a higher utility. Suppose all sellers continue to run their business in the same product market as before, now they have no options but to participate in Taobao. Since there is no screening effect, the consumer's expectation of a seller's unobservable quality equals to the population mean: $\mathbb{E}(\theta) = 0$. Therefore, the consumer i's utility from purchasing seller j is reduced to

$$\begin{aligned} \tilde{u}_{i,j}^t &= u_{i,j}^t + \epsilon_{i,j}^t \\ &= x_j^t \alpha + price_j^t \gamma + \epsilon_{i,j}^t \end{aligned}$$

, in which the impact of the seller's unobservable quality vanishes.

Figuring out the utility consumers obtain from an individual seller, now I compute the consumer's expected utility from Alibaba's platform(s) before and after Alibaba would remove quality screening. I characterize the consumer's choice using a Multinomial Logit model. It is well known that in this type of model, if two alternatives are merged into one, the number of consumers that choose the new alternative decreases. This fact holds even if the characteristics of the two alternatives remain the same before and after merger. Back to our case, suppose Alibaba merges Tmall and Taobao into one platform but still imposes quality screening on Tmall sellers, there will be fewer consumers visiting the merged platform. This change is not due to quality screening but only because we reduce the choice set of consumers from three options (Tmall, Taobao and the outside option) to two options (merged platform and the outside option).

In order to fairly investigate the effect of quality screening, I construct the following variable to compare the change of consumers' utility before and after Alibaba would cancel quality screening:

$$CU^t = \frac{\sum_{m\in\{S,NS\}} \sum_{j'\in\mathcal{J}^t(m)} \exp\left[x^t_{j'}\alpha + \mathbb{E}\left[\theta^t_{j'}|x^t_{j'},\ m\right]\beta + price^t_{j'}\gamma\right]}{\sum_{\text{all j in market t}} \exp\left[X^t_j\alpha + price^t_j\gamma\right]}$$

. The numerator captures the total utility provided by all sellers of product t, when Alibaba employs quality screening to discriminate sellers. This number can also be understood as the proportion of consumers visiting sellers located on Alibaba compared to the proportion of consumers that choose the outside option. The denominator measures the total utility offered by sellers if there is no quality screening.

If this ratio is bigger than one, it means the quality screening policy improves consumers' welfare, and to some extent also helps Alibaba to achieve a better position in the competition with the outside option. In Table 6, I present the statistics of CU^t. The variable CU^t of the 89 camera models has a mean of 2.86, which suggests consumers can enjoy larger utility from sellers when Alibaba screens products for quality. For 74% camera models, CU takes values larger than one, which indicates the quality screening alleviates the problem of asymmetric information, benefits consumers, and also prompts Alibaba's market share.

5 Conclusion

In two-sided markets, a user's utility depends not only on the number of users on the other side, but also the quality provided by the other side. Previous studies of two-sided markets have mainly focused on the quantity externality. Less attention has been paid to end-users' concerns about quality. In this paper I propose a model which incorporates both quality and quantity in the end-users' interactions, and analyze how platforms can use quality screening to alleviate the problem of asymmetric information and motivate end-users' participation. I study this question from both the theoretical and empirical perspective.

First, I build a theory in which sellers compete for selling a particular product to consumers through online trading platforms. The product quality is heterogeneous and imperfectly observable to other players other than the seller itself. I model duopoly platform competition where one platform (Platform S) uses quality screen and the other (Platform NS) does not. I show that the quality screening influences the consumer's belief of the quality and the choice of sellers and platforms. These choices result in the screening, network and competition effects in turn drive sellers to different platforms.

Next, I formalize the end-users' choices in the theory as a simultaneous entry game with incomplete information, and carry out the estimation using the data from Alibaba's Tmall and Taobao. The estimation results are consistent with theory: the screening effect varies by the observable quality, and the screening cost decreases with quality. Moreover, I find the network effect overrides the competition effect in the case of Alibaba.

Do birds of a feather flock together? Not necessarily. The comparative statics illustrates that under the influence of the screening effect, the seller's incentive to submit to quality screening first increases with the observable quality. Once the observable quality reaches a certain point, the network and competition effects take over and may drive sellers to the platform which does not use screening. In other words, the seller of high observable quality may be willing to pool with the seller with low observable quality, in order to enjoy the network effect or avoid the competition effect. In this sense, the birds of a feather do not always flock together.

Does the quality screening benefit consumers and the platform provider? Yes. Using the estimation results, I conduct the counterfactual analysis and study the change of the consumers' utility if Alibaba did not employ quality screening. I find that with the presence of quality screening, the total utility that consumers obtain from the entire seller group gets increased in 70% of camera markets in the data. Quality screening helps Alibaba to attract more consumers and increases its market share.

