



## Marketing Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

# Quantifying Cross and Direct Network Effects in Online Consumer-to-Consumer Platforms

Junhong Chu, Puneet Manchanda

To cite this article:

Junhong Chu, Puneet Manchanda (2016) Quantifying Cross and Direct Network Effects in Online Consumer-to-Consumer Platforms. Marketing Science 35(6):870-893. <https://doi.org/10.1287/mksc.2016.0976>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Quantifying Cross and Direct Network Effects in Online Consumer-to-Consumer Platforms

Junhong Chu

NUS Business School, National University of Singapore, Singapore 119245, [bizcj@nus.edu.sg](mailto:bizcj@nus.edu.sg)

Puneet Manchanda

Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109, [pmanchan@umich.edu](mailto:pmanchan@umich.edu)

Consumer-to-consumer (C2C) platforms have become a major engine of growth in Internet commerce. This is especially true in countries such as China, which are experiencing a big rush toward e-commerce. The emergence of such platforms gives researchers the unique opportunity to investigate the evolution of such platforms by focusing on the growth of both buyers and sellers. In this research, we build a utility-based model to quantify both cross and direct network effects on Alibaba Group's Taobao.com, the world's largest online C2C platform (based in China). Specifically, we investigate the relative contributions of different factors that affect the growth of buyers and sellers on the platform. Our results suggest that the direct network effects do not play a big role in the platform's growth (we detect a small positive direct network effect on buyer growth and no direct network effect on seller growth). More importantly, we find a significant, large and positive cross-network effect on both sides of the platform. In other words, the installed base of either side of the platform has propelled the growth of the other side (and thus the overall growth). Interestingly, this cross-network effect is asymmetric with the installed base of sellers having a much larger effect on the growth of buyers than vice versa. The growth in the number of buyers is driven primarily by the seller's installed base and product variety with increasing importance of product variety. The growth in the number of sellers is driven by buyer's installed base, buyer quality, and product price with increasing importance of buyer quality. We also investigate the nature of these cross-network effects over time. We find that the cross-network effect of sellers on buyers increases and then decreases to reach a stable level. By contrast, the cross-network effect of buyers on sellers is relatively stable. We discuss the policy implications of these findings for C2C platforms in general and Taobao in particular.

Data, as supplemental material, are available at <https://doi.org/10.1287/mksc.2016.0976>.

**Keywords:** platforms; two-sided markets; cross-network effect; direct network effect; e-commerce; emerging markets; China

**History:** Received: September 13, 2014; accepted: November 14, 2015; Preyas Desai served as the editor-in-chief and Michel Wedel served as associate editor for this article. Published online in *Articles in Advance* July 6, 2016.

## 1. Introduction

Consumer-to-consumer (C2C) platforms such as eBay, Amazon's Marketplace, Taobao.com, and OLX.in have become a major engine of growth in e-commerce. This is especially true in countries such as China that are experiencing a big rush toward e-commerce. The emergence of such platforms represents a new phenomenon because they have scaled up to very large numbers very quickly. For example, the Chinese C2C network, Taobao.com, had 435 million consumers participating as buyers and 7.1 million as sellers in less than a decade after its formation in 2003. The factors that have enabled this growth and size have been novel revenue generating mechanisms, e.g., charging sellers only for value-added services, and the platforms' agnostic attitude toward product assortment, allowing buyers and sellers to make choices on what to offer. Although there is a rich body of work on platform economies and

two-sided markets, starting with the pioneering work of Rochet and Tirole (2003), the focus has typically been on platform competition, pricing structure, and business model determination (e.g., Caillaud and Julien 2003, Armstrong 2006, Rochet and Tirole 2006) and less on the factors determining platform evolution and growth. In addition, most empirical work on platform markets has usually been set in "conventional" or offline markets, such as VCRs, game consoles, personal digital assistants (PDAs), media (TV, newspaper, and magazines), payment systems, and yellow pages (e.g., Ohashi 2003, Rysman 2004, Nair et al. 2004, Clements and Ohashi 2005, Wilbur 2008, Dubé et al. 2010, Liu 2010, Sun et al. 2015). Much of the extant research on online C2C platforms such as eBay has focused on the auction mechanism and recommendation system, rather than on the evolution and growth of the platform.

In this paper, we focus on the evolution and growth of online C2C platforms. Specifically, we investigate the evolution of the platform both from the buyer's and the seller's perspective as well as the nature of buyer and seller interactions over the platform's life cycle. We look at the following novel questions. First, how large is the cross-network effect (CNE) on both sides of the platform? As for any network, the growth and evolution of one side has a direct impact—the CNE—on the growth and evolution of the other side. Our objective is to quantify the CNEs—the impact of the installed base of sellers on the growth of buyers and the impact of the installed base of buyers on the growth of sellers. Second, this quantification leads us to discover whether the two CNEs are asymmetric. This asymmetry, if it exists, allows us to pinpoint the side of the platform that is more important for the overall growth (of the platform). Third, in two-sided markets where CNEs are likely to exist because the platform has no stand-alone value to either side of the market, can we detect a direct network effect (DNE) on each side of the platform? If there is a DNE, how large is it vis-à-vis CNEs? Fourth, we examine how the nonnetwork factors (e.g., product variety, product price, and buyer quality) affect the growth of the two sides of the network. We contrast the effect of nonnetwork factors with the network effects toward the growth of the network. Finally, we allow both the CNEs and nonnetwork effects to vary over time beginning from the platform's inception.

To do this, we exploit a new data set from Alibaba Group's <http://www.taobao.com>, the world's largest online C2C platform based in China. Taobao.com (referred to as Taobao for the rest of the paper) has 7.1 million sellers and 435 million buyers (as of December 2012). Each day there are 728 million unique items on the "shelf" for sale and 75 million unique viewers, generating 13 million transactions and 1.61 billion yuan (USD \$258 million)<sup>1</sup> in revenues. A major distinguishing feature of our data set is that our data start from the first day of Taobao's operations (May 11, 2003). Our data set contains daily observations on the number of platform participants, the assortment of products on offer, and the revenue from buyers and sellers. Interestingly, Taobao allows both buyers and sellers to participate for free on the platform. Industry reports (e.g., Morningstar 2014) have noted that the rapid growth of the platform is due to the *strong network effect* (italics ours) where the value

of the platform to consumers increases with a greater number of sellers (and vice versa).

Using a utility-based approach to model buyer's and seller's platform joining decisions jointly, we identify a large, significant, and positive CNE on both sides of the platform market. However, we find that the CNE is asymmetric: the installed base of sellers has a much larger effect on the growth of buyers than vice versa, implying that the platform's growth is driven more by sellers than by buyers. There is also a small positive and significant DNE on the buyer's side, and a negative but insignificant DNE on the seller's side. Furthermore, the growth in the number of buyers is driven primarily by the seller's installed base and product variety with increasing importance of product variety over time. By contrast, the growth in the number of sellers is driven by the buyer's installed base, buyer quality, and product price with increasing importance of buyer quality over time. The two CNEs demonstrate different temporal patterns. Specifically, the CNE of sellers on buyers increases and then decreases to reach steady state. By contrast, the CNE of buyers on sellers is relatively stable. We examine the policy implication of our findings.

Overall, our paper makes the following contributions to the two-sided markets literature. First, the paper is one of the few papers that is able to pin down the CNE and DNE in one holistic model. Second, the CNE is allowed to be time varying and estimated from the inception of the network, something that is novel to the literature. Third, there is also little work that quantifies the relative importance of one side over the other side of the network. Our finding, that sellers play a much bigger role in growing such networks, has implications for both academics and managers. Fourth, we are also able to estimate the effect of nonnetwork factors on the platform's growth. Fifth, most extant work on two-sided markets has focused on the role of price in growing the network. Our setting is unique in that the platform charges zero price for participation (for both the buyer and the seller). Thus, the platform has no direct instrument to enable growth. In such settings, it is important to understand the drivers of growth. Finally, our paper sheds new light on a business model—C2C—that is becoming increasingly prevalent in e-commerce markets.

The rest of the paper is organized as follows. In §2, we review the relevant literature. We describe the institutional setting in §3, set up the econometric model in §4, and summarize the data and explain the variable operationalization in §5, and describe the estimation in §6. We report the main findings, results of the robustness checks, and managerial implications in §7 and conclude in §8.

<sup>1</sup> For the sake of exposition, we use 6.23 yuan to \$1 USD as the exchange rate (the rate reported at xe.com on December 31, 2012) throughout the paper. This rate was around 8.50 yuan at the time of Taobao's inception, dropped to 6.8 yuan in 2008 and then was steady till about 2010, and then declined to 6.23 yuan at the end of 2012 (all data from xe.com).

## 2. Literature Review

Research on two-sided markets has a relatively short history (see Rysman 2009 and Sriram et al. 2015 for an overview). Rochet and Tirole (2003, 2006), Caillaud and Jullien (2003), and Armstrong (2006) each provide a theoretical framework for two-sided markets to explain how the price structure is determined when either a monopoly platform sets prices or two platforms compete. A common feature present in all this theory work is that the benefit of joining a platform for any agent depends on the total number of agents from the other side on the same platform. This relationship can be summarized by the CNE, testifying to the importance of the existence and magnitude of the CNE. Our paper does not investigate the price structure because the platform under study adopts free pricing for both sides of the market; however, the theoretical work cited here guides in determining the drivers of platform growth as well as the possible functional forms for capturing buyer and seller utility.

The empirical work on network effects is somewhat limited, though growing at a rapid pace.<sup>2</sup> One stream (e.g., Shankar and Bayus 2003, Ohashi 2003, Park 2004) has focused on direct network effects. In other words, the estimated network effect quantifies the benefit (or cost) that agents obtain from the presence of other agents on the same side, rather than those on the complementary side (usually due to lack of data). Gandal et al. (2000) are among the first to explicitly model cross-network externalities. They measure the effect of hardware prices and software titles in the diffusion of CD players, and find that a 10% increase in CD titles would have as large an effect as a 5% price cut. Rysman (2004) estimates the importance of CNEs in the market for yellow pages and finds two-way positive cross-network externalities whereby advertisers value consumer usage and consumers value advertising. Akerberg and Gowrisankaran (2006) estimate the size and importance of network externalities in the automated clearing house (ACH) banking industry, and find that most of the impediments to ACH adoption is from the large customer fixed cost of adoption. Wilbur (2008) explicitly models the two-way cross-network interactions in the television industry and finds a negative effect of the number of advertisements on audience size (viewers are ad averse) and a positive

effect of audience size on advertiser demand (advertisers are viewer loving). In contrast to this literature, we estimate two-way CNEs and two-way DNEs in two-sided markets and compare their relative magnitudes.

A few studies have extended this literature by quantifying the evolution of cross-network externalities. Nair et al. (2004) quantify the size of CNEs in the PDA market with competing incompatible technology standards, and find significant and growing effect of software provision on hardware adoption. Clements and Ohashi (2005) measure the effects of hardware price and software variety in the diffusion of video game systems in the U.S. market between 1994 and 2002. They find that introductory pricing is an effective practice at the beginning of the product cycle, and expanding software variety becomes more effective later.

Our paper is closely related to these two studies, but with some notable differences. First, our empirical context is a unique and different two-sided market, i.e., a monopoly C2C online platform that has adopted free pricing for both sides of the market throughout the platform's life cycle. Thus the drivers for the platform's growth are very different from other platforms. Second, we explicitly quantify two-way CNEs, i.e., buyers on sellers and sellers on buyers, and their evolution over the platform's life cycle, and thus we are able to pin down which side of the platform is more important for the platform's growth. Nair et al. (2004) and Clements and Ohashi (2005) focus on one side, i.e., software titles on hardware adoptions, so they are not able to pinpoint the relative importance of one side over the other. Third, we measure CNEs together with DNE for the same platform, and find that CNEs are much larger than DNE, whereas they only measure CNEs. Ours is the first to examine both CNEs and DNEs in the same framework. Fourth, in contrast to their separate estimation of the two sides, we model the decisions of the two sides *jointly*. Fifth, our finding on the relative importance of network factors versus nonnetwork factors provides qualitative new insights. For example, both Nair et al. (2004) and Clements and Ohashi (2005) find an *increasing* importance of CNE in the diffusion of hardware over the platform's life cycle. By contrast, we find a *decreasing* importance of network factors and increasing importance of nonnetwork factors in the growth of the platform. Therefore, our research adds to and complements the literature on the network effects and their evolution.

## 3. Institutional Setting

As noted earlier, our data are provided by <http://www.taobao.com>, a China-based online platform.<sup>3</sup>

<sup>2</sup> Given that our setting is a monopoly platform that does not charge sellers and buyers, we only focus on work related to network effects. For work that has focused on competition, price structure, and market power, see Kaiser and Wright (2006), Chandra and Collard-Wexler (2009), Jin and Rysman (2015), Seamans and Zhu (2014), Argentesi and Filistrucchi (2007), Liu (2010), and Pattabhiramaiah et al. (2013). For work on market outcomes and consumer welfare, see Chen and Xie (2007), Dubé et al. (2010), Fan (2013), and Song (2013).

<sup>3</sup> The data are provided to us under a non-disclosure agreement that allows us to publish analyses and results but not the raw data.



Taobao is the world's largest online C2C platform, both by registered users and by revenues. By December 31, 2012, Taobao had 7.1 million sellers and 435 million buyers. Its transactions in 2012 totaled 590 billion yuan or \$95 billion. Given these numbers, Taobao essentially represents the C2C platform market. We now provide a brief introduction to the platform, its history, organizational structure, and business model as many of our modeling choices are based on these details.

Taobao is a subsidiary of Alibaba (China) Group, a publicly owned company<sup>4</sup> that does business in many areas of e-commerce, including business-to-business (B2B, <http://www.alibaba.com> for international trade, and <http://www.1688.com> for China's domestic trade) for small and midsize enterprises, C2C (<http://www.taobao.com>), business-to-consumer (B2C, <http://www.tmall.com>), online payment system (<http://www.alipay.com>), group buying (<http://www.juhuasuan.com>), cloud computing (<http://www.aliyun.com>), and others.<sup>5</sup>

Taobao began operations in May 2003. The first seller registered on May 11, 2003. In the early years, Taobao's growth was slow. Taobao adopted a "free" policy—free registration and free transactions for buyers, and free registration, free listing, and free transactions for sellers. It created Aliwangwang, a Skype-like communication device that allows buyers and sellers to fully communicate and exchange information to facilitate transactions. It also created Alipay, a PayPal-like escrow payment system that resolved the payment and trust issue for Internet commerce in a country where credit card use was far from universal and buyers and sellers had mutual distrust for each other in online transactions.<sup>6</sup> As a result, Taobao quickly gained market acceptance. As of the end of 2012, Taobao accounted for about 75% of China's online retailing and had over a 95% market share in China's online C2C commerce. It can therefore be considered as having a virtual monopoly in C2C platforms.<sup>7</sup>

<sup>4</sup> The Alibaba Group went public in Hong Kong in 2007, and privatized in 2012. It went public on the New York Stock Exchange on September 19, 2014. It raised \$25 billion, representing the largest initial public offering worldwide to date.

