Decomposing the Congestion Effect and the Inference Effect in Two-Sided Networks: A Field Experiment

June 18, 2008

Catherine Tucker

Assistant Professor of Marketing
MIT Sloan School of Management
1 Amherst Street, E40-167
Cambridge, MA 02142
Phone: (617) 252-1499
Fax: (617) 258-7597
Email: cetucker@mit.edu

Juanjuan Zhang

Assistant Professor of Marketing
MIT Sloan School of Management
1 Amherst Street, E40-171
Cambridge, MA 02142
Phone: (617) 452-2790
Fax: (617) 258-7597
Email: jjzhang@mit.edu

Decomposing the Congestion Effect and the Inference Effect in Two-Sided Networks: A Field Experiment

Abstract

Are businesses more likely or less likely to enter a two-sided market crowded with competitors? On the one hand, a strong rival presence may dissipate payoff (a "congestion effect"). On the other hand, rivalry may signal high demand (an "inference effect"). To distinguish between these two effects, we use field experiment data from a B2B website that brings together buyers and sellers of used equipment and real estate opportunities. Before each seller made a posting request, the website randomized whether to disclose the number of buyers and/or sellers, and the exact number to disclose. We find that when presented together with the number of buyers, a larger number of sellers makes sellers less likely to list their products, indicating a negative congestion effect. However, when presented in isolation, the number of sellers has a significantly less negative effect on entry, indicating a positive inference effect. More intense buyer search amplifies the moderating role of demand uncertainty. The results suggest that information on the number of users can be a powerful tool to grow two-sided networks. A network can attract more users by advertising dense competition when demand is not transparent, especially in search-intensive markets.

Keywords: Competition, Entry, Inference, Congestion, Decision-making Under Uncertainty, Two-Sided Markets

JEL Classifications: C93, D83, L11, M31

1 Introduction

Let us suppose that a business is considering whether to sell used medical equipment through a B2B website that brings together both buyers and sellers. The website advertises that it has already signed up 200 sellers of used medical equipment. Would knowing this number affect a business's decision to sell products through this website? On the one hand, the business may balk at the idea of ferocious competition. On the other hand, it may be more likely to sign up, reasoning that the 200 other firms may have joined the website because of a large buyer base, or (to the same effect) that the volume of sellers on the website might have attracted a large number of buyers.

This business's dilemma is an example of a broader question: how the numbers of participants in two-sided networks affect potential customers' participation decisions, and whether firms that profit from two-sided networks should publicize their number of users. A two-sided network is, loosely, a common platform of exchange for traders, such as singles looking for a match, recruiting firms and job seekers, and sellers and buyers, whose benefit from joining the network depends on the the number of traders on the other side of the market. Although it has been a popular practice on the internet to publicize the number of users, partly due to the number being easy to calculate, it is unclear whether this practice hinders or accelerates the growth of two-sided networks.

This paper aims to decompose the effects of peer presence on participation in two-sided markets. We identify "demand uncertainty" as a key moderator of how peer presence impacts network entry decisions. In the medical equipment example, the potential seller's demand uncertainty is its uncertainty about the number of buyers that will browse its listing on the website. We distinguish between two effects of peer presence that are governed by the availability of demand information: A "congestion effect" where potential entrants avoid payoff-dissipating competition, and an "inference effect" where potential entrants infer high demand from heavy peer presence, either because high demand has attracted substantial entry or because substantial entry has created high demand.

Most empirical research on two-sided networks is plagued by endogeneity. It is difficult to establish causality when a seller enters a two-sided network packed with other sellers, given the abundance of alternative explanations as to why entry decisions are correlated. Equally challenging is the sparse observations of potential sellers who choose not to participate and of the factors that lead to this nonparticipation decision (see Bradlow and Park (2007) for a discussion of latent auction participants and a model of imputing the competitor set). Meanwhile, previous empirical studies using historical data on entry decisions have not been able to isolate the effects of demand uncertainty, due to the lack of data on the degree of uncertainty or on the variation of uncertainty across time or market.[1]

We circumvent these empirical challenges by using data from a field experiment where potential network entrants were randomly informed about demand. The field experiment was conducted by a website that brings together sellers and buyers of various categories of used goods and real estate properties in one metropolitan area. Before each potential seller decided whether to list their good, the website randomized whether to display the number of buyers and/or sellers and, if so, how many buyers and sellers to claim.

We find that when information on both the number of buyers and the number of sellers is

[1]Substantial recent research in competition and entry has focused on estimating equilibrium models of negatively correlated entry decisions with full demand information (e.g., Seim (2006), Orhun (2007), Zhu, Singh, and Dukes (2005)). A few studies take the opposite approach and estimate the equilibrium assuming that firms lack information about market conditions (e.g., Toivanen and Waterson (2005) and Vitorino (2007)). Both approaches rely on assumptions about demand information availability and then interpret entry correlations in light of such assumptions. This reliance may open two questions. First, it is unusual for researchers to observe the exact information structure of potential entrants. Therefore, the validity of the informational assumptions key to this stream of research is hard to test. Second, even if researchers have precise information on firm's knowledge about market conditions, there is a deeper concern that information acquisition itself may be an endogenous variable (e.g., Hitsch (2006)). This endogeneity problem may further confound the results because firms' decision to acquire information is affected by their (often unobserved to the researcher) knowledge about the products' chance of success, which in turn affects their subsequent entry decisions. We adopt the field experimental approach to address both questions by exogenously controlling the level of market uncertainty and tracing the causal effect of information.

presented, a larger number of sellers reduces a potential seller's posting propensity. However, when information about the number of sellers is presented in isolation, it has a less negative effect on the seller's posting decision. The results suggest that the effect of peer presence on entry can be negative or positive, depending on how much potential entrants rely on peer presence to resolve demand uncertainty.