References

Ackerberg, D. A. and G. Gowrisankaran (2006). Quantifying equilibrium network externalities in the ach banking industry. *RAND Journal of Economics 37*(3), 738–761.

Aguirregabiria, V. and P. Mira (2002). Swapping the nested fixed point algorithm: A class of estimators for discrete markov decision models. *Econometrica 70*(4), 1519–1543.

Aguirregabiria, V. and P. Mira (2007). Sequential estimation of dynamic discrete games. *Econometrica 75*(1), 1–53.

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

Ambrus, A. and R. Argenziano (2004). Network markets and consumer coordination. Technical report, CESifo working papers.

An, M. Y. (1996). Log-concave probability distributions: Theory and statistical testing. *Centre for Labour Economics at University of Aarhus and Aarhus School of Business 96*(1).

Argentesi, E. and L. Filistrucchi (2007). Estimating market power in a two-sided market: The case of newspapers. *Journal of Applied Econometrics 22*(7), 1247–1266.

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.

Bagnoli, M. and T. Bergstrom (2005). Log-concave probability and its applications. *Economic Theory 26*(2), 445–469.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica: Journal of the Econometric Society 60*(4), 889–917.

Bresnahan, T. F. and P. C. Reiss (1990). Entry in monopoly market. *Review of Economic Studies 57*(4), 531–553.

Bresnahan, T. F. and P. C. Reiss (1991). Entry and competition in concentrated markets. *Journal of Political Economy 99*(5), 977–1009.

Brock, W. A. and S. N. Durlauf (2001). Discrete choice with social interactions. *Review of Economic Studies 68*(2), 235–260.

Caillaud, B. and B. Jullien (2003). Chicken & egg: Competition among intermediation service providers. *RAND Journal of Economics*, 309–328.

Chao, Y. and T. Derdenger (2013). Mixed bundling in two-sided markets in the presence of installed base effects. *Management Science 59*(8), 1904–1926.

Damiano, E. and H. Li (2007). Price discrimination and efficient matching. *Economic Theory 30*(2), 243–263.

Hagiu, A. (2006). Pricing and commitment by two-sided platforms. *RAND Journal of Economics 37*(3), 720–737.

Heckman, J. J. and B. E. Honore (1990). The empirical content of the roy model. *Econometrica: Journal of the Econometric Society*, 1121–1149.

Hotz, V. J. and R. A. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies 60*(3), 497–529.

Katz, M. L. and C. Shapiro (1985). Network externalities, competition, and compatibility. *American Economic Review 73*(3), 424–440.

Katz, M. L. and C. Shapiro (1986). Technology adoption in the presence of network externalities. *Journal of Political Economy 94*(4), 822–841.

Laffont, J.-J., P. Rey, and J. Tirole (1998a). Network competition: I. overview and nondiscriminatory pricing. *RAND Journal of Economics 29*(1), 1–37.

Laffont, J.-J., P. Rey, and J. Tirole (1998b). Network competition: Ii. price discrimination. *RAND Journal of Economics 29*(2), 38–56.

Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review 103*(7), 2960–3000.

Mazzeo, M. J. (2002). Product choice and oligopoly market structure. *RAND Journal of Economics 33*(2), 221–242.

Pakes, A., M. Ostrovsky, and S. Berry (2007). Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *RAND Journal of Economics 38*(2), 373–399.

Pesendorfer, M. and P. Schmidt-Dengler (2008). Asymptotic least squares estimators for dynamic games. *Review of Economic Studies 75*(3), 901–928.

Rochet, J.-C. and J. Tirole (2002). Cooperation among competitors: Some economics of payment card associations. *Rand Journal of Economics 33*(4), 549–570.

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 37*(3), 645–667.

Rust, J. (1996). Estimation of dynamic structural models, problems and prospects: discrete decision processes. In *Advances in econometrics: Sixth world congress*, Volume 2, pp. 119–170. Cambridge Univ Pr.

Rysman, M. (2004). Competition between networks: A study of the market for yellow pages. *Review of Economic Studies 71*(2), 483–512.

Rysman, M. (2007). An empirical analysis of payment card usage. *Journal of Industrial Economics 55*(1), 1–36.

Rysman, M. (2009). The economics of two-sided markets. *Journal of Economic Perspectives 23*(3), 125–143.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics 37*(3), 619–640.

Stavins, J. (1995). Model entry and exit in a differentiated-product industry: The personal computer market. *Review of Economics and Statistics 77*(4), 571–584.

Toivanen, O. and M. Waterson (2005). Market structure and entry: Where's the beef? *RAND Journal of Economics 36*(3), 175–191.

Vitorino, M. A. (2012). Empirical entry games with complementarities: An application to the shopping center industry. *Journal of Marketing Research 49*(2), 175–191.