<sup>5</sup> For more information, please visit <http://www.alibabagroup.com/cn/global/home>.

<sup>6</sup> For example, in the Chinese market context, it was very novel that buyers pay *before* seeing the actual goods they buy and sellers deliver the goods *before* receiving payment.

<sup>7</sup> Taobao was launched in May 2003 as part of a defensive action by the Alibaba Group against eBay that, in 2003, was firming up a deal to enter China in collaboration with a Chinese partner, eachnet.com. eBay Eachnet adopted a business model similar to its U.S. counterpart—transactions cleared via an auction process, sellers had to pay registration and listing fees whereas buyers did not pay registration and transacted for free. eBay Eachnet did not employ an

Taobao continues its free policy to date. Specifically, buying at Taobao is free. To register as a buyer, an agent has to provide a valid cell phone number or an email address. Once a person chooses a user name and a password, Taobao sends an activation code or link to the phone number or email account, typically on the same day (or occasionally, the next day). Once activated, a buyer remains registered as a buyer, even if she does not transact. Selling at Taobao is also free—free registration, no membership fee, no annual fee, no listing fee, and no transaction commissions. Registering as a seller at Taobao is via a "real-name authentication" registration process.<sup>8</sup> The seller must be at least 18 years old, hold a valid photo ID, and pass a simple test (mainly on Taobao's rules and regulations). The process of verification and approval takes from two to seven days. This variation in the approval time is a function of the amount of transaction activity, seller registration volume, and unrelated corporate activity.<sup>9</sup>

Taobao also provides a lot of information about its activities on its website. Specifically, it regularly

escrow based system and also forbade buyers and sellers from communicating directly with each other. Because of the lack of localization, eBay Eachnet never enjoyed the success and popularity in China that it did in the United States and Germany and was quickly overtaken by the local upstart, Taobao. In three short years, Taobao had over two-thirds of the Chinese C2C market and eBay exited China, dissolving the partnership in December 2006 (Wang 2012). Therefore, it is clear that Taobao was a virtual monopoly after 2006 but did face some competition in the 2003–2006 period. Because we do not have any data on the number of participants on eBay Eachnet, the outside good described in §§4.1 and 4.2 between 2003–2006 combines not joining any C2C platform (i.e., shopping at physical stores) and joining eBay Eachnet. In §7.2.6, we drop the initial years from our analysis to see if that affects our results.

<sup>8</sup> Details on the seller registration procedure (in Chinese) are at [http://service.taobao.com/help/seller/shop\\_step1\\_01\\_03.html?spm=0.0.0.0.6vM1SF](http://service.taobao.com/help/seller/shop_step1_01_03.html?spm=0.0.0.0.6vM1SF).

<sup>9</sup> Note that as Taobao is a C2C platform, an agent can function as a buyer *and* a seller. If an agent registers as a buyer first, then the agent is counted as a seller only if she applies to be a seller and the application is approved. On the other hand, if an agent applies to be a seller first (and is approved), then the agent is counted as a buyer only when the agent makes the first purchase. The marginal impact of the presence of such agents could be different from the impact of agents who are pure buyers or sellers in the evolution of the network. Taobao estimates that in general, the number of such agents (acting as both buyer and seller) is about 10% of all active sellers. So at the end of 2012, 700,000 of the 7.1 million sellers are also buyers. We were also able to obtain more disaggregate data on three reasonably large product categories—women's shoes, baby care products, and cell phones. The number of sellers who acted as buyers in these three product categories was 0.15%, 0.24%, and 0.17% in women's shoes, baby care products, and cell phones. Given the aggregate nature of our data, we cannot control for this potential difference explicitly. Although this remains a limitation of our data, the relatively small proportion of such agents is unlikely to "contaminate" the average estimated effect in any significant manner.

posts the information on the number of buyers and sellers transacting at Taobao. In addition, it provides details on the total transaction volume. The website also carries a list of all available products, organized as a hierarchy of category, subcategory, etc., all the way to the individual item. Taobao also makes publicly available several of its indices, including (a) the Taobao consumer price index that tracks and publishes the overall price of products sold on Taobao; (b) the Taobao Index that provides information on searches, transactions, and characteristics of buyers at the product category level; and (c) the Taobao Interest Index that tracks searches, bookmarks, and transactions by category, by day, and by week. Other supplementary data on the status of Taobao's platform can also be obtained relatively easily at search engines such as [www.baidu.com](http://www.baidu.com).<sup>10</sup> Thus, sellers and buyers have access to quite a lot of information before they decide to join the platform. Taobao also advertised on TV during the 2003–2005 period to inform consumers about the existence of the platform—unfortunately, we do not have access to these data.

All transactions at Taobao are made via Alipay, which is linked to buyers and sellers' accounts in many banks in China. Using Alipay is free both for Taobao buyers and sellers. After a buyer places an order with a seller and pays Alipay, Alipay notifies the seller of the purchase and asks the seller to fulfill the order. The seller then arranges for logistics and delivery and notifies the buyer of shipping details (shipping date, expected delivery date, tracking information, etc.). Alipay holds the payment for one month or upon buyer's confirmation of delivery. The money paid to Alipay is held in escrow by a Chinese national bank. The funds held by Alipay are not available to Taobao for any use under Chinese regulations.

The free buying and selling policy means that Taobao does not earn money from buyer and seller registration and transactions. Taobao's revenues are based on three sources. The primary source is from seller online advertising expenditure on Taobao.com. The second source is seller participation fees in Taobao's special marketing channels and promotional activities, such as "Taobao Golden Coins," "Everyday Special Prices," "Trial Center," etc. The third source is fee-based shop management tools (such as software) and value-added services for sellers. Taobao estimates that there is the usual 80:20 split across sellers with approximately 20% of the sellers accounting for about 80% of the transactional revenue. Not surprisingly, the majority of Taobao's own revenue comes from this heavy seller group.

<sup>10</sup> A set of annotated screenshots illustrating the availability of this information is available from the authors on request.

## 4. Model

We take a utility-based approach to model the platform's evolution by focusing on the growth of buyers on one side of the market and that of sellers on the other side. We consider a monopoly platform that provides a marketplace for buyers  $B$  and sellers  $S$  to transact with each other. It charges buyers and sellers, respectively,  $P^B$  and  $P^S$  (fixed) membership fees and  $a^B$  and  $a^S$  commissions per transaction. Both membership fees and transaction charges can be zero or negative (subsidies). Note that the key decision in our model is whether to join the platform or not (for both the buyer and seller). In other words, we are not modeling the buyer's decision to buy an item at a given price or a seller's decision to sell an item at a given price. We next derive a buyer's and a seller's probability of joining the platform and the platform's market shares on the buyer and seller sides.

### 4.1. Buyer Side Model

A representative buyer's utility of joining the platform is based on (i) her intrinsic preference for the platform  $b^B$ ; (ii) the number of sellers on the platform or the installed base of sellers at time  $t - \tau$ ,  $N_{t-\tau}^S$ ; (iii) product variety  $V_t$ , which increases the chance of match between buyers and sellers; (iv) the platform's time-varying marketing activities and other facilitators for online shopping such as the advancement of the logistics industry  $Y_t$ ; (v) seasonality and holiday factors  $X_t$ ; (vi) a price index representing the price image of goods sold on the platform at time  $t$ ,  $p_t^B$ ; (vii) the price of joining the platform,  $P^B$ ; (viii) unobserved (to the researcher) factor(s)  $\xi_t^B$ ; and (ix) a buyer idiosyncratic factor  $\varepsilon_{it}^B$ . In addition, a large literature on new product adoption (e.g., Bass 1969, Mahajan et al. 1995) has shown that an individual's adoption decision is influenced (typically by word of mouth) by how many others have adopted the product or the installed base of buyers,  $N_{t-\kappa}^B$ , i.e., there is a direct network effect. The net indirect utility of a representative buyer at time  $t$  is

$$U_{it}^B = f(b^B)f(N_{t-\tau}^S, V_t, Y_t, X_t, p_t^B, \xi_t^B, \varepsilon_{it}^B)f(N_{t-\kappa}^B) - P^B. \quad (1)$$

In our context,  $P^B = 0$ .<sup>11</sup> Since we do not have information on  $Y_t$ , the platform's marketing activities and the logistics industry, we include a linear and a quadratic time trend to capture their effects. Assuming the buyer's net utility from joining the platform

<sup>11</sup> As noted earlier, the actual cost of joining the platform is zero as per Taobao's policy. To ensure that there was no "hassle" or "time" cost of joining the platform, we recruited 100 Chinese individuals and asked them to register on Taobao as buyers. We found that, on average, it took them 1.92 minutes (SD = 0.44 minutes) to do so, suggesting that there are no hassle or time costs incurred in joining the platform.

takes the Cobb–Douglas form (see Berry et al. 1995, Petrin 2002, Rysman 2004), we have buyer  $i$ 's indirect utility of joining the platform as

$$U_{it}^B = \beta_0 + \gamma_1 t + \gamma_2 t^2 + \beta_1 \ln(N_{t-\kappa}^B) + \beta_{2i} \ln(N_{t-\tau}^S) + \beta_3 \ln(p_t^B) + \beta_4 V_t + \beta_5 X_t + \xi_t^B + \varepsilon_{it}^B. \quad (2)$$

In this setup,  $\beta_0 \equiv f(b^B)$  represents buyer's intrinsic preference for the platform, and  $\gamma_1$  and  $\gamma_2$  are the effect of time trend, capturing the influence of all other time-varying variables ( $Y_t$ ) that are not included in the model. The coefficient  $\beta_1$  measures the effect of buyer's installed base on the growth of buyers or the DNE, and  $\beta_{2i}$  measures the effect of seller's installed base on buyer's utility, i.e., the CNE of sellers on buyers. To capture the evolution of CNE over time, we allow this coefficient to be time (year and month) varying— $\bar{t}$  refers to the calendar month in which day  $t$  falls. The coefficient  $\beta_3$  represents the effect of product price index or price image on buyers,  $\beta_4$  is the marginal effect of product variety on buyers, and  $\beta_5$  stands for the effect of seasonality and holidays. Assuming  $\varepsilon_{it}^B$  follows independent and identically distributed (i.i.d.) extreme value distribution and the utility of not joining the platform is normalized to zero, we have the buyer's probabilities of joining and not joining the platform, respectively, as

$$\Pr_{it}^B = \left[ \exp(\beta_0 + \gamma_1 t + \gamma_2 t^2 + \beta_1 \ln(N_{t-\kappa}^B) + \beta_{2i} \ln(N_{t-\tau}^S) + \beta_3 \ln(p_t^B) + \beta_4 V_t + \beta_5 X_t + \xi_t^B) \right] \cdot \left[ 1 + \exp(\beta_0 + \gamma_1 t + \gamma_2 t^2 + \beta_1 \ln(N_{t-\kappa}^B) + \beta_{2i} \ln(N_{t-\tau}^S) + \beta_3 \ln(p_t^B) + \beta_4 V_t + \beta_5 X_t + \xi_t^B) \right]^{-1}, \quad (3)$$

$$\Pr_{it}^{B,0} = \left[ 1 + \exp(\beta_0 + \gamma_1 t + \gamma_2 t^2 + \beta_1 \ln(N_{t-\kappa}^B) + \beta_{2i} \ln(N_{t-\tau}^S) + \beta_3 \ln(p_t^B) + \beta_4 V_t + \beta_5 X_t + \xi_t^B) \right]^{-1}. \quad (4)$$

Under the assumption that buyers “single home” (a reasonable assumption in our empirical context as described above), a buyer's probability of joining the platform is the same as the platform's market share of buyers,  $z_t^B$ .<sup>12</sup> Thus, the platform's relative market share is

$$\ln \frac{z_t^B}{z_t^{B,0}} = \ln \frac{n_t^B / M_t^B}{(M_t^B - n_t^B) / M_t^B} = \beta_0 + \gamma_1 t + \gamma_2 t^2 + \beta_1 \ln(N_{t-\kappa}^B) + \beta_{2i} \ln(N_{t-\tau}^S) + \beta_3 \ln(p_t^B) + \beta_4 V_t + \beta_5 X_t + \xi_t^B, \quad (5)$$

<sup>12</sup> As we discuss in §5 we only have access to aggregate data. We tried to accommodate heterogeneity by allowing the intrinsic preference for the platform to vary across buyers and sellers, but were unable to obtain a meaningful estimate for the heterogeneity term. All of the other estimates were materially unaffected (details are available from the authors on request).

where  $n_t^B$  is the number of new buyers in time period  $t$  and  $M_t^B$  is the market potential for buyers at the beginning of time  $t$ .