We then turn to investigate how demand uncertainty interacts with firm's expectations of the behavior of buyers on the other side of the network. We examine how buyers' likelihood of browsing multiple listings affects seller entry behavior. When the website discloses the number of users on both sides of the market, a large number of sellers discourages entry in search-intensive categories. However, when the website supplies information on only the number of sellers, a large number of sellers leads to a greater boost in posting propensities in search-intensive categories. This second result initially seems counter-intuitive as firms would presumably be more concerned of competition where customers are inclined towards comparison shopping. However, in a two-sided network setting with cross-group network externalities, the high density of sellers can be an attraction to buyers who have the need for intensive search. This is analogous to a retailer choosing to locate in a mall with many other rivals if customers enjoy browsing shops and as a consequence prefer malls with more options (e.g., Dudey (1990), Gould, Pashigian, and Prendergast (2005)).

There is a growing body of research that models participation decisions in two-sided networks. Rochet and Tirole (2006) and Armstrong (2006) provide an excellent overview of theoretical developments in this area. This research underlines that the driver of two-sidedness for exchange networks are the transaction costs buyers and or sellers incur when seeking out multiple matching options. Fath and Sarvary (2003) explicitly model these insights and investigate the adoption dynamics in buyer-side exchange platforms, and find it optimal for platforms to subsidize buyers rather than sellers to encourage participation. Chen and Xie (2007) show that in markets with cross-market network effects that customer loyalty can have ambiguous effects in driving demand.

Tucker (2008) investigates how the success of federal government intervention to subsidize the customer base of electronic payment systems led to more adoption by banks. However, how the availability of information about the number of customers affects network growth has been neglected, despite the salience of such information and its increasing use in practice.

Our results illuminate how the strategic release of information can help grow traffic for two-sided networks. Two-sided networks as a contemporary business model have been a magnet to entrepreneurs. Some are able to expand at a furious pace. Match.com started from scratch in 1994 and now has over 15 million members. EBay's revenue exploded from $4 million in 1998 to $7.67 billion in 2007. However, there are also numerous well-funded two-sided networks that have never gained traction. Chemdex.com, despite pioneering the B2B portal model and raising $112.5 million through its IPO, never accumulated enough clients to make profit. Similarly, Amazon Auctions, while feted as an "eBay Killer," was quietly dropped two years past inception after attracting less than two percent of auction listings.[2] Behind all these tales of high-stake hits and flops is the key question of how to grow network participation. This paper suggests ways to grow participation using one of the simplest yet trickiest tools—information on the number of network users itself. In particular, the results suggest that a large number of peer users is more likely to help a network recruit new customers when demand on the other side of the market is less transparent. That is, high popularity among peers can turn out to hurt a network when potential customers are well informed about demand. Furthermore, the decision of whether to publicize the number of users is critically determined by demand uncertainty especially in categories where buyers tend to search intensively.

This paper also sheds light on a growing debate on how competition affects entry. Research in both marketing and industrial organization has emphasized the entry deterrence role of competition, both theoretically and empirically (e.g., Salop (1979), Bresnahan and Reiss (1991), Berry (1992)). However, this received wisdom has been questioned by recent findings of "competition

[2] "Auctions getting lost in Amazon's jungle," CNET News, July 31, 2002.

neglect" (e.g., Camerer and Lovallo (1999), Simonsohn (2006)), and "competition contagion" (e.g., Narasimhan and Zhang (2000), Debruyne and Reibstein (2005)), where firms are indifferent or even more likely to enter heavily congested markets. This paper helps reconcile the controversy by identifying and validating demand uncertainty as a moderator of how entrants respond to existing competition. Meanwhile, while many studies rely on bounded rationality, such as limited iterative thinking capacity (Camerer, Ho, and Chong (2004)), to explain excess entry and high incidence of post-entry failure,[3] our conceptualization of competition provides an angle to interpret non-negative correlations in entry within a rational framework. It can even explain inefficiencies in entry. If potential entrants indeed infer demand from prior firms' entry decisions, early entrants could initiate a socially irrational bandwagon of repeated entry, even if inference is a rational engagement for each individual firm.[4] In this sense, our approach echoes that of Wernerfelt (1995) who reinterprets the compromise effect, another well-known bounded-rationality phenomenon, as rational behavior governed by limited information.

The rest of the paper is organized as follows. §2 describes the field experiment and §3 presents the data. §4 first discusses the main results where sellers react differently to the level of competition, depending on whether they are given demand information. We then discuss the augmented analysis where buyer search intensity amplifies the impact of demand uncertainty. §5 summarizes the paper and discusses potential directions for future research. In addition, the Appendix collects an analytical decomposition of the congestion effect and the inference effect, data to verify experimental randomization, and robustness checks of empirical results.

[3] Urban and Hauser (1993) report that across industries an average of 35% of new products fail after launch.

[4] Please refer to Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992) for models of how individually rational observational learning triggers irrational aggregate decisions due to information externalities.

2 Business Context and Field Experiment

We obtained field experiment data from a B2B website that in appearance resembles craigslist.org.[5] The website provides a common platform for sellers and buyers of used equipment and real estate opportunities to advertise these items and to read the advertisements. The target customers are largely one-person businesses and small-time entrepreneurs. Figure 1 presents the span and size of product categories. More than 40 major metropolitan areas are served. Conversation with website management suggests that most transactions take place locally. This means that each metropolitan area roughly corresponds to an isolated market. The website draws revenues mainly from banner advertisements on their main page, and does not charge sellers for using its posting service or buyers