Weyl, E. G. (2010). A price theory of multi-sided platforms. *American Economic Review 100*(4), 1642–1672.

Zhou, Y. (2013). Failure to launch in two-sided markets: A study of the us video game market. *SSRN Scholarly Paper ID 2163948*.

Zhu, T. and V. Singh (2009). Spatial competition with endogenous location choices: An application to discount retailing. *Quantitative Marketing and Economics 7*(1), 1–35.

Appendix

Proposition 1. (Equilibrium)

Proof: I establish the existence of the equilibrium by first proving the following lemmas.

Lemma 1. A seller joins Platform S if and only if its quality satisfies the following condition

$$\ln\frac{1-t_S}{1-t_{NS}}+\frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}}+E(\theta|\mu,\ k_S)-E(\theta|\mu,\ k_{NS})+\ln(1-F_{\varepsilon}(k_S-\mu-\theta))\geq 0$$

Proof:

Seller j's profit on Platform S can be expressed as

$$\pi_{j,S}(\mu_j,\theta_j;\,k_S,t_S)=p*d_{j,S}*(1-t_S)*\Pr(\varepsilon_j\geq k_S-\mu_j-\theta_j)$$

and seller j's profit on Platform NS equals to

$$\pi_{j,NS}(\mu_j,\theta_j;\,k_{NS},t_{NS})=p*d_{j,NS}*(1-t_{NS})$$

Therefore, the seller enters Platform S if and only if the log-profit on Platform S is larger than that on Platform NS, i.e.,

$$\begin{aligned}\triangle\ln\pi_j&=\ln\pi_{j,s}(\mu_j,\theta_j;\,k_S,t_s)-\ln\pi_{jns}(\mu_j,\theta_j;\,k_{NS},t_{ns})\\&=\ln\frac{1-t_S}{1-t_{NS}}+\ln\frac{d_{j,S}}{d_{j,NS}}+\ln(1-F_{\varepsilon}(k_S-\mu_j-\theta_j))\\&\geq 0\end{aligned}\tag{22}$$

Recall the demand function of seller j on market m is

$$\begin{aligned}d_{j,m}&=d_{j|j\in\mathcal{J}(m)}*d_m\\&=\frac{\exp[u_{j,m}]}{\sum_{\text{all }j'\,\in\,\mathcal{J}(m)}\exp[u_{j',m}]}\frac{\exp\left[\lambda EU_m\right]}{1+\sum_{\,m'\in\{S,NS\}}\exp\left[\lambda EU_{m'}\right]}\\&=\exp[E(q_j|\mu_j,\,k_m)-price_j]\frac{\exp\left[(\lambda-1)EU_m\right]}{1+\sum_{\,m'\in\{S,NS\}}\exp\left[\lambda EU_{m'}\right]}\end{aligned}$$

. Besides,

$$\ln\frac{d_S}{d_{NS}}=\lambda(EU_S-EU_{NS})$$

The difference of log market shares on Platform S and NS, therefore, equals to:

$$\begin{aligned}
\ln\frac{d_{j,S}}{d_{j,NS}} &= E(q|\mu_j,\, k_S) - E(q|\mu_j,\, k_{NS}) + (\lambda-1)(EU_S - EU_{NS}) \\
&= E(q|\mu_j,\, k_S) - E(q|\mu_j,\, k_{NS}) + \frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}} \\
&= E(\theta|\mu_j, k_S) - E(\theta|\mu_j, k_{NS}) + \frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}} \qquad (23)
\end{aligned}$$

where the last equality follows the consumer's belief of the seller's unobservable quality on the two platforms

$$E(q|\mu_j,\, k_S) = \mu_j + E(\theta|\mu_j, k_S)$$

and

$$E(q|\mu_j,\, k_{NS}) = \mu_j + E(\theta|\mu_j, k_{NS})$$

Combine (22) and (23), a seller with quality pair $(\mu,\, \theta)$ joins Platform S if and only if

$$\ln\frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}} + E(\theta|\mu,\, k_S) - E(\theta|\mu,\, k_{NS}) + \ln(1 - F_\varepsilon(k_S - \mu - \theta)) \geq 0 \qquad (24)$$

Q.E.D

Lemma 2. When the observable quality $\mu \in (-\infty,\, k_S - \underline{\theta} - \underline{\varepsilon})$, a seller chooses Platform S if and only if the unobservable quality θ is larger than the threshold $\theta^*(\mu)$, where $\theta^*(\mu)$ is the minimum point such that

$$\ln\frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}} + E(\theta|\theta \geq \theta^*(\mu)) - E(\theta|\theta \leq \theta^*(\mu)) + \ln(1 - F_\varepsilon(k_S - \mu - \theta^*(\mu))) \geq 0$$

If $\theta^*(\mu) \in [\underline{\theta},\, \bar{\theta}]$, there is a positive measure of type-μ sellers on Platform S. If $\theta^*(\mu) > \bar{\theta}$, none of type-$\mu$ sellers enter Platform S. If $\theta^*(\mu) < \underline{\theta}$, all type-$\mu$ sellers participate in Platform S.