#### 4.2. Seller Side Model

We derive the seller's probability of joining the platform and the platform's market share of sellers in a similar manner. A seller's utility of joining the platform depends on (i) her intrinsic preference for the platform  $b^S$ ; (ii) the number of buyers on the platform or the installed base of buyers at time  $t - \kappa$ ,  $N_{t-\kappa}^B$ ; (iii) buyer's quality  $Q_t^B$ , which increases the attractiveness of the platform; (iv) the platform's time-varying marketing activities and other facilitators for online shopping such as the advancement of the logistics industry  $Y_t$ ; (v) seasonality and holiday factors  $X_t$ ; (vi) a price index representing the price image of goods sold on the platform at time  $t$ ,  $p_t^S$ ; (vii) the price of joining the platform,  $P^S$ ; (viii) some unobserved factors  $\xi_t^S$ ; and (ix) seller idiosyncratic factor  $\varepsilon_{jt}^S$ . A seller's decision to join the platform may also be influenced by how many other sellers have joined the platform or the seller's installed base  $N_{t-\tau}^S$ . On one hand, a prospective seller can learn about the prospect of doing business on the platform from existing sellers;<sup>13</sup> on the other hand, the prospective seller may be concerned about potential competition from existing sellers. Thus, the DNE of seller's installed base can be positive or negative, depending on which of the two effects dominates. The net indirect utility of a representative seller  $j$  joining the platform at time  $t$  is

$$U_{jt}^S = f(b_t^S) f(N_{t-\kappa}^B, Q_t^B, Y_t, X_t, p_t^S, \xi_t^S, \varepsilon_{jt}^S) f(N_{t-\tau}^S) - P^S. \quad (6)$$

In our context,  $P^S = 0$ .<sup>14</sup> Assuming the seller's net utility from joining the platform takes the Cobb–Douglas form (see Berry et al. 1995, Petrin 2002, Rysman 2004), we have seller  $j$ 's indirect utility of joining the platform as

$$U_{jt}^S = \alpha_0 + \lambda_1 t + \lambda_2 t^2 + \alpha_1 \ln(N_{t-\kappa}^S) + \alpha_{2i} \ln(N_{t-\kappa}^B) + \alpha_3 \ln(p_t^S) + \alpha_4 Q_t^B + \alpha_5 X_t + \xi_t^S + \varepsilon_{jt}^S. \quad (7)$$

In this utility setup,  $\alpha_0 \equiv f(b^S)$  represents seller's intrinsic preference for the platform, and  $\lambda_1$  and  $\lambda_2$

<sup>13</sup> There are reports in the business press about the so-called “Taobao villages,” where the existence of local sellers encourages other local sellers to join Taobao as well (see <http://www.economist.com/news/china/21602755-one-small-hamlet-teaching-people-how-sell-online-cash-cow-taobao>). We thank an anonymous referee for bringing this phenomenon to our attention.

<sup>14</sup> Although the hassle cost of signing up is slightly higher for sellers, based on the results of the survey carried out by Peking University and Aliresearch as well as in-depth interviews of 10 Taobao sellers (carried out by the authors), it seems that the overall cost (nonfinancial) of joining is considered negligible by sellers.



are the effect of time trend, capturing the influence of all other time-varying variables ( $Y_t$ ) that are not included in the model. The coefficient  $\alpha_1$  measures the effect of seller's installed base or the DNE, and  $\alpha_{2\bar{t}}$  measures the effect of the installed base of buyers on seller's utility, i.e., the CNE of buyers on sellers. To capture the evolution of CNE over time, we allow this coefficient to be time (year and month) varying— $\bar{t}$  refers to the calendar month in which day  $t$  falls. The coefficient  $\alpha_3$  represents the effect of product price image on seller's utility,  $\alpha_4$  denotes the effect of buyer quality, and  $\alpha_5$  stands for the effect of seasonality and holidays. Assuming  $\varepsilon_{jt}^S$  follows i.i.d. extreme value distribution and the utility of not joining the platform is normalized to zero, we have the seller's probabilities of joining and not joining the platform, respectively, as

$$\Pr_{jt}^S = \left[ \exp(\alpha_0 + \lambda_1 t + \lambda_2 t^2 + \alpha_1 \ln(N_{t-\tau}^S) + \alpha_{2\bar{t}} \ln(N_{t-\kappa}^B) + \alpha_3 \ln(p_t^S) + \alpha_4 Q_t^B + \alpha_5 X_t + \xi_t^S) \cdot [1 + \exp(\alpha_0 + \lambda_1 t + \lambda_2 t^2 + \alpha_1 \ln(N_{t-\tau}^S) + \alpha_{2\bar{t}} \ln(N_{t-\kappa}^B) + \alpha_3 \ln(p_t^S) + \alpha_4 Q_t^B + \alpha_5 X_t + \xi_t^S)]^{-1}, \quad (8)$$

$$\Pr_{jt}^{S,0} = [1 + \exp(\alpha_0 + \lambda_1 t + \lambda_2 t^2 + \alpha_1 \ln(N_{t-\tau}^S) + \alpha_{2\bar{t}} \ln(N_{t-\kappa}^B) + \alpha_3 \ln(p_t^S) + \alpha_4 Q_t^B + \alpha_5 X_t + \xi_t^S)]^{-1}. \quad (9)$$

Under the assumption that sellers “single home,” the seller's probability of joining the platform is the same as the platform's market share of sellers,  $Z_t^S$ . Thus, the platform's relative market share is

$$\ln \frac{Z_t^S}{Z_t^{S,0}} = \ln \frac{n_t^S / M_t^S}{(M_t^S - n_t^S) / M_t^S} = \alpha_0 + \lambda_1 t + \lambda_2 t^2 + \alpha_1 \ln(N_{t-\tau}^S) + \alpha_{2\bar{t}} \ln(N_{t-\kappa}^B) + \alpha_3 \ln(p_t^S) + \alpha_4 Q_t^B + \beta_5 X_t + \xi_t^S, \quad (10)$$

where  $n_t^S$  is the number of new sellers in time period  $t$  and  $M_t^S$  is the market potential for sellers at the beginning of time  $t$ . We now have the system of Equations (5) and (10) that can be taken to the data for estimation. We collect the notation into Table 1 for ease of exposition.

## 5. Data and Variable Operationalization

As noted earlier, our data are novel, especially in the sense that we have data from Taobao's inception. Specifically, we have daily observations from May 11, 2003, the day when the first seller registered on Taobao, to December 31, 2012. For each day,

**Table 1** Notation

Notation	Definition
$B$	Buyer
$S$	Seller
$t$	Time (day)
$n_t^B$	New registered buyers during time $t$
$n_t^S$	New registered sellers during time $t$
$N_t^B$	Total number (the installed base) of buyers at the beginning of $t$
$N_t^S$	Total number (the installed base) of sellers at the beginning of $t$
$M_t^B$	Potential market size for buyers at the beginning of time $t$
$M_t^S$	Potential market size for sellers at the beginning of time $t$
$U_{it}^B$	Buyer's utility of joining the platform
$U_{it}^S$	Seller's utility of joining the platform
$P_{it}^B$	Buyer's probability of joining the platform
$P_{it}^S$	Seller's probability of joining the platform
$P_t^B$	Buyer's price of joining the platform
$P_t^S$	Seller's price of joining the platform
$p_t^B$	Price index for buyers
$p_t^S$	Price index for sellers
$Z_t^B$	The platform's market share of buyers
$Z_t^S$	The platform's market share of sellers
$V_t$	Product variety index
$Y_t$	Time-varying factors such as the platform's marketing activities and other facilitators for online shopping (e.g., the advancement of the logistics industry)
$Q_t^B$	Buyer quality
$X_t$	Seasonality and holiday factors
$CF(r_t^B)$	Control function for buyer installed base
$CF(r_t^S)$	Control function for seller installed base
$\xi_t^B$	Unobservable buyer factors
$\xi_t^S$	Unobservable seller factors
$\varepsilon_{it}^B$	Buyer idiosyncratic factor
$\varepsilon_{it}^S$	Seller idiosyncratic factor
$\sigma_t^B$	Standard deviation for the buyer's equation error in the bivariate normal distribution
$\sigma_t^S$	Standard deviation for the seller's equation error in the bivariate normal distribution
$\rho$	Correlation coefficient for the bivariate normal distribution

we observe the number of new buyers, new sellers, transacting buyers, transacting sellers, transactions, unique items sold, total items sold, mean transaction prices, expenditures per buyer, expenditure per transaction, and total revenues. These variables are aggregated across all products. At the product category level (Taobao defines its own product categories), we observe numbers of new items added, total number of items on shelf, mean item price, and total transactions for each product category.<sup>15</sup> Unfortunately, we do not

<sup>15</sup> Taobao also allows buyers to rate sellers on a five “star” scale and reports the percentage of good ratings. It is quite possible that buyers decide to join Taobao based on average seller ratings across the platform. We approached the company about getting data on ratings. The company did not provide us the ratings for the following four reasons. First, Taobao executives told us that, during this period, given the large number of transactions, only about 40% of them had actual ratings by buyers. For the remaining 60%, Taobao would assign them five stars (the maximum) as the default rating. Second, the average rating across all sellers on a daily basis



**Table 2** Daily New Buyers and New Sellers

	New sellers (1,000)				New buyers (1,000)			
	Mean	Std dev	Sum	Cum sum	Mean	Std dev	Sum	Cum sum
2003 <sup>a</sup>	0.01	0.01	3	3	0.00	0.01	1	1
2004	0.12	0.08	39	42	1.42	1.76	473	473
2005	0.78	0.52	286	328	24.19	12.67	8,831	9,304
2006	2.07	0.51	754	1,082	43.92	6.32	16,031	25,335
2007	2.92	0.73	1,064	2,147	56.82	13.44	20,739	46,075
2008	4.66	1.20	1,707	3,854	115.75	31.20	42,364	88,439
2009	7.68	2.02	2,805	6,659	156.11	47.24	56,981	145,419
2010	10.54	3.43	3,848	10,507	197.30	45.09	72,014	217,434
2011	15.85	4.47	5,785	16,292	236.47	57.65	86,312	303,746
2012	13.86	3.30	5,072	21,364	359.46	91.74	131,562	435,308
All years	6.15	5.96	21,364		125.38	119.49	435,308	

<sup>a</sup>New sellers start from May 11, 2003, and new buyers from October 15, 2003.

have numbers of sellers or buyers for each product category.

### 5.1. Data Summary

Table 2 summarizes daily new sellers and new buyers, their annual totals and growth; Figure 1(a) plots the evolution of daily registrations over time. There are huge variations in daily registrations. During October to December 2003, the average number of new sellers and buyers on a day was 15 and 3, respectively. Daily new sellers reached three digits and daily new buyers reached four digits in 2004. The platform really started to take off in 2007—nearly 3,000 sellers and 57,000 buyers registered each day, and over one million sellers and 20 million buyers registered in that year. The seller installed base reached two million and the buyer installed base exceeded 46 million. Since then, both buyer and seller numbers continued to grow. In 2012, there were 14,000 new sellers and 360,000 new buyers added to the platform each day. By the end of 2012, the installed base was 21 million (sellers) and 435 million (buyers). Note that this installed base here does not account for seller and buyer attrition (we discuss this in detail in §5.2.1).

Over time, the total number of transactions per day has gone from 2,000 per day in 2004 to 13 million in 2012. Table 3 reports percentages of sellers and buyers with transactions over total sellers and buyers as well

as total transactions per day. The share of sellers with a transaction has remained stable in the last three or four years at around 5% (around 11% once we account for seller attrition). On the other hand, the share of buyers making purchases has been rising slowly since 2006, culminating at about 1.4 out of 100 registered buyers making a purchase in end 2012.

Table 4 shows some characteristics of daily transactions, including mean item price, size of each transaction, and revenues. The daily transaction revenue has been increasing rapidly and reached 1.61 billion yuan (USD \$258 million) in 2012. The average item price stabilized to around 13 yuan (USD \$2.09) by 2006 after some initial fluctuation. The value of each transaction<sup>16</sup> has also stabilized to around 125 yuan (USD \$20) with the expenditure per buyer being around 325 yuan (USD \$52.17).

### 5.2. Variable Operationalization

Because of the nature of the research methodology, data, and institutional setting, we need to construct many of the variables that we use. We discuss these below. In §6.3, we explore the robustness of our results to alternative operationalization of these variables.

**5.2.1. Buyer and Seller Installed Base.** We use the cumulative sum of registered buyers each day as the installed base of buyers, and that of registered sellers each day as the installed base of sellers.<sup>17</sup> As noted

did not have much variation (e.g., using a supplementary data set, we found that the mean weekly percentage of good ratings for a million sellers over an 18 month period beginning November 2011 is 80.9% with a standard deviation of 0.9%). Third, Taobao had noticed that some buyers were using the Aliwangwang communication tool to “intimidate” sellers into giving them better deals in return for more favorable ratings, i.e., these ratings were not truly reflective of seller quality. Fourth, at a seller’s request and provision of evidence that a certain rating is a result of buyer’s (failed) intimidation attempt, Taobao can revise or erase bad ratings. These reasons suggest that aggregate ratings data are not likely to be diagnostic.

<sup>16</sup> This includes shipping fees that range from 1% to 15% of transaction size depending on product category. Generally, the smaller the total basket value in yuan, the higher the percentage shipping fee.

<sup>17</sup> Given the modest transaction size, it is possible that transactions on Taobao skew local, i.e., buyers tend to buy from local sellers. In that case, both parties would care about the local installed base rather than the national installed base. We were able to obtain some supplemental data from Taobao.com vis-à-vis this issue. For the women’s shoe product category, across China’s 31 provinces, the average percentage of buyers *outside of a seller’s province* is 92.01%

Downloaded from informs.org by [130.115.158.36] on 10 October 2017, at 01:03 . For personal use only, all rights reserved.