Proof: We only prove the first part of Lemma 2. The second part is obvious.

As shown in Lemma 1, a type-μ seller enters Platform S if and only if the inequality specified in (24) holds. When $\mu + \underline{\theta} + \underline{\varepsilon} < k_S$, $\ln(1 - F_\varepsilon(k_S - \mu - \theta))$ is a non-decreasing with respect to θ, so does the left-hand side of (24). Moreover, since $\ln(1 - F_\varepsilon(k_S - \mu - \theta))$ is bounded within $(-\infty,\, 0]$, there exists a minimum point $\theta^*(\mu) \in \mathbb{R}$ such that

$$\ln\frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}} + E(\theta|\mu,\, k_S) - E(\theta|\mu,\, k_{NS}) + \ln(1 - F_\varepsilon(k_S - \mu - \theta^*(\mu))) \geq 0$$

if

$$\ln\frac{1-t_S}{1-t_{NS}}+\frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}}+\sup\{E(\theta|\mu,\,k_S)-E(\theta|\mu,\,k_{NS})\}>0$$

. Correspondingly, the consumer's equilibrium belief of θ takes a form as follows:

$$E(\theta|\mu,\,k_S)=E(\theta|\theta\geq\theta^*(\mu))$$

and

$$E(\theta|\mu,\,k_{NS})=E(\theta|\theta\leq\theta^*(\mu))$$

and $\theta^*(\mu)$ is determined by

$$\ln\frac{1-t_S}{1-t_{NS}}+\frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}}+E(\theta|\theta\geq\theta^*(\mu))-E(\theta|\theta\leq\theta^*(\mu))+\ln(1-F_\varepsilon(k_S-\mu-\theta^*(\mu)))=0$$

Q.E.D

Lemma 3. If the observable quality $\mu\in(k_S-\underline{\theta}-\underline{\varepsilon},+\infty)$, consumers' belief of θ equals to

$$E(\theta|\mu,k_S)=E(\theta|\mu,k_{NS})=0$$

and the seller enters Platform S if

$$\ln\frac{1-t_S}{1-t_{NS}}+\frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}}>0$$

and otherwise it chooses Platform NS.

Proof:

When $\mu\in(k_S-\underline{\theta}-\underline{\varepsilon},+\infty)$, for any $\theta\in[\underline{\theta},\,\bar{\theta}]$ and $\varepsilon\in[\underline{\varepsilon},\,\bar{\varepsilon}]$, we have

$$\Pr(\mu+\theta+\varepsilon\geq k_S)=1$$

, which means the seller of type-μ can always pass the quality screening set by Platform S, no matter what the unobservable quality θ is. Anticipating this, consumers form the belief of θ which equals to the population mean:

$$\mathbb{E}(\theta|\mu,k_S)=\mathbb{E}(\theta|\mu,k_{NS})=\mathbb{E}(\theta)=0$$

.

Given consumers' belief of θ, the seller's comparative profit on Platform S can be written as

$$\triangle\ln\pi=\ln\frac{1-t_S}{1-t_{NS}}+\frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}}$$

. Therefore, if $\ln \frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda} \ln \frac{d_S}{d_{NS}} > 0$, the seller enters Platform S, i.e., $\theta^*(\mu) = \underline{\theta}$. Otherwise, the seller enters Platform NS, i.e., $\theta^*(\mu) = \overline{\theta}$.

Q.E.D.

Using Lemma 1-3, now I prove the Proposition 1. Given two platforms' prices $\{t_S, t_{NS}\}$ and quality screening policies $\{k_S, -\infty\}$, the seller chooses a platform, in anticipation of $\{d_S, d_{NS}\}$ and consumers' belief $\mathbb{E}(\theta|\mu, k_m)$. It enters Platform S iff

$$\theta_j \quad \geq \quad \theta^*(\mu_j, d_S, d_{NS};\, t_S, t_{NS}, k_S)$$

After all N sellers make decisions, the numbers of sellers on the two platforms, N_S and N_{NS} are known where

$$N_S = \sum_{j=1}^{N} 1(\theta_j \geq \theta^*(\mu_j, d_S, d_{NS};\, t_S, t_{NS}, k_S))$$

and

$$N_{NS} = N - N_S$$

and the sets of sellers on the two platforms $\mathcal{J}(S)$ and $\mathcal{J}(NS)$ are also available where

$$\mathcal{J}(S) = \{\text{all } j \text{ s.t } \theta_j \geq \theta^*(\mu_j, d_S, d_{NS};\, t_S, t_{NS}, k_S)\}$$

and

$$\mathcal{J}(NS) = \{\text{all } j \text{ s.t } \theta_j < \theta^*(\mu_j, d_S, d_{NS};\, t_S, t_{NS}, k_S)\}$$

Observing $\{N_S, N_{NS}\}$ and $\{\mathcal{J}(S), \mathcal{J}(NS)\}$, consumers can calculate the utility provided by platforms and sellers and make choices. According to (4) and (6), the number of consumers on two platforms can be expressed as follows:

$$\begin{aligned} \begin{bmatrix} d_S \\ d_{NS} \end{bmatrix} &= \begin{bmatrix} G_1(\{N_S, N_{NS}\}, \{\mathcal{J}(S), \mathcal{J}(NS)\};\, t_S, t_{NS}, k_S) \\ G_2(\{N_S, N_{NS}\}, \{\mathcal{J}(S), \mathcal{J}(NS)\};\, t_S, t_{NS}, k_S) \end{bmatrix} \\ &= G(\theta^*(\mu_j, d_S, d_{NS};\, t_S, t_{NS}, k_S); j = 1, ...N) \end{aligned}$$

, where the second equality is followed by the determinations $\{N_S, N_{NS}\}$ and $\{\mathcal{J}(S), \mathcal{J}(NS)\}$. So the equilibrium distribution of consumers on two platforms is a fixed point defined by this equation. Since $d_S, d_{NS} \in [0, 1)$, by Brouwer Fixed Point Theorem, there must exist a pair of $[d_S^*, d_{NS}^*]$ satisfying above equation.

Also in equilibrium consumers make correct inference about the private information θ:

$$\mathbb{E}(\theta|\mu, k_S) = \mathbb{E}(\theta|\theta \geq \theta^*(\mu, d_S^*, d_{NS}^*; t_S, t_{NS}, k_S)$$

and

$$\mathbb{E}(\theta|\mu, k_{NS}) = \mathbb{E}(\theta|\theta < \theta^*(\mu, d_S^*, d_{NS}^*; t_S, t_{NS}, k_S)$$

Q.E.D

Proposition 2. (Comparative Statics)

Proof:

According to Proposition 1, when $\mu \in (-\infty,\ k_S - \underline{\theta} - \underline{\varepsilon})$ and $\varepsilon \sim \mathrm{Exp}(l)$, $\theta^*(\mu)$ is determined by

$$\begin{aligned} 0 &= \ln\frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}} + E(\theta|\theta \geq \theta^*(\mu)) - E(\theta \leq \theta^*(\mu)) + \ln(1 - F_\varepsilon(k_S - \mu - \theta^*(\mu))) \\ &= \ln\frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda}\ln\frac{d_S}{d_{NS}} + E(\theta|\theta \geq \theta^*(\mu)) - E(\theta \leq \theta^*(\mu)) + l\,(\mu + \theta^*(\mu) - k_S) \end{aligned}$$

By taking partial derivative with respect to μ and θ, we get

$$\frac{\partial\theta^*(\mu)}{\partial\mu} = -\frac{l}{\frac{\partial E(\theta|\theta\geq x)}{\partial x}|_{x=\theta^*(\mu)} - \frac{\partial E(\theta|\theta\leq x)}{\partial x}|_{x=\theta^*(\mu)} + l}$$

Since θ follows log-concave distribution, we know

$$0 \leq \frac{\partial E(\theta|\theta \geq x)}{\partial x} \leq 1$$

and

$$0 \leq \frac{\partial E(\theta|\theta \leq x)}{\partial x} \leq 1$$

Therefore,

$$l - 1 \leq \frac{\partial E(\theta|\theta \geq x)}{\partial x}|_{x=\theta^*(\mu)} - \frac{\partial E(\theta|\theta \leq x)}{\partial x}|_{x=\theta^*(\mu)} + l \leq l + 1$$

.Since $l \geq 1$,

$$\frac{\partial\theta^*(\mu)}{\partial\mu} < 0$$

When $\mu \in (k_S - \underline{\theta}, +\infty)$, as shown by Lemma 3, the threshold is given by

$$\theta^*(\mu) = \begin{cases} \underline{\theta} & \text{if } \ln\frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda}\ln\frac{d_S^*}{d_{NS}^*} \geq 0 \\ \bar{\theta} & \text{if } \ln\frac{1-t_S}{1-t_{NS}} + \frac{(\lambda-1)}{\lambda}\ln\frac{d_S^*}{d_{NS}^*} < 0 \end{cases}$$