**Table 3** Summary of Daily Transacting Buyers and Sellers

	Transacting sellers/ Total sellers (%)		Transacting sellers/ Discounted total sellers (%)		Transacting buyers/ Total buyers (%)		No. of transactions ('000)		No. of transactions per 100 buyers	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
2003	0.42	0.31	0.52	0.39	10.47	15.04	0.01	0.01	12.51	18.00
2004	4.99	4.11	6.34	5.09	1.93	0.97	2.07	2.68	2.56	1.29
2005	11.43	1.78	14.85	2.35	0.76	0.20	42.18	29.34	1.27	0.29
2006	8.36	1.58	11.82	1.97	0.62	0.11	201.16	77.18	1.15	0.21
2007	7.09	0.98	12.16	1.74	0.80	0.15	560.65	194.88	1.59	0.33
2008	6.86	0.98	13.33	1.90	0.98	0.16	1,480.11	505.55	2.24	0.45
2009	6.52	1.02	13.55	2.05	1.05	0.19	3,480.62	789.10	3.00	0.62
2010	5.81	0.80	13.35	1.88	1.11	0.20	5,685.80	1,676.95	3.13	0.67
2011	4.79	0.78	11.14	1.76	1.27	0.29	8,860.72	2,549.62	3.44	0.84
2012	4.66	0.74	13.01	2.40	1.42	0.33	13,015.08	4,013.32	3.54	0.92
All	6.63	2.75	12.02	3.68	1.26	2.40	3,674.00	4,682.64	2.61	2.94

**Table 4** Summary of Daily Transactions

Year	Mean item price (yuan)		Expenditure per transaction (yuan)		Expenditure per buyer (yuan)		Daily revenues (million yuan)	
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
2003	102.54	265.44	110.73	290.90	121.14	307.07	0.00	0.00
2004	160.17	140.90	232.91	129.10	292.33	134.37	0.36	0.50
2005	30.11	10.34	166.91	27.53	279.84	36.79	6.61	4.32
2006	13.69	3.11	144.61	12.17	265.42	23.98	29.38	12.40
2007	11.54	1.66	162.52	10.71	319.48	24.45	91.72	33.82
2008	13.99	1.99	142.52	20.43	319.63	28.81	203.71	51.78
2009	12.23	1.52	117.48	17.94	329.22	31.45	404.68	103.27
2010	12.62	1.46	122.21	16.01	339.07	26.12	684.99	205.94
2011	13.82	1.81	125.00	13.52	335.27	31.40	1,104.46	339.76
2012	14.74	1.64	125.13	16.44	307.40	27.95	1,612.17	512.77
All years	34.7	91.8	145.68	90.65	298.17	104.47	435.54	574.67

earlier, Taobao's policy is that once a buyer activates her account, the buyer remains a buyer, regardless of transaction activity. Unlike buyers however, Taobao has data on whether a seller is present and active on the platform. Sellers exit either voluntarily from the platform (typically for business reasons, e.g., they are not profitable) or involuntarily (usually because they violate Taobao rules and regulations and the platform shuts them down). By December 31, 2012, the total number of sellers ever registered exceeded 21 million, but the total number of sellers in normal state (defined as transacting and/or engaging in

merchandising activity at least once a quarter) was only 7.1 million, i.e., about one-third the cumulative sum of registered sellers. We therefore need to adjust the cumulative sum of registered sellers to be consistent with the number of sellers in the normal state.<sup>18</sup> Based on our discussion with the company, we assumed that sellers drop out in a manner consistent with an exponential decay. Specifically, if there are  $n_{t-\tau}^S$  sellers registered at  $t - \tau$ , by time  $t$ , there will be  $n_{t-\tau}^S / (1 + r^S)^\tau$  sellers left where  $r^S$  is the decay parameter, and the resulting number is termed "discounted number of sellers," and their cumulative sum is termed "total discounted sellers." To estimate this parameter, we equate the adjusted number of sellers (using this parameter) with the actual number of normal state sellers on December 31, 2012. The best-fit value for  $r^S$  is 0.0018, i.e., every day 1.8 out of 1,000 sellers drop out (we test the robustness of the model estimates to this adjustment in §7.2.3). Figure 1(b)

(with a range of 29.26% to 100%). Taobao also reported to us that for the cell phone category, buying is nationwide (following the population distribution) whereas selling is concentrated with 80% of sellers based in Guangdong. This suggests that agent utility is based on the national installed base, not a local one. In fact, feedback from the company's surveys suggests that sellers wanted to go online at Taobao because it gave them access to a national market of buyers (as opposed to a local market for a physical store)—virtually no sellers on Taobao maintain a physical store. Buyers on the other hand went on Taobao to get the best prices from sellers nationwide.

<sup>18</sup> The company was unable to provide us an exact count of the number of normal sellers on each day because of the cost involved in extracting these data.

shows the buyer installed base and seller installed base (with and without the adjustment).

**5.2.2. Product Variety Index.** We first compute the platform's category concentration in the number of product items (equivalent to stock keeping unit). Analogous to the industry concentration Herfindahl–Hirschman Index (HHI), category  $c$ 's share in the number of items is  $S_c^i = I_c^i / I_i$ , where  $I_c^i$  is the number of items in category  $c$  and  $I_i$  is the number of items across all product categories. The category's item concentration HHI is calculated as

$$HHI_i = \sum_{c=1}^C (S_c^i)^2.$$

Product variety index is defined as  $V_i = 1 - HHI_i$ . The product variety index,  $V_i$ , lies between  $[0, 1]$ . When all items are concentrated in one category,  $V_i = 0$ , and when all items are evenly distributed across categories,  $V_i = 1 - 1/C$ . The product variety index,  $V_i$ , approaches 1 as the number of categories increases. A similar index is used to measure variety in other studies (e.g., Fan 2013). Product variety at Taobao has been increasing. It fluctuates substantially in the beginning years and gradually stabilizes at a high level.

**5.2.3. Buyer Quality.** We define buyer quality as the number of transactions per 100 buyers in the installed base, calculated by dividing the number of transactions each day by the installed base of buyers ( $\times 100$ ). We test the sensitivity of model parameter estimates to other measures of buyer quality. The right-most columns of Table 3 report the average daily buyer quality and its standard deviation for each year. Most Taobao buyers are not active. On an average day, there are only 2.5 transactions per 100 buyers. Even during the peak promotion days such as "Double 11" (November 11) and "Double 12" (December 12) promotions, the number of transactions is still less than 10 per 100 buyers. However, buyer quality has been gradually improving over the years.

**5.2.4. Product Price.** We take a representative consumer approach in the model setup. We observe the average transaction prices across all items for each day. For sellers, we can use this price because sellers are assumed to be more informed of product prices.<sup>19</sup> However, using average prices for a representative

buyer is equivalent to requiring her to know prices in each product category, which is too strong an assumption. Instead, we construct a price index using a fixed basket. We compute basket shares based on the total revenues of the 50 categories that are sold throughout the entire period, and compute price index using the basket shares.<sup>20</sup> We check the robustness on the number of categories included, and on fixed basket versus time-varying basket. Over the years, average basket prices have been declining to within 200 yuan (USD \$32), corroborating Taobao's low-price image.

**5.2.5. Seasonal, Promotional Event, and Holiday Controls.** Taobao started Double 11 and Double 12 promotions in 2010. Ever since, these two days have become the biggest promotional activities for the platform and for its buyers and sellers as well. We create dummies for these two days. Many households in China may not have computers with Internet access at home. They often surf the Web from offices. They may have more engagements on weekends, such as visiting friends, shopping at physical stores, or simply relaxing. We create a dummy for each day of the week to account for these effects. Many holidays in China run over several days, some for even a week or more. Logistics companies nearly stop operations during these holidays, particularly during the Chinese Lunar New Year, which greatly hinders online shopping behavior. We create dummies for all Chinese holidays to capture these effects.

## 6. Estimation

We use maximum likelihood to estimate the model parameters. In the estimation, we need to address two issues, one is how the market for buyers and sellers evolve over time, and the other is how to resolve the potential simultaneity and endogeneity of the buyer's and seller's installed bases.

### 6.1. Potential Market Sizes for Buyers and Sellers

We allow buyers and sellers to have the option of not joining the platform. We use the number of Internet users in China as the base of the potential market for buyers and scale it by 1.3 because an average buyer has 1.3 accounts at Taobao.<sup>21</sup> Buyer's market size evolves as follows: At the beginning of time  $t$ ,

<sup>20</sup> On July 5, 2008, Taobao started to publish CPI based on product prices and sales on its website. However, this index is not available to us. Our method of computing the price index is similar to how Taobao computes its CPI.

<sup>21</sup> The data on the number of Internet users in China is obtained from the China Internet Network Information Center (CNNIC). Internet use in China has grown rapidly over the last decade. In June 2003, there were 68 million Internet users with a penetration rate of 5.6%. By December 2012, there were 564 million Internet users with a penetration rate of 42.1%.

<sup>19</sup> We do not adjust prices for inflation in our main model. We did run a model with adjusted prices and the correlation between the reported results and the one with the inflation adjusted prices is 0.99. Relative to the initial period, i.e., over a 10-year period, the consumer price index (CPI) went up 37%. The monthly (and hence daily movement) in CPI is therefore relatively very small, resulting in it not having a meaningful impact on the results (which are available from the authors on request).



there are  $M_t^B$  buyers. During time period  $t$ ,  $n_t^B$  buyers join the platform and drop out of the market, and there are  $m_t^B$  new Internet users joining the potential market. At the end of time  $t$  (or beginning of time  $t+1$ ), the market size is  $M_{t+1}^B = M_t^B - n_t^B + m_t^B = M_t^B(1 - z_t^B)$ .

The great majority of Taobao sellers are individual entrepreneurs, and it is quite common for a husband and wife to start a Taobao business (an online equivalent of the mom-and-pop store), either full-time or part-time. Therefore, we use the number of households in China as the base of the potential market for sellers. Based on internal surveys carried out by Taobao, we scale this by 0.1, with the assumption being that the potential number of sellers for Taobao is likely to be a maximum of 10% of households (to check for robustness, we vary this proportion and also look at an alternative market size for the number of sellers in §7.2.5).<sup>22</sup> From May 2003 to December 2012, households in China increased from 374 million to 448 million. The seller market size evolves as follows: At the beginning of time  $t$ , there are  $M_t^S$  sellers. During time  $t$ ,  $n_t^S$  sellers join the platform and drop out of the market, and there are  $m_t^S$  households joining the potential market. At the end of time  $t$  (or beginning of time  $t+1$ ), the market size becomes  $M_{t+1}^S = M_t^S - n_t^S + m_t^S = M_t^S(1 - z_t^S)$ .

## 6.2. Temporal Lags Vis-à-Vis Buyer and Seller Installed Bases

A key aspect of our model is the installed base of both buyers and sellers affect the joining decision of a prospective buyer and seller. As described in §3, buyers can make purchases immediately after registration approval (typically on the same day or occasionally the next day). Thus both buyers and sellers can see the buyer installed base contemporaneously or at worst, with a one day lag. In our main specification, we set this lag to zero, i.e., in  $N_{t-\kappa}^B$ ,  $\kappa = 0$ . On the other hand, after registration, seller approval takes anywhere between two to seven days (see §3). Thus the buyers and sellers can see the seller installed base after a lag of two to seven days. In our main specification, we set this lag to the modal value of four days, i.e., in  $N_{t-\tau}^S$ ,  $\tau = 4$  (we also test the robustness to other choices of both lags in §7.2.4).

## 6.3. Identification

The main parameters of interest, the monthly CNEs, are identified from the monthly variation in the installed base of buyers and sellers each month (after controlling for the time trends). The DNE in the buyer's (seller's) model is identified from the variation in the installed base of buyers (sellers) across the

entire data period (recall that our main model specification has a time invariant DNE). This was also confirmed via simulation studies. In addition, we address three other broad concerns that are typically raised for simultaneous equation models with respect to identification of model parameters.<sup>23</sup> These are simultaneity, omitted variables, and common shocks. We discuss how we handle each of these in §§6.3.1–6.3.3.

**6.3.1. Simultaneity.** In a typical simultaneous system of equation approach where the actions of one agent affect the actions of the other, there is a possibility of a simultaneity confound, where it is not clear which agent's behavior causally affects the other because both act simultaneously. A typical solution to this confound is the use of excluded variables—these are variables that affect the utility of one agent but not the other, and vice versa. In our setting, the joining decision of buyer (seller) is not a function of the joining decision of a seller (buyer), only of the installed base of sellers and buyers. Thus, the potential for a simultaneity confound is limited. However, we do use excluded variables to control for this issue. Specifically, in our case, buyer quality affects a seller's utility of joining the platform, but not buyer's utility of joining, because buyer quality directly affects seller performance and there is no reason for a prospective buyer to care about buyer quality of the platform (recall buyer quantity is controlled for using the installed base). On the other hand, product variety across the entire platform affects the buyer's, but not the seller's, propensity to join the platform because it increases the chance of product match for buyers. Previous studies (e.g., Boatwright and Nunes 2001, Briesch et al. 2009, Sun et al. 2015) have found that product variety affects buyer's purchase, store choice, or platform choice behavior. Finally, the price image of the platform is constructed differently for buyers and sellers (see §5.2.4), thus representing another set of excluded variables.

**6.3.2. Omitted Variables.** It is indeed possible that our model does not capture all of the variables that drive the buyers and sellers joining utility. There is potential of bias in our estimates if these omitted variables are correlated with our variables of interest (the buyer and seller installed bases) and the error

<sup>22</sup> Data on the number of households over time in China is from the State Statistics Bureau of China.

<sup>23</sup> We also investigated the time-series features of the data to ensure that our results were not driven by "spurious regression." First, although we found modest autocorrelation (even though the Durbin-Watson statistic suggests that we cannot reject the null hypothesis of a random walk), simulation results showed that our parameter recovery was not affected. Second, we were able to replicate our main results using an AR(1) specification (though we could not include DNE in the specification because of multicollinearity). These results are available, on request, from the authors. We thank an anonymous reviewer for suggesting these tests.

term, leading to the classic endogeneity problem. The typical solution is to use instrumental variables—variables that affect the endogenous regressor but not the error term.

We use the following instrumental variables. Specifically, for the buyer installed base, we use national level consumer sentiment indices—the consumer expectation index, the consumer confidence index, and the consumer satisfaction index—as instruments. The intuition for using these is that consumer sentiment is likely to have a material impact on the consumption decisions both online and offline and therefore will affect the probability of engaging in consumption, including via e-commerce, leading to an impact in the buyer installed base. These three monthly indices are compiled by the State Statistics Bureau of China, and jointly explain 58% of variation in the buyer installed base (this is the incremental  $R^2$  as defined in Rossi 2014).

For the seller installed base, we focus on the motivational and operational attributes that drive agents to engage in selling on Taobao. We use the entrepreneur confidence index, compiled by the State Statistics Bureau of China, and the component indices of China's purchasing managers' index (PMI), including new orders index, inventory index, and suppliers' delivery time (to vendors) index, as instruments. These latter indices measure the difficulty, speediness, and costs for sellers to obtain goods for online sales. These data are obtained from the China Federation of Logistics and Purchasing (<http://www.chinawuli.com.cn>) and the Hong Kong Shanghai Banking Corporation (<http://www.hsbc.com/news-and-insight/emerging-markets>). Taken together, these variables jointly explain 36% of the variation (incremental  $R^2$ ) in the seller installed base.

Note that since Taobao buyers and sellers essentially come from the same population, it is a challenging task to find unique instrumental variables that affect one, but not the other side of the platform. We checked the correlation between the instrumental variables for the buyer and seller installed bases and find that the correlations are typically low (the mean absolute correlation is 0.12 and the median is 0.11).