Q.E.D

Table 1: Variable Statistics

	89 Camera Models				
	Mean	σ	Min	Max	N
Number of Sellers	558.88	285.6	101	1272	35384
Rating	4.61	1.15	0	5	35384
Prices (RMB)	8405.12	7803.58	1101	99999	35384
Tmall (0/1)	.0986	.2982	0	1	35384
Monthly Sales	2.19	21.09	0	1220	35384
Positive Sales (0/1)	.1284	.3345	0	1	35384
Market Share	.00125	.00975	0	.466	35384

Notes: Market share is defined as $\frac{\text{seller j's sales of camera t}}{10{,}000}$ under the assumption that in the data period there are 10,000 potential consumers for each camera model. The variable of positive sales is a dummy which indicates whether the seller has non-zero sales.

Table 2: Comparisons between Tmall and Taobao

	Variable: Prices				
	Mean	σ	Min	Max	N
Tmall	8044	6857	2080	51479	3492
Taobao	8444	7899	1	99999	31892

	Variable: Individual Sales				
	Mean	σ	Min	Max	N
Tmall	4.61	23.14	0	389	3492
Taobao	2.08	20.98	1	1220	31892

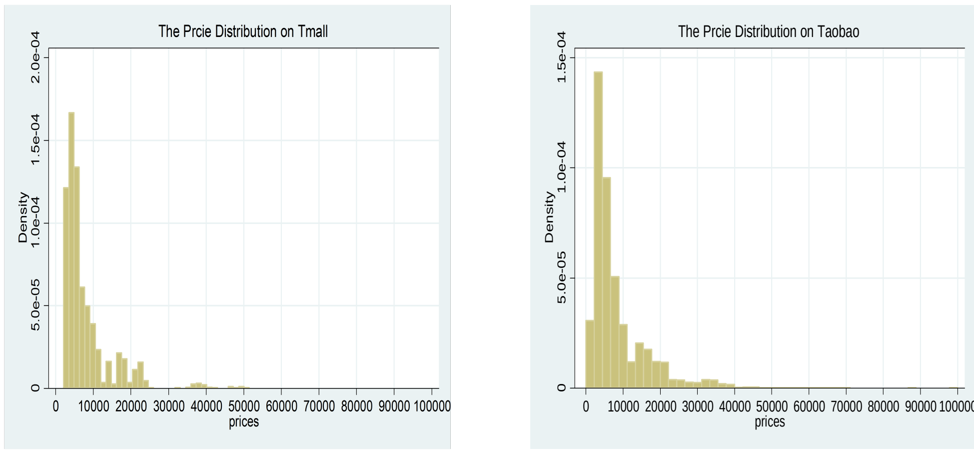
	Variable: Positive Sales (0/1)				
	Mean	σ	Min	Max	N
Tmall	.2860	41.10	0	1	3492
Taobao	.1210	.3261	0	1	31892

	Variable: Individual Market Share				
	Mean	σ	Min	Max	N
Tmall	.002572	.01171	0	.3240	3492
Taobao	.001114	.0095	0	.4667	31892

	Variable: Total Sales on the Platform				
	Mean	σ	Min	Max	N
Tmall	383.8	430.3	2	1815	89
Taobao	1200.3	1628.8	5	7696	89

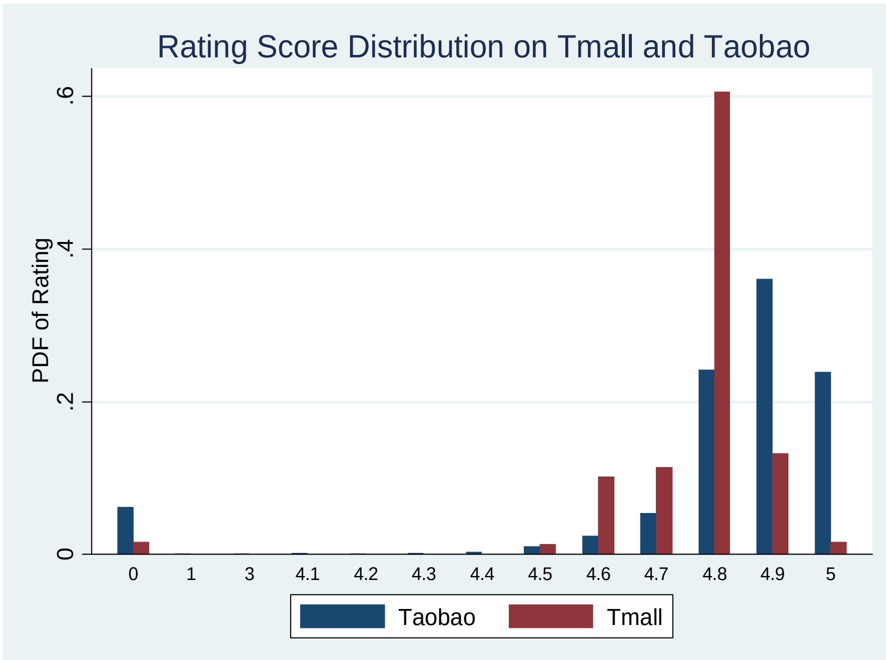
Notes: The first four tables show the statistics of variables of individual sellers located on Tmall and Taobao. The last table presents the statistics of the total sales of the 89 camera models on Tmall and Taobao.

Figure 1: **The Distribution of Prices on Tmall and Taobao**



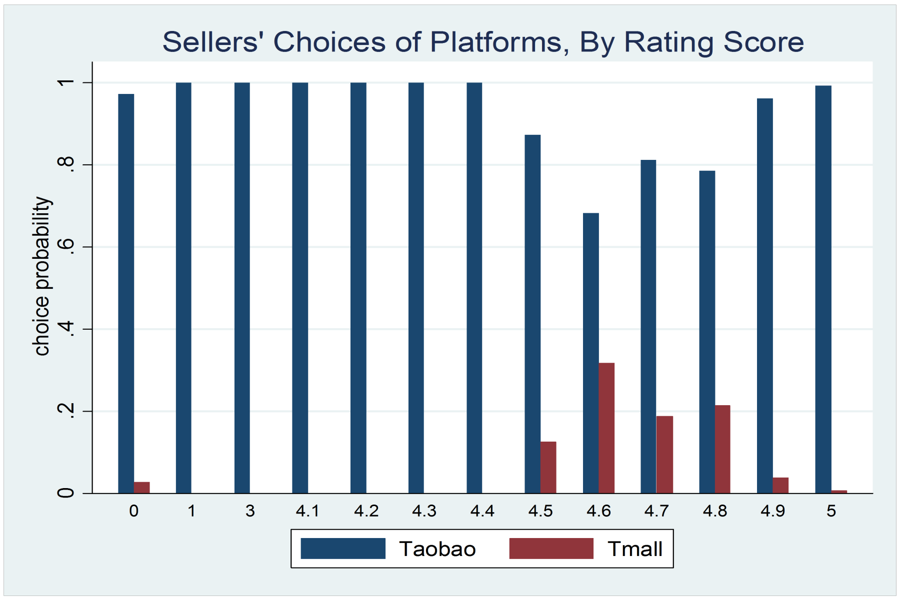
Notes: The figure on the left displays the distribution of prices on Tmall; The right figure shows the distribution of prices on Taobao; The x-axis is the value of prices and the y-axis represents the density of prices.

Figure 2: **The Distribution of Rating Score on Tmall and Taobao**



Notes: The figure displays the distribution of sellers' rating scores by platform. The x axis is the rating score. The blue bar stands for the proportion of sellers with corresponding rating score in the population of Taobao. The red bars present the distribution of the rating score for sellers located on Tmall.

Figure 3: **Sellers' Choices of Platforms, By Rating Score**



Notes: This figure displays sellers' entry decisions by the rating score. The x-axis is the rating score. For each value of rating score, the blue bar represents the proportion of sellers that choose Taobao, and the red bar displays the proportion of sellers on Tmall.

Table 3: **The Proportion of Tmall Sellers, By Rating Group**

	Group of Rating Score		
	Group_1	Group_2	Group_3
Rating Score	$[0 \sim 4.7)$	$[4.7 \sim 4.9)$	$[4.9 \sim 5.0]$
Tmall Sellers (%)	12.13	18.35	1.03
N	3,785	15,620	15,979

Notes: The sellers, according to their rating scores, are classified into three groups: Group_1 is composed by sellers with rating scores less than 4.7; Group_2 sellers have rating scores between 4.7 and 4.9 and Group_3 sellers' rating scores are equal or larger than 4.9. This table presents the proportion of Tmall sellers and the number of observations in each rating group.

Table 4: **Comparisons between Tmall and Taobao, By Rating Group**

	Prices							
	Tmall				Taobao			
	Mean	σ	Min	Max	Mean	σ	Min	Max
Rating Group_1	7870.1	7072.9	2080	48688	8401.2	7634.1	1	88888
Rating Group_2	8015.6	6717.1	1120	51479	7887.4	7476.4	1120	99999
Rating Group_3	9072.1	8446.1	2398	49000	8903.1	8249.8	499	99999

	Individual Monthly Sales							
	Tmall				Taobao			
	Mean	σ	Min	Max	Mean	σ	Min	Max
Rating Group_1	1.113	3.482	0	38	.04239	.5101	0	25
Rating Group_2	5.552	26.759	0	393	1.592	14.83	0	677
Rating Group_3	.6667	3.628	0	29	2.579	26.16	0	1220