Using the excluded and the instrumental variables, we take the control function approach (Petrin and Train 2010) to address the potential simultaneity/endogeneity problem. Specifically, we regress the buyer installed base on its instruments and other exogenous variables and compute the regression residuals  $r_t^B$ ; we regress the seller installed base on its instruments and other exogenous variables and compute regression residuals  $r_t^S$ . We then put functions of the residuals back into the relative market share equations (Equations (5) and (10)), as shown in Equations (11) and (12), and reestimate the model parameters. The control function includes both the linear and

the quadratic term of the residuals (the results are also robust to other forms of the control functions)

$$\begin{aligned}\ln \frac{z_t^B}{z_{t,0}^B} &= \ln \frac{n_t^B/M_t^B}{(M_t^B - n_t^B)/M_t^B} \\ &= \beta_0 + \gamma_1 t + \gamma_2 t^2 + \beta_1 \ln(N_{t-\kappa}^B) + \beta_{2i} \ln(N_{t-\tau}^S) \\ &\quad + \beta_3 \ln(p_t^B) + \beta_4 V_t + \beta_5 X_t + \beta_6 CF(r_t^S) + \xi_t^B, \quad (11)\end{aligned}$$

$$\begin{aligned}\ln \frac{z_t^S}{z_{t,0}^S} &= \ln \frac{n_t^S/M_t^S}{(M_t^S - n_t^S)/M_t^S} \\ &= \alpha_0 + \lambda_1 t + \lambda_2 t^2 + \alpha_1 \ln(N_{t-\tau}^S) + \alpha_{2i} \ln(N_{t-\kappa}^B) \\ &\quad + \alpha_3 \ln(p_t^S) + \alpha_4 Q_t^B + \alpha_5 X_t + \alpha_6 CF(r_t^B) + \xi_t^S. \quad (12)\end{aligned}$$

**6.3.3. Common Shocks.** There is also a potential for common shocks to affect both buyers and sellers. We control for these by allowing the econometric error terms  $\xi_t^B$  in Equation (11) and  $\xi_t^S$  in Equation (12) to be correlated. We assume they follow bivariate normal distribution as in Equation (13) and estimate model parameters jointly via the maximum likelihood

$$\begin{pmatrix} \xi_t^B \\ \xi_t^S \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^{B2} & \rho \sigma^B \sigma^S \\ \rho \sigma^B \sigma^S & \sigma^{S2} \end{bmatrix} \right). \quad (13)$$

#### 6.4. Measurement of Cross-Network Effect and Nonnetwork Effect

Following the literature (e.g., Gandal et al. 2000), we use elasticities to measure CNE and DNE. We compute the impact on the number of new buyers (sellers) when seller's (buyer's) installed base increases by 1%. The equations to compute cross-network elasticities are as follows:

$$\begin{aligned}e_{S2B} &= \beta_{2i}(1 - Z_t^B), \\ e_{B2S} &= \alpha_{2i}(1 - z_t^S).\end{aligned} \quad (14)$$

The equations to compute direct network elasticities are as follows:

$$\begin{aligned}e_{B2B} &= \beta_1(1 - Z_t^B), \\ e_{B2S} &= \alpha_1(1 - z_t^S).\end{aligned} \quad (15)$$

The equations to compute the elasticities of product price are as follows:

$$\begin{aligned}e_{P2B} &= \beta_3(1 - Z_t^B), \\ e_{P2S} &= \alpha_3(1 - z_t^S).\end{aligned} \quad (16)$$

The equation to compute elasticities for the effect of product variety index on buyers is

$$e_{V2B} = \beta_4 V_t(1 - Z_t^B). \quad (17)$$

The equation to compute elasticities for the effect of buyer quality on sellers is

$$e_{Q2S} = \alpha_4 Q_t^B(1 - z_t^S). \quad (18)$$

Table 5 Main Parameter Estimates

	OLS				IV approach			
	Buyer's model		Seller's model		Buyer's model		Seller's model	
	Est.	Std error	Est.	Std error	Est.	Std error	Est.	Std error
Intercept	−29.893	2.168	−16.445	1.492	−29.890	2.233	−16.054	1.632
$t/100$	−0.126	0.171	0.146	0.123	−0.124	0.175	0.176	0.131
$(t/100)^2$	−0.001	0.003	−0.005	0.003	−0.001	0.003	−0.005	0.003
DNE	0.127	0.053	−0.121	0.106	0.125	0.063	−0.167	0.240
Price	0.006	0.016	0.084	0.012	0.003	0.018	0.048	0.016
Product variety	1.436	0.060			1.439	0.061		
Buyer quality $\times$ 100			0.026	0.004			0.025	0.009
New year	−0.005	0.036	−0.149	0.032	−0.005	0.037	−0.148	0.030
Lunar new year	−0.316	0.022	−0.673	0.019	−0.315	0.023	−0.690	0.024
Labor day	−0.127	0.030	−0.337	0.026	−0.127	0.030	−0.356	0.026
National day	−0.125	0.025	−0.380	0.022	−0.126	0.025	−0.376	0.021
Chingming	−0.026	0.047	−0.098	0.042	−0.026	0.048	−0.082	0.039
Dragon boat	−0.066	0.047	−0.112	0.041	−0.065	0.048	−0.120	0.039
Mid-autumn	0.080	0.047	−0.198	0.042	0.080	0.048	−0.205	0.039
Double 11	0.545	0.100	−0.171	0.089	0.545	0.103	−0.188	0.090
Double 12	0.140	0.101	−0.028	0.089	0.140	0.103	−0.017	0.097
Monday	0.070	0.011	0.099	0.010	0.070	0.011	0.100	0.010
Tuesday	0.050	0.012	0.144	0.010	0.050	0.011	0.142	0.010
Wednesday	0.047	0.012	0.144	0.010	0.047	0.012	0.143	0.010
Thursday	0.036	0.012	0.140	0.010	0.036	0.012	0.139	0.010
Friday	0.047	0.012	0.096	0.009	0.047	0.012	0.095	0.009
Saturday	0.013	0.012	0.022	0.009	0.013	0.012	0.021	0.009
CF					0.003	0.118	0.205	0.048
CF <sup>2</sup>					−0.025	0.177	0.136	0.015
$\sigma^B$	0.200	0.013			0.200	0.013		
$\sigma^S$	0.153	0.012			0.150	0.012		
$\rho$	0.185	0.029			0.180	0.028		
LL		−2,207.563				−2,253.701		
Durbin–Watson statistic	1.218		1.010		1.261		1.032	

## 7. Results

We estimate the models both without (ordinary least squares (OLS)) and with the instrumental variables (two-stage least squares (2SLS)) and find the parameter estimates are only slightly different (Table 5). Our discussion focuses on the 2SLS results. Note that we report heteroscedasticity consistent standard errors (White 1980). In this section, we first present the main parameter estimates, report multiple robustness checks, and then explore the implications of the overall results for managers.

### 7.1. Parameter Estimates

**7.1.1. Cross-Network Effects.** Even though the platform was open for transactions in May 2003, there were very few transactions until November 2003. We therefore use data from November 2003 to December 2012 for our estimation (this also allows for the use of lagged variables without the initial condition problem). To capture the evolution of CNEs over time, we interact the installed base with year and month dummies (November 2003–December 2012). Thus, we have 110 for the seller installed base in the buyer's

equation and 110 coefficients for the buyer installed base in the seller's equation. All of the cross-network coefficients are statistically significant (the mean  $t$ -statistic for the buyer equation is 5.01 with a standard deviation of 0.32 and that of the seller equation is 5.41 with a standard deviation of 0.30). The evolution of the CNEs is shown in Figure 2(a). We now discuss four aspects of these results in detail.

First, there exists a large, significant, and positive CNE on both sides of the platform, indicating that the installed base of either side of the platform has propelled the growth of the other side. Specifically, we find that when the installed base of sellers increases by 1%, the number of new buyers will on average increase by 1.53% (SD = 0.08%, min = 1.45%, max = 1.72%); when the installed base of buyers increases by 1%, the number of new sellers will on average increase by 0.44% (SD = 0.05%, min = 0.31%, max = 0.53%). Our finding of significant positive CNEs on the C2C online platform is analogous to the findings in other settings such as yellow pages (Rysman 2004) and magazines (Kaiser and Wright 2006, Song 2013). However, the magnitudes of these effects are much larger in the online C2C platform than in these other

settings. The mutually enhancing CNEs imply that in the introduction stage, the platform needs to take necessary measures, e.g., free or subsidized pricing, to encourage the growth of the installed base. They also imply that once the installed bases become large enough, the positive externalities will accelerate the growth on both sides without too much intervention from the platform.

Second, the CNE is asymmetric. The seller installed base has a much larger impact on buyer growth than vice versa. This suggests that the platform is much more seller driven than vice versa, especially in the early stages. In Figure 2(b), we plot the ratio of seller's cross-network externality on buyers over buyer's cross-network externality on sellers. On average, seller's CNE is 3.56 times ( $SD = 0.53$ ) as large as buyer's CNE, ranging from 2.84 to 5.26. This ratio declined over time, at a faster speed in the initial two years, which was primarily driven by the decreasing seller's CNE. The ratio started to stabilize around 3.0 from 2010. This asymmetry in the CNE implies that a more preferential policy for the side (sellers) with a larger CNE will be more effective for the platform's growth than the other way around (Rochet and Tirole 2003, Armstrong 2006).

Third, the buyer's CNE on sellers is relatively stable over time. On the other hand, the seller's CNE on buyers, first increases (2003–2004) and then decreases (from mid-2005). It becomes stable after 2010. Thus, it appears that in the introduction phase, the platform's growth is primarily seller driven: seller growth induces buyers to register, which in turn leads to more sellers to register, which further encourages more buyers to register, etc. In the growth stage, the seller's CNE is declining, but it is still well above the buyer's CNE.

Finally, the significance and magnitude of CNE suggests that the installed base of sellers matters to potential buyers even after we control for the general level of price on the platform and product variety available to buyers. Our reasoning as to why it matters goes like this. In general, the magnitude of the installed base of sellers can impact the buyer joining decision directly or indirectly. The direct impact comes from nonmeasurable (to the buyer) attributes, e.g., the size of the installed base could provide the buyer with confidence with respect to carrying out transactions on the platform or the buyers may get consumption utility from knowing that they are shopping at the largest shopping platform in China (and indeed in the world). The indirect impact comes from measurable attributes such as assortment, price, service quality, store layout, delivery cost, shipping speed, etc. In our case, we have data on some of these attributes (e.g., price, assortment) but not on others (e.g., service quality, store layout, shipping speed). In addition, even for the attributes that we have

measures on, they are not perfect in the sense that the measures are constructed. For example, although we use a price index, buyers may be looking at other transformations of price to make their buying decision. Thus, the installed base variable represents both the direct impact as well as the missing and imperfectly measured parts of the indirect impact. Similar arguments can be made for the impact of the buyer installed base on the seller joining decision.<sup>24</sup>

**7.1.2. Direct Network Effects.** There exists a positive and significant, albeit a small DNE on the buyer side, implying that a buyer's decision to join the platform is also influenced by others who have joined the platform. One possible reason for this is word of mouth, wherein a potential buyer may learn about the Taobao website and shopping at Taobao from other Taobao buyers. However, at 0.12%, the DNE is about one order of magnitude smaller than the magnitude of the CNE. This is not surprising because the value of joining to an individual buyer is obtained more from the complementary side of the platform.

The DNE of seller's installed base on seller's growth is negative, but not statistically significant. This means that although there might be some competition effect when sellers decide to join the platform, it has not yet become a barrier to entry. This is probably due to the explosive growth of the platform over its first decade.

**7.1.3. Nonnetwork Factors.** In addition to cross-network externalities and direct network effects, non-network factors such as product variety, product price, and buyer quality also contribute to the growth of the platform. In Table 5, we report the parameter estimates of nonnetwork factors from the joint estimation.

In terms of the time trends, for both the buyer and the seller model, neither the linear nor quadratic terms are statistically significant. This means that the growth of buyers and sellers is driven by other factors such as installed base or variety.

As can be seen from Table 5, product price does not have a significant effect on buyer growth. This might be because Taobao has been positioned as a low-price platform from the very beginning and has successfully established this price image. In addition, based on our discussion with Taobao managers, it turns out that absolute price levels may not matter much as long as Taobao has lower prices than other options, typically physical stores. By contrast, product price does affect seller growth in a significant manner. Specifically, a 10% increase in product price will lead to a 0.48% increase in sellers. Sellers do care about price levels because they affect their profits directly.

<sup>24</sup> We thank an anonymous reviewer for prompting this discussion.



Figure 2(a) (Color online) Evolution of Cross-Network Effects

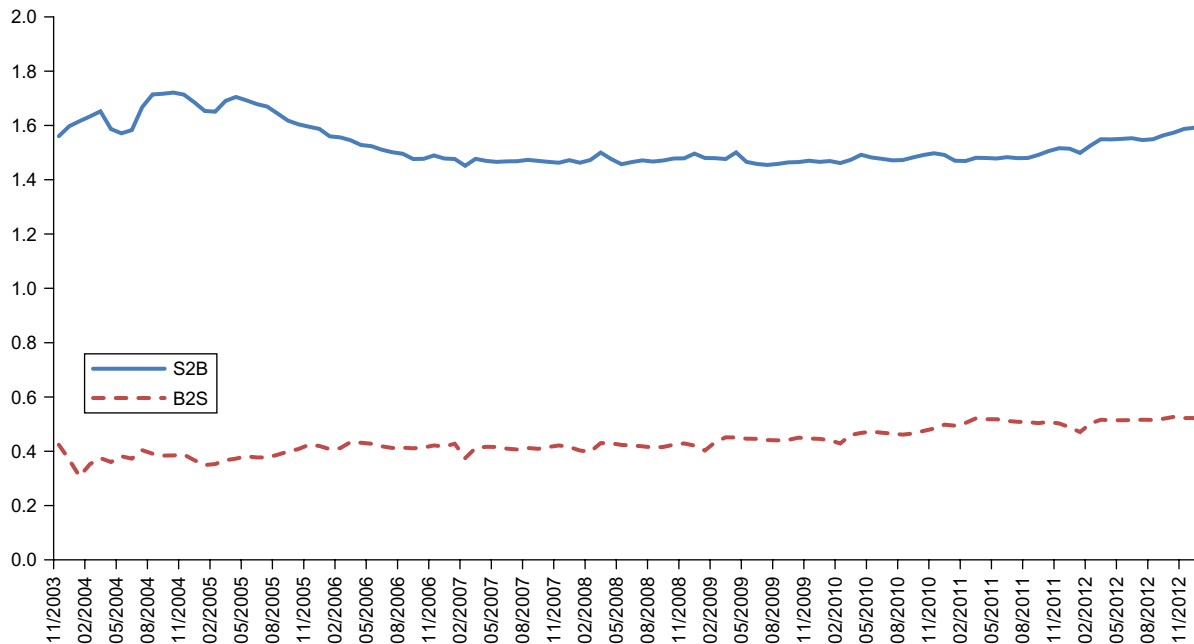
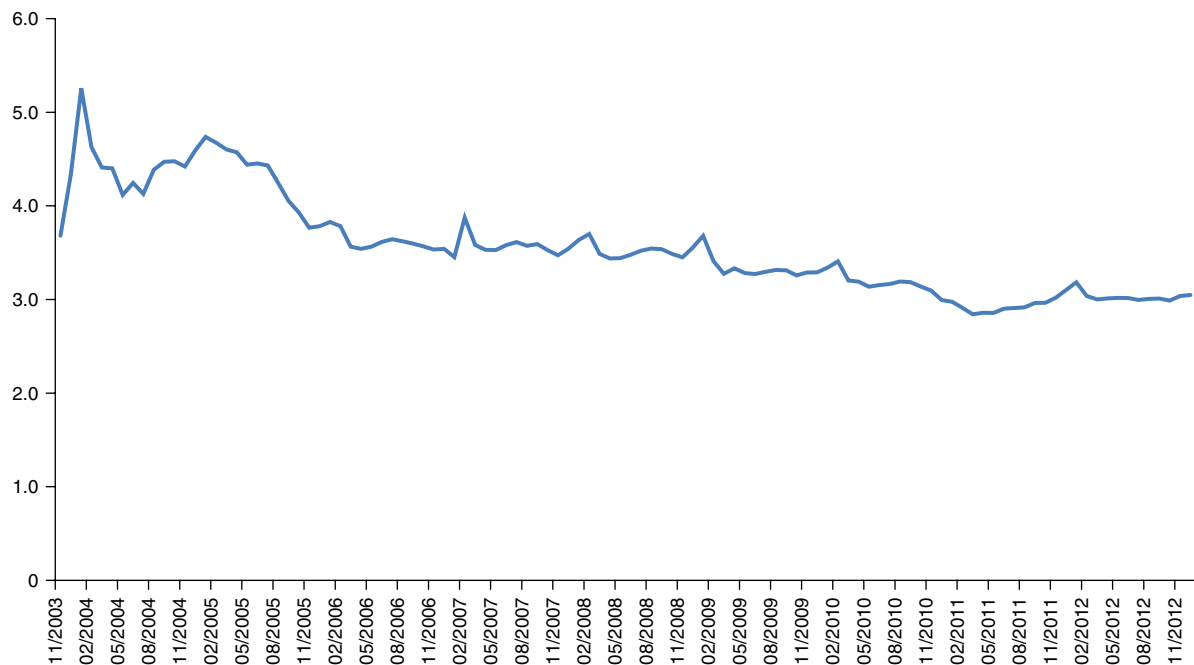


Figure 2(b) (Color online) Ratio of Cross-Network Effects: Sellers on Buyers/Buyers on Sellers



Product variety has a large, significant, and positive effect on buyer growth, next only to the CNE. When the product variety index increases by 1%, new buyers will increase by 1.31% (SD = 0.10%). Besides the large installed base of users (buyers and sellers), product variety is another biggest differentiator between Taobao and all other retailing channels. Many consumers patronize Taobao because they can buy nearly everything there (“Taobao” literally means

“treasure hunt” in Chinese). As Taobao’s positioning catchphrase goes, “there is no treasure that cannot be hunted out” in Taobao. Thus, it may not be surprising that product variety has a large effect on buyer growth.

Buyer quality has a significant and positive effect on seller growth. When buyer’s quality increases by 10%, new sellers will grow by 0.63% (SD = 0.37%). Given the buyer installed base, when buyers make

Figure 3(a) (Color online) Relative Contribution to Buyer Growth

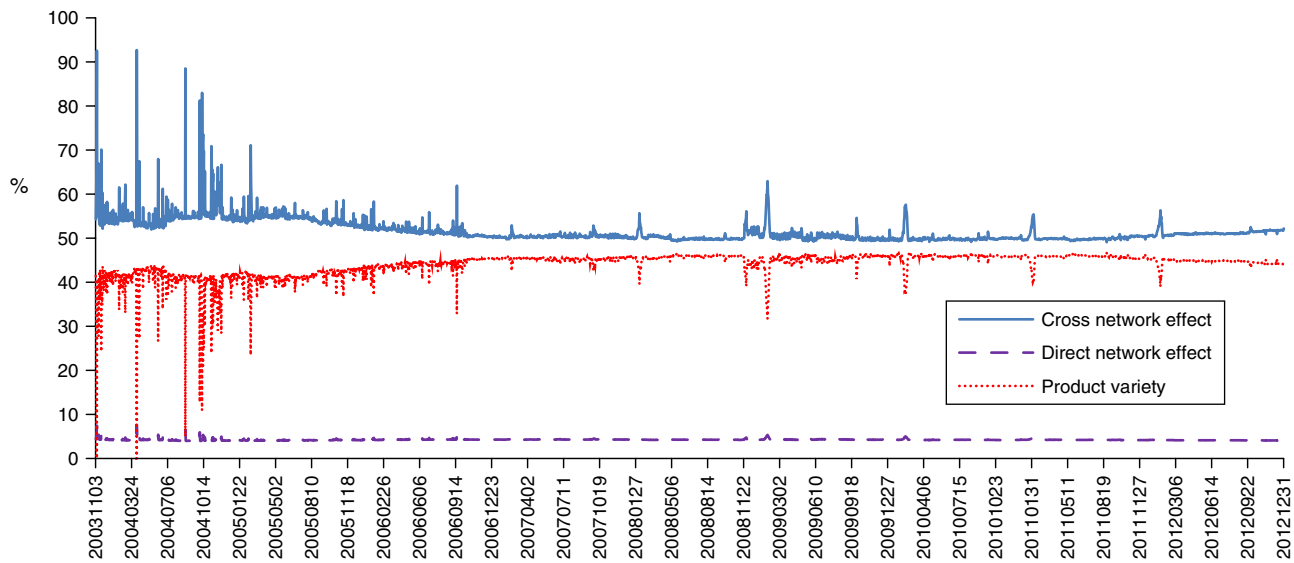
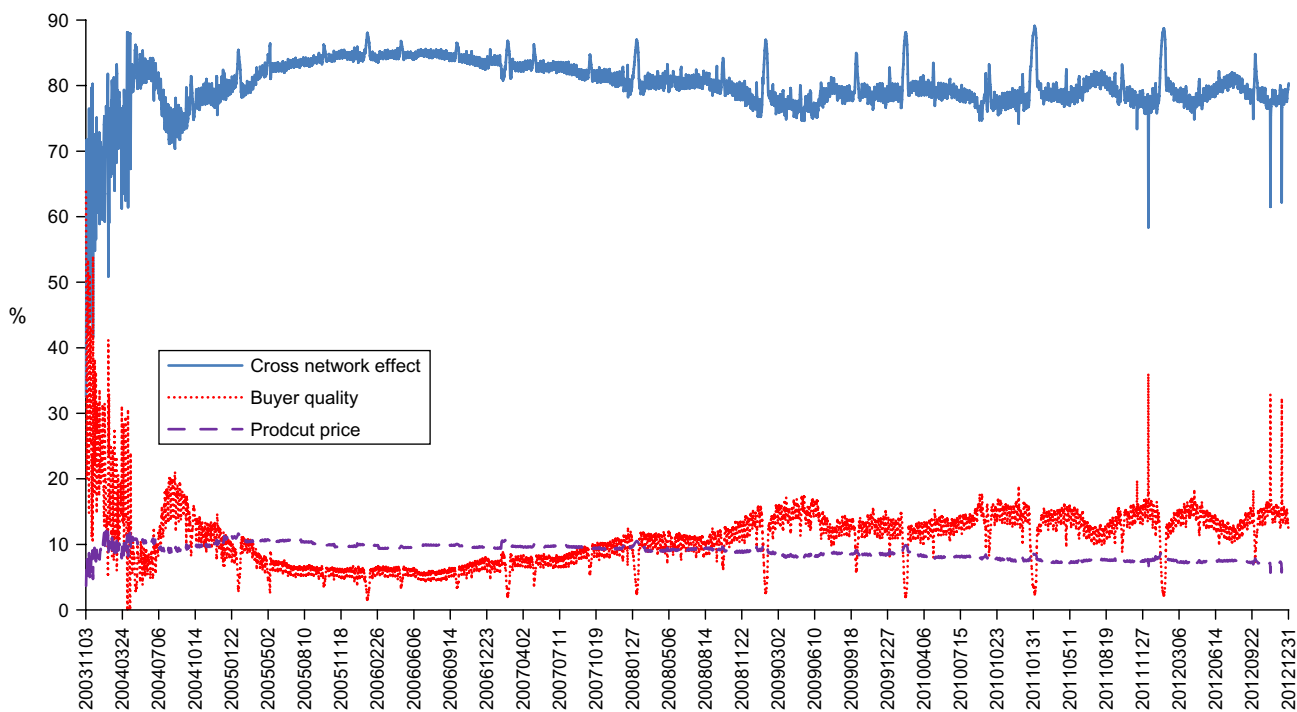


Figure 3(b) (Color online) Relative Contribution to Seller Growth



more transactions, it definitely increases the platform's attractiveness to sellers. Since seller growth will lead to more buyers, it is important for the platform to take measures to induce buyers to transact more at the platform.

Most holidays have a significant dampening effect on buyer and seller growth, particularly the latter. Interestingly, sellers are more responsive to holidays and seller registrations go down dramatically on all holidays. Buyer registrations go down substantially

during long holidays such as the Chinese Lunar New Year, National Day, and Labor Day and do not change much during other short holidays. The deepest drop occurs on the Chinese Lunar New Year when new seller registrations are 49.9% lower and new buyer registrations are 27.0% lower than other days. China's National Day Holiday is another low day with seller registrations going down by 29.9% and buyer registrations decreasing by 11.9%.

Buyers are mostly likely to register on Monday, followed by Tuesday, Wednesday, and Friday, and least likely to register on weekends. There are 7.3% more buyer registrations on Monday, 5.1% more on Tuesday, and 4.8% more on Wednesday and Friday than on Sunday. Sellers are most likely to register on Tuesday through Thursday, followed by Monday and Friday, and least likely to register on weekends. There are about 15% more seller registrations on Tuesday through Thursday, 10% more on Monday, and 9.5% more on Friday than on Sunday.

The Double 11 promotion has a large and significant positive impact on buyer growth, but a large and significant negative impact on seller growth. On that day, buyer registrations increase by 72.5% because buyers want to take advantage of Taobao's biggest annual price promotion. On the other hand, seller registrations decrease by 17.1%. This is because it takes several days for seller registration to be verified and approved and sellers thus advance their registrations so they can sell on the biggest promotion day. Surprisingly, the Double 12 promotion does not significantly affect buyer and seller registration. One reason might be that it is too close to the Double 11 promotion, not giving both buyers and sellers enough time to fully absorb the previous promotional effect.

**7.1.4. Relative Contribution of Network and Nonnetwork Factors.** To compare the relative contributions of network and nonnetwork factors to the platform's development, we compare the relative magnitudes of their elasticities in the growth of new buyers and new sellers. Gandal et al. (2000) use a similar approach to compare the relative effectiveness of hardware price cuts versus software provision in driving hardware adoptions. For buyer growth, we focus on three statistically significant factors—seller installed base, buyer installed base, and product variety. For seller growth, we focus on three statistically significant factors—buyer installed base, buyer quality, and product price.

In Figure 3(a), we plot the evolution of cross network, direct network, and nonnetwork effects on the growth of buyers. Through the platform's entire history, the cross-network factor has been the primary driving force for buyer's growth. However, its effect is declining gradually. Product variety, on the other hand, is exercising increasing influence in buyer growth. The effect of buyer's installed base or DNE is stable over time. The decomposition of the three effects show that, by December 2012, the CNE accounts for 52% of the growth with the balance coming from product variety (44%) and direct network effect (4%).

In Figure 3(b), we plot the evolution of network and nonnetwork effects on the growth of sellers. Similarly, network factor has been the dominant force for

seller's growth. However, its effect is declining over the platform's life cycle. The effect of product price is relatively stable. Buyer quality, on the other hand, has a growing impact over time. In the first half of the data period, buyer quality is the third most important factor, lower than product price, whereas in the second half of the period, it rises to be the second important factor, exceeding the impact of product price.

The finding of declining network effect is at variance with that in the U.S. video game console market where expanding software variety (CNE) becomes more effective over time (Clements and Ohashi 2005) as well as in the PDA market (Nair et al. 2004) where software provision has a growing effect on hardware adoption. This difference may be due to the fact that we focus on platforms (where the intermediary does not produce or own any goods) and/or the fact that we model nonnetwork factors explicitly and/or due to the specific institutional setting in our study (Internet commerce, Chinese market, etc.).

## 7.2. Robustness Checks

We conduct a series of robustness checks, including the functional form of time trend, DNE, and CNE, the scale factor for buyer's and seller's potential market sizes, the definition of seller's potential market size, the discount factor for seller's installed base, the registration approval duration for buyers and sellers, and definition of buyer quality. In the interest of brevity, we do not report the results of these robustness checks in the main paper. We collect some of them into appendices, and the rest are available from the authors on request.

**7.2.1. Functional Form of Time Trend, DNE, and CNE.** We compare six models with different combinations of time trend, DNEs, and CNEs as follows:

Model 1: Without time trend or DNE, but with  $\text{year} \times \text{month} \times \text{CNEs}$ .

Model 2: Without DNE, but with time trend and  $\text{year} \times \text{month} \times \text{CNEs}$ .

Model 3: Without time trend, but with constant DNE and  $\text{year} \times \text{month} \times \text{CNEs}$ .

Model 4: With time trend, constant CNE, and  $\text{year} \times \text{month} \times \text{DNEs}$ .

Model 5: With time trend, constant DNE, and  $\text{year} \times \text{month} \times \text{CNEs}$  (proposed model).

Model 6: With  $\text{year} \times \text{month}$  fixed effects, constant CNE, and constant DNE.