	Positive Sales (0/1)							
	Tmall				Taobao			
	Mean	σ	Min	Max	Mean	σ	Min	Max
Rating Group_1	.2527	.4350	0	1	.02585	.1587	0	1
Rating Group_2	.3017	.4591	0	1	.1394	.3464	0	1
Rating Group_3	.1212	.3274	0	1	.1061	.3079	0	1

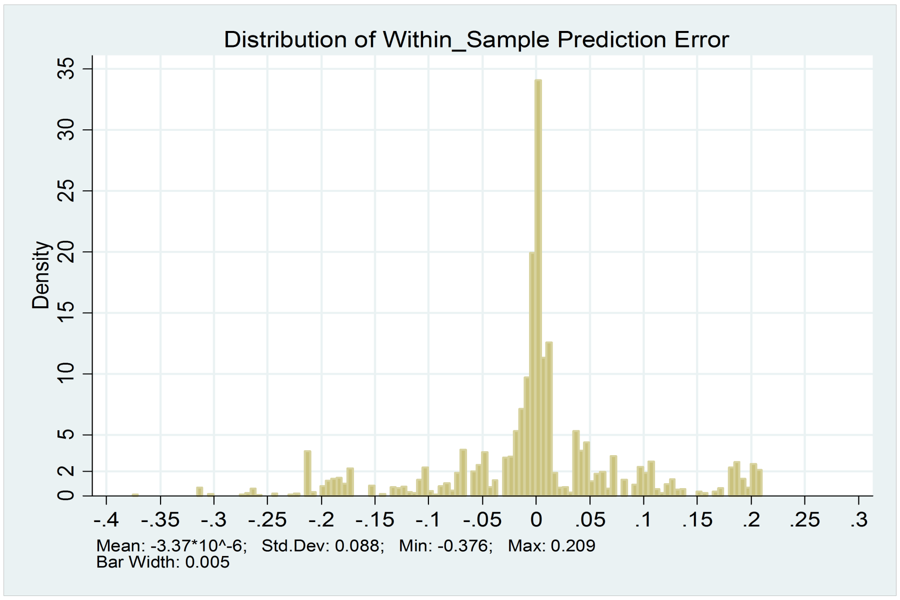
	Individual Market Share							
	Tmall				Taobao			
	Mean	σ	Min	Max	Mean	σ	Min	Max
Rating Group_1	.001449	.007286	0	.1042	.0001405	.001840	0	.07143
Rating Group_2	.002885	.01256	0	.3241	.001234	.008414	0	.3333
Rating Group_3	.0002654	.001202	0	.007576	.001221	.01113	0	.4667

Table 5: **Estimation Results**

	Coefficients (Std. Error)
Variables of Demand Function	
Rating Group_2	3.049^{***} (0.0710)
Rating Group_3	1.583^{***} (0.0089)
Price (1000 RMB)	-0.0514^{***} (0.0181)
$\mathbb{E}(\theta\|X,m)$	2.598^{***} (0.0334)
Rating Group_2*$\mathbb{E}(\theta\|X,m)$	-0.7712^{***} (0.0316)
Rating Group_3*$\mathbb{E}(\theta\|X,m)$	-2.863^{***} (0.0501)
$\ln EU(m)$	1.714^{***} (0.0601)
Constant	-19.27^{***} (0.49)
Variables in Entry Function	
$\mathbb{E}\left[\ln \frac{d_{j,S}}{d_{j,NS}}\right]$	0.6179^{***} (0.0387)
Rating Group_2	1.296^{***} (0.0547)
Rating Group_3	2.333^{***} (0.0184)
Constant	-4.821^{***} (0.1667)

Notes: Standard errors are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. $\mathbb{E}(\theta|X,m)$ stands for the consumer's expectation of θ conditional on the seller's observable rating score, X and the platform m on which the sellers is located. $EU(m)$ is the consumer's expected maximum utility from platform m. $d_{j,m}$ is seller j's market share on platform m.

Figure 4: **Distribution of Prediction Errors, Sellers' Choice Probabilities**



Notes: The prediction error is defined as the difference between the seller's choice probability predicted by the model and the probability observed in the data. The statistics of prediction errors are presented at the bottom of the figure.

Table 6: **Counterfactual: Change of Consumer Utility**

	Mean	σ	Min	Max	N (camera models)
CU	2.86	1.89	.562	7.99	89
$1(CU > 1)$	.742	.441	0	1	89

Notes: The variable CU is defined as the ratio of total utility that sellers provide to consumers before and after Alibaba would remove the quality screening on Tmall. The number of camera models in the data is 89.