We include a linear and a quadratic time trend in Models 1 through 5, and the results of these models are in Online Appendix A (available as supplemental material at <https://doi.org/10.1287/mksc.2016.0976>). We find that after accounting for CNEs and time trend, DNE becomes either insignificant or much smaller (between one-seventh to one-twelfth) than

CNEs (Online Appendix Tables A1 and A2). In addition, as can be seen from Tables A1 and A2, the choice of specification does materially affect the CNEs.

**7.2.2. Scale Factors for the Potential Market Size (Buyer and Seller).** We tried different scale factors for the buyer and seller potential market size (see §5.2.1). For the buyer potential market size, we used a scale factor of 1, 1.5, and 2 (we use 1.3 in the main model). For the seller potential market size, we used a scale factor of 0.05, 0.20, and 1 (we use 0.1 in the main model) as also the number of (individual) Internet users. We find the change of scale factors only shifts the intercepts up or down and does not affect the estimates of other parameters much (this is consistent with previous work; see Chu et al. 2007, Chu and Chintagunta 2009, Chu 2013). We also looked at the ratio of the CNEs based on the number of Internet users for the seller market size and the CNEs based on the scaled number of households for the seller market size and found the mean to be 0.991 (SD = 0.011), suggesting that there is no material change to our findings. Details can be found in Online Appendix B.

**7.2.3. Discount Factor for the Seller Installed Base.** We also estimated our model without adjusting for the difference between registered sellers and normal state sellers (see §5.2.1). In other words, we use the cumulative sum of all registered sellers as the seller installed base. We find that this affects only the estimate of the intercept. There is no material change in the other coefficients. The mean of the ratios of CNEs based on the discounted seller's installed base to those based on nondiscounted seller's installed base is 1.006 and the standard deviation is 0.002 (min = 1.003 and max = 1.009), and all other coefficients remain nearly identical.

**7.2.4. Buyer and Seller Registration Approval Duration.** As noted in §6.2, there is a difference in how quickly buyer registrations and seller registrations are approved by Taobao. We have used zero days for buyers to appear in the installed base for sellers and new buyers to consider and four days for sellers to appear in the seller installed base for buyers and new sellers to consider. We estimated our model with different approval times spanning the entire range of approval times. Specifically, for buyers, we looked at a one day approval and for sellers we looked at two, three, five, six, and seven day approval. We find that our estimates are not sensitive to the choice of approval period. We computed the mean absolute percentage error (MAPE) based on different lags and found they only differ in the fifth decimal.

**7.2.5. Measurement of Buyer Quality.** Recall that we used the number of transactions per 100 buyers in the installed base as a measure of buyer quality (§5.2.2). We also used alternative measures of buyer quality—the number of transactions per transacting buyer and the percentage of transacting buyers in the installed base. We obtained similar results on the network factors and nonnetwork factors (see the chart in Online Appendix C).

**7.2.6. The Role of Initial Conditions.** The early years of Taobao were characterized by a slightly different competitive situation (see §3) and fluctuations in the average item price and transaction value relative to its later years. To make sure that our steady-state estimates were not affected by these factors, we reestimated the model for two data periods—2003–2005 and 2006–2012. We find that the estimated CNEs do not differ significantly, especially for the latter period.

**7.2.7. Market Share vs. Quantity.** Our choice of dependent variable is the probability of signing up for an individual buyer or seller, which is then aggregated to a market share and taken to the estimation. The use of the market share in the estimation allows us to accommodate the changing market size on both the buyer (individuals in China with Internet access) and sellers (10% of all households in China). This allows us to scale the dependent variable in a manner that makes it comparable over time. We also estimated a model formulation using just the count of new buyers and sellers via a log-linear regression of log (new buyers or new sellers) on the same set of independent variables as in our proposed model. We do not see a material difference for our main results.

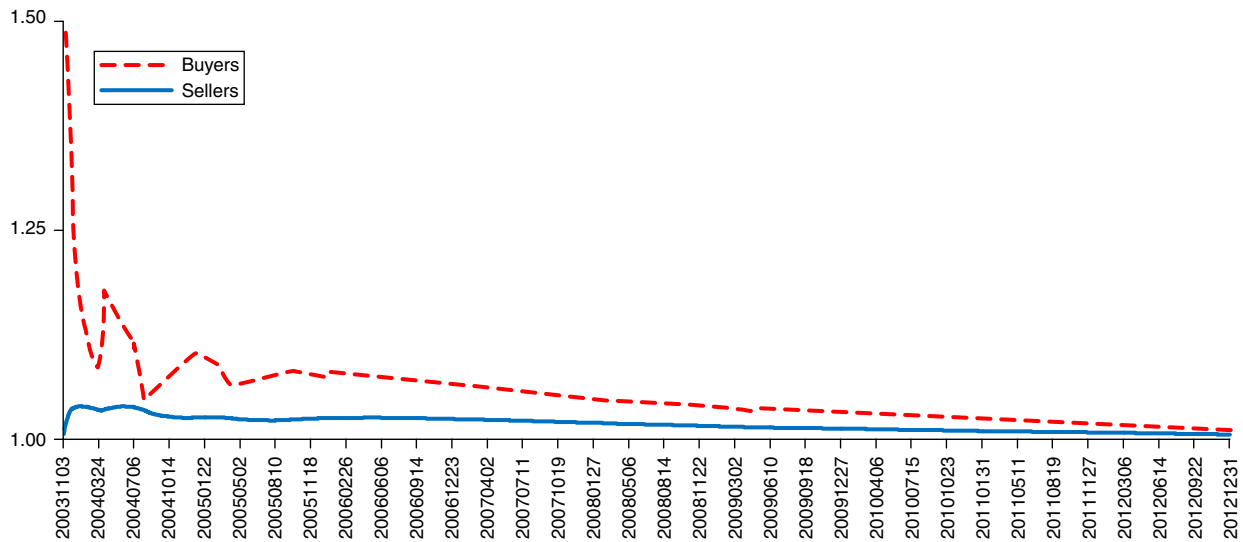
### 7.3. Managerial Implications

Managers of platforms are typically concerned with understanding the primacy of one side versus the other. If they know the size and asymmetry in the CNEs, they can allocate resources more efficiently. In our case, we find that sellers are relatively more important and thus should get more resources (including nonfinancial resources such as managerial attention). Interestingly, this result is generally consistent with institutional practice in (offline) retail settings. For example, the literature on shopping mall development (the mall acts as the “platform,” bringing stores and consumers together), suggests that the mall developer's prime focus early on is to find strong sellers (typically called anchor stores) rather than on consumer traffic (Bean et al. 1988, Pashigian and Gould 1998, Gould et al. 2005, Vitorino 2012).

In addition, platform managers can also try and influence factors that are more under their control. We



Figure 4 (Color online) Ratios of Simulated/Observed Installed Base by Seeding Buyers and Sellers on the First Day



focus on three such factors—initial sellers and buyers, product variety, and buyer quality—to show the impact that changes in these factors can have on the growth of the network. We also discuss qualitatively the impact of our findings on Taobao’s practices.

**7.3.1. Seeding More Sellers and Buyers in the Introduction Stage.** The impetus for this simulation comes from the fact that our model estimates show the existence of positive CNEs on both sides of the platform. This suggests that having more buyers and sellers in the *early periods* of the platform’s operation will have a larger and longer-lasting impact on the platform’s growth. The implication for the platform is that it should try to encourage buyers and sellers to register in the introduction stage of the platform’s operation via marketing and economic incentives such as subsidized pricing, cash bonus, referral bonus, etc. It is noteworthy that platforms such as Alipay, Uber, GrabTaxi, and Didi-Kuaidi took such an approach.<sup>25</sup> Matchmaking platform AshleyMadison.com even went to the extreme of faking the installed base (of women) to incentivize more men to join.<sup>26</sup>

To simulate the impact of early members, we seed one seller and 60 buyers (the overall ratio between sellers and buyers in the data is about 1:60) on the first day of the data period to see how this will affect the installed base of buyers and sellers over time. We plot the ratios of simulated over observed installed base(s)

in Figure 4. As can be seen from Figure 4, relative to the observed installed base, seeding additional buyers and sellers right at the beginning has a significant and long-lasting impact in terms of growing the installed base(s). The impact is larger on buyers than on sellers, due to the much bigger CNE of seller installed base on buyers. The effects decline over time but remain apparent even at the end of the data period.

**7.3.2. Changes in Product Variety.** Changes in product variety have both direct effects and indirect effects. Since buyers value product variety, a deterioration (an improvement) in product variety will lead to fewer (more) new buyers to register on the platform. This is the direct effect. Fewer (more) buyer registrations will reduce (increase) *buyer’s* installed base in all future periods, which will discourage (encourage) new *sellers* to register, which will decrease (increase) *seller’s* installed base in all future periods, which will lead to fewer (more) new *buyers*. This forms the indirect effect. On the other hand, although seller registrations are not directly affected by changes in product variety, they will be indirectly affected by the resultant changes in the buyer installed base brought by changes in buyer registrations.

We disentangle the direct and indirect effects of a change in product variety using two scenarios. In the first scenario, we fix the buyer and the seller installed bases at their observed values in the data (direct effect), and in the second, we allow buyer and seller installed bases to change in the future by responding to changes in new buyer and new seller registrations. For each scenario, we simulate new buyers and new sellers using the cross-network and direct network parameter estimates as well as nonnetwork parameter estimates reported in Table 5 and compute the corresponding buyer’s installed base and seller’s installed

<sup>25</sup> <http://www.wsj.com/articles/inside-ubers-fight-with-its-chinese-nemesis-didi-kuaidi-1441234010>.

<sup>26</sup> <https://www.washingtonpost.com/news/the-intersect/wp/2015/08/25/ashley-madison-faked-female-profiles-to-lure-men-in-hacked-data-suggest/>.

base for each day from November 1, 2003 to December 31, 2012. The first scenario does not account for the changes in new buyers and new sellers brought by the changed seller and buyer installed bases, so it measures the effect of product variety changes net of network effect. The second scenario allows sellers to respond to changes in the buyer's installed base (CNE) and in the seller's installed base (DNE), and buyers to respond to changes in the seller's installed base (CNE) and in the buyer's installed base (DNE), so it measures the total effect, i.e., direct and indirect effects of product variety. The difference between these two scenarios can be taken as the effect of installed base, primarily CNE.

We simulate the effect of reducing product variety by setting product variety level to zero, which is akin to forcing all products sold on Taobao to be in one category. In Figure 5, we plot the ratio of simulated seller installed base over observed seller installed base for the scenario without network effect and the scenario with network effect, as well as the ratio of simulated buyer's installed base over observed buyer's installed base for these two scenarios.

Several observations are in order. First, minimizing product variety will substantially discourage buyer and seller registrations, leading to considerable reductions in buyer and seller installed bases. The reduction in the installed base was small in the beginning, but became very dramatic as time went by. By the end of the period, the buyer installed base without any product variety would be only about 5% of the actual buyer installed base, and the seller installed base would be around 26% of the actual seller installed base. Second, the CNE compounds the effect of product variety, both on buyers and sellers. The simulated buyer installed base would be around 26% of the actual installed base, if there were no CNEs and DNE, primarily CNEs, as compared to around 5% with CNEs and DNE. Since product variety does not directly affect seller registration, the reduction in seller's installed base is totally due to CNEs and DNE, primarily CNEs. Third, product variety has a much larger impact on buyers than on sellers, both directly and indirectly. The buyer installed base would be much more negatively affected by reducing product variety than the seller installed base.

**7.3.3. Changes in Buyer Quality.** Similarly, changes in buyer quality have both direct effects and indirect effects. Since sellers value buyer quality, an increase in buyer quality will lead to more new sellers to register on the platform. This is the direct effect. More seller registrations will increase the seller installed base in all future periods, which will encourage new buyers to register, which will increase buyer's installed base in all future periods, which will lead to more new sellers. This forms the indirect effect. On the other hand,

although buyer registrations are not directly affected by changes in buyer quality, they will be indirectly affected by the resultant changes in seller's installed base brought on by changes in seller registrations, and to a lesser extent by the resultant changes in buyer's installed base.

The direct and indirect effects of a change in buyer quality can also be disentangled in the same way as the change in product variety. We simulate the effect of doubling buyer quality. In Figure 6, we plot the ratio of the simulated seller installed base over observed seller installed base for the scenario without network effect and the scenario with network effect, as well as the ratio of the simulated buyer installed base over the observed buyer installed base for these two scenarios.

We observe the following. First, enhancing buyer quality will encourage sellers and buyers to register, leading to sizable increases in the seller and buyer installed bases. The seller installed base would be nearly 10%–14% higher than the actual seller installed base, with first an increasing and then a flat effect over time. The buyer installed base would be 10%–30% higher than the actual installed base, with first an increasing and then a decreasing effect over time. Second, CNE compounds the effect of buyer quality, both on sellers and buyers. The simulated seller installed base would be about 6%–9% higher than the actual installed base, if there were no CNE or DNE, primarily CNE, as compared to nearly 10%–14% higher with CNEs and DNE. Since buyer quality does not directly affect buyer registration, the increase in buyer installed base is completely due to network effects, primarily the CNE. Third, although buyer quality does not affect buyer registration directly, it has a larger impact on buyers than on sellers, except for the beginning month. This is because the seller installed base has a much larger effect on buyers than vice versa, and the CNE on buyers outweighs the direct effect of buyer quality on sellers.

The last two simulations also demonstrate that CNEs are a double-edged sword. They can accelerate or decelerate outcomes. Thus, it is crucial for platform managers to understand, quantify, and manage the trajectory of the installed base.

**7.3.4. Impact at Taobao.** We shared our analysis and findings with Taobao. One aspect of their reaction is particularly noteworthy. The generally held wisdom in the company was that buyers were more important than sellers because they had a bigger impact on sellers rather than the other way around. Our finding—that the seller on buyer CNE was 3.6 times as big as the CNE of buyer on seller—was seen as a very surprising finding. In a separate conversation with Savio Kwan, the ex-COO of the Alibaba Group, we discovered the reason for this view. He

Figure 5 (Color online) Simulated Impact of Product Variety on Buyer and Seller Installed Bases

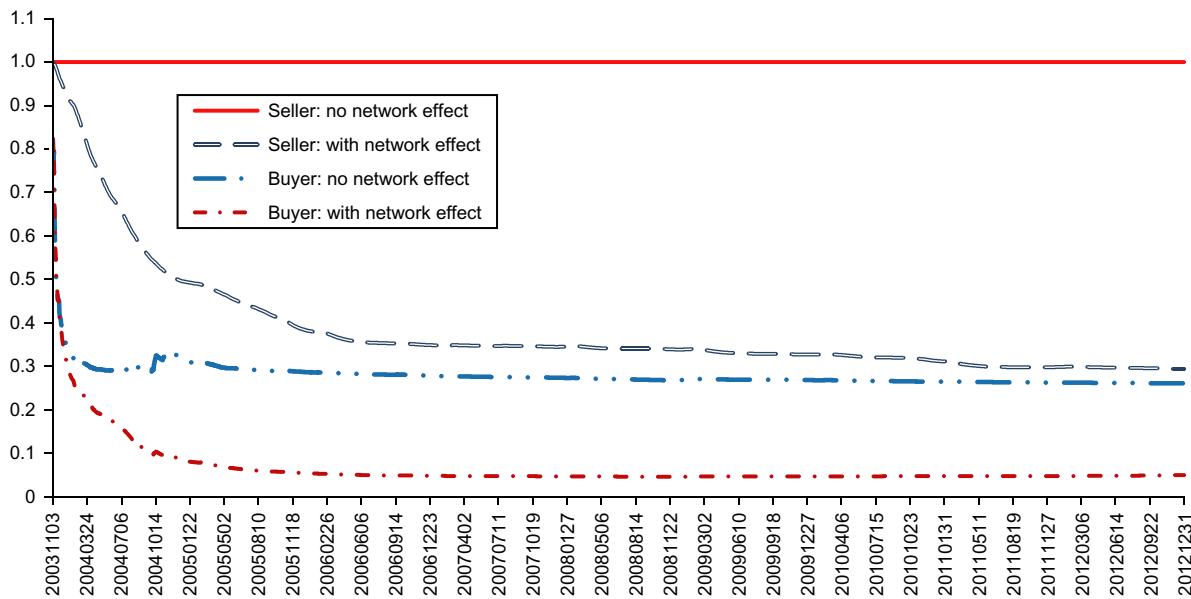
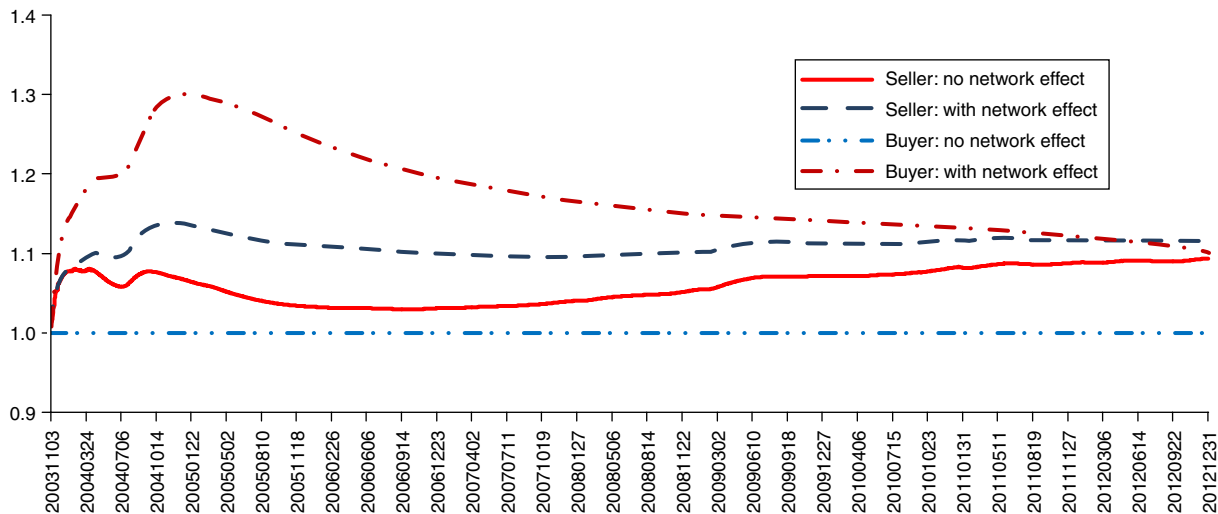


Figure 6 (Color online) Simulated Impact of Buyer Quality on Buyer and Seller Installed Bases



noted that in the early days of Taobao, the belief was that buyers had the purchasing power and hence needed to be nurtured over sellers (who were after all making profits and so were getting rewarded for participating on the platform). This belief had become rooted in company culture over time.

As a result of our findings, the company's managers started to become more "seller friendly." They lowered the emphasis on seller ratings and generally focused on improving seller welfare. In addition, they started exploring mechanisms to reactivate buyers to improve buyer quality. To improve product variety, on the margin, they encouraged sellers who provide more variety (relative to what was already available on the platform).

## 8. Conclusion

This paper adds to the small but growing empirical literature on platforms (or two-sided markets), especially in online settings. We use novel data that span the entire history of the world's largest C2C platform—Taobao in China—to model its growth. Specifically, we take a utility-based approach to track the growth as a function of network and nonnetwork factors. We focus on the quantification of the CNEs over the platform's life cycle and compare the relative importance of network and nonnetwork factors in the platform's growth. We find a large, significant, and positive CNE on both sides of the platform market, but the CNE is asymmetric with the installed base of sellers having a much larger effect on the growth of

buyers than vice versa. We also find a positive and significant albeit small DNE on the buyer side and a negative but insignificant DNE on the seller's side. The growth in the number of buyers is driven primarily by the seller installed base and product variety with increasing importance of product variety. By contrast, the growth in the number of sellers is driven by the buyer installed base, buyer quality, and product price with increasing importance of buyer quality. We further find that the CNE of sellers on buyers increases and then decreases to reach a stable level. By contrast, the CNE of buyers on sellers is relatively stable. Finally, we carry out analyses to show how seeding more sellers and buyers in the introduction stage, increasing product variety and buyer quality have a material direct and indirect effect on the installed base.

Our paper suffers from a few limitations, mostly driven by the nature of the available data. First, our measures of price and product variety are aggregates across the platform. Second, we cannot control for differences across buyers and sellers given the lack of individual level data. Similarly, our model also uses data aggregated over product categories and therefore the estimates cannot be used for category specific inference or policy counterfactuals. Third, we assume that both sellers and buyers are myopic in their decision to join the platform. In the Taobao setting, this is perhaps not a first-order issue because the platform's free-pricing policy together with nearly hassle-free registration greatly reduces sellers' and especially buyers' risk of joining and transacting on the platform and thus their incentives to look forward. Fourth, we do not have seller quality in the buyer's model because of data unavailability (see the caveat to this in §4.2). We hope that future research can address these limitations.

### Supplemental Material

Supplemental material to this paper is available at <https://doi.org/10.1287/mksc.2016.0976>.

### Acknowledgments

The authors are very grateful to Taobao.com for providing data for this research. The authors also thank Tat Chan, John Deighton, Savio Kwan, Thomas Otter, S. Sriram, Xu Zhang, participants at University of Cambridge, Erasmus University, Goethe University Frankfurt, University of Houston, KU Leuven, London Business School, LMU Munich, University of Michigan, Nanyang Technological University, University of Oxford, Temple University, Tilburg University, University of Texas at Dallas, University of Connecticut, Xi'an Jiaotong University, Shanghai University of Finance and Economics, the 2015 University of Leeds CDRC International Big Data Symposium, the 2014 NYU Center for U.S.–China Relations Conference, the 2014 AMA-ECMI-EMAC Marketing Symposium, the 2013 Yale China India

Insights Conference, the 10th and 11th Marketing Dynamics Conferences, and the 9th Invitational Choice Symposium at Erasmus University for valuable comments and suggestions. Finally, the authors gratefully acknowledge financial support from the Asian Consumer Insights Institute at Nanyang Business School, the National University of Singapore [HSS Seed Fund R-316-000-102-646], and the Phelps Faculty Fund at the Ross School of Business, University of Michigan.

### References

- Akerberg DA, Gowrisankaran G (2006) Quantifying equilibrium network externalities in the ACH banking industry. *RAND J. Econom.* 37(3):738–761.
- Argentesi E, Filistrucchi L (2007) Estimating market power in a two-sided market: The case of newspapers. *J. Appl. Econometrics* 22(7):1247–1266.
- Armstrong M (2006) Competition in two-sided markets. *RAND J. Econom.* 37(3):668–691.
- Bass FM (1969) A new product growth for model consumer durables. *Management Sci.* 15(5):215–227.
- Bean JC, Noon CE, Ryan SM, Salton GJ (1988) Selecting tenants in a shopping mall. *Interfaces* 18(2):1–9.
- Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.
- Boatwright P, Nunes JC (2001) Reducing assortment: An attribute-based approach. *J. Marketing* 65(3):50–63.
- Briesch RA, Chintagunta PK, Fox EJ (2009) How does assortment affect grocery store choice? *J. Marketing Res.* 46(2):176–189.
- Caillaud B, Jullien B (2003) Chicken and egg: Competition among intermediation service providers. *RAND J. Econom.* 34(2):309–328.
- Chandra A, Collard-Wexler A (2009) Mergers in two-sided markets: An application to the Canadian newspaper industry. *J. Econom. Management Strategy* 18(4):1045–1070.
- Chen Y, Xie J (2007) Cross-market network effect with asymmetric customer loyalty: Implications for competitive advantage. *Marketing Sci.* 26(1):52–66.
- Chu J (2013) Quantifying network equity with sales data: A structural approach. *Internat. J. Res. Marketing* 30(1):19–35.
- Chu J, Chintagunta PK (2009) Quantifying the economic value of warranties in the U.S. server market. *Marketing Sci.* 28(1):99–121.
- Chu J, Chintagunta PK, Vilcassim NJ (2007) Assessing the economic value of distribution channels: An application to the personal computer industry. *J. Marketing Res.* 44(1):29–41.
- Clements MT, Ohashi H (2005) Indirect network effects and the product cycle: Video games in the U.S., 1994–2002. *J. Indust. Econom.* 53(4):515–542.
- Dubé J-PH, Hitsch GJ, Chintagunta PK (2010) Tipping and concentration in markets with indirect network effects. *Marketing Sci.* 29(2):216–249.
- Fan Y (2013) Ownership consolidation and product characteristics: A study of the U.S. daily newspaper market. *Amer. Econom. Rev.* 103(5):1598–1628.
- Gandal N, Kende M, Rob R (2000) The dynamics of technological adoption in hardware/software systems: The case of compact disc players. *RAND J. Econom.* 31(1):43–61.
- Gould ED, Pashigian BP, Prendergast CJ (2005) Contracts, externalities, and incentives in shopping malls. *Rev. Econom. Statist.* 87(3):411–422.
- Jin GZ, Rysman M (2015) Platform pricing at sports card conventions. *J. Indust. Econom.* 63(4):704–735.
- Kaiser U, Wright J (2006) Price structure in two-sided markets: Evidence from the magazine industry. *Internat. J. Indust. Organ.* 24(1):1–28.
- Liu H (2010) Dynamics of pricing in the video game console market: Skimming or penetration? *J. Marketing Res.* 47(2):428–443.



- Mahajan V, Muller E, Bass FM (1995) Diffusion of new products: Empirical generalizations and managerial uses. *Marketing Sci.* 14(3-suppl):G79–G88.
- Morningstar (2014) Powerful network effect makes Alibaba a rare find. (July 9), <http://www.morningstar.com/topics/t/94162328/powerful-network-effect-makes-alibaba-a-rare-find.htm>.
- Nair H, Chintagunta P, Dubé J-P (2004) Empirical analysis of indirect network effects in the market for personal digital assistants. *Quant. Marketing Econom.* 2(1):23–58.
- Ohashi H (2003) The role of network effects in the U.S. VCR market, 1978–1986. *J. Econom. Management Strategy* 12(4):447–494.
- Park S (2004) Quantitative analysis of network externalities in competing technologies: The VCR case. *Rev. Econom. Statist.* 86(4):937–945.
- Pashigian BP, Gould ED (1998) Internalizing externalities: The pricing of space in shopping malls. *J. Law Econom.* 41(1):115–142.
- Pattabhiramaiah A, Sriram S, Sridhar H (2013) Rising prices under declining preferences: The case of the U.S. print newspaper industry. Working paper, University of Michigan, Ann Arbor.
- Petrin A (2002) Quantifying the benefits of new products: The case of the minivan. *J. Political Econom.* 110(4):705–729.
- Petrin A, Train K (2010) A control function approach to endogeneity in consumer choice models. *J. Marketing Res.* 47(1):3–13.
- Rochet J-C, Tirole J (2003) Platform competition in two-sided markets. *J. Eur. Econom. Assoc.* 1(4):990–1029.
- Rochet J-C, Tirole J (2006) Two-sided markets: A progress report. *RAND J. Econom.* 37(3):645–667.
- Rossi PE (2014) Even the rich can make themselves poor: A critical examination of IV methods in marketing applications. *Marketing Sci.* 33(5):655–672.
- Rysman M (2004) A study of the market for yellow pages. *Rev. Econom. Stud.* 71(2):483–512.
- Rysman M (2009) The economics of two-sided markets. *J. Econom. Perspect.* 23(3):125–143.
- Seamans R, Zhu F (2014) Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers. *Management Sci.* 60(2):476–493.
- Shankar V, Bayus BL (2003) Network effects and competition: An empirical analysis of the home video game industry. *Strategic Management J.* 24(4):375–384.
- Song M (2013) Estimating platform market power in two-sided markets with an application to magazine advertising. Working paper, University of Rochester, Rochester, NY.
- Sriram S, Manchanda P, Bravo ME, Chu J, Ma L, Song M, Shriver S, Subramanian U (2015) Platforms: A multiplicity of research opportunities. *Marketing Lett.* 26(2):141–152.
- Sun L, Rajiv S, Chu J (2015) Beyond the more the merrier: Variety effects in system markets. *Internat. J. Res. Marketing*, Forthcoming.
- Vitorino MA (2012) Empirical entry games with complementarities: An application to the shopping center industry. *J. Marketing Res.* 49(2):175–191.
- Wang HH (2012) Looking at the world upside down. Wang HH, ed. *The Chinese Dream: The Rise of the World's Largest Middle Class and What It Means to You*, 2nd ed. (CreateSpace Independent Publishing Platform, Seattle), 153–170.
- White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4):817–838.
- Wilbur KC (2008) A two-sided, empirical model of television advertising and viewing markets. *Marketing Sci.* 27(3):356–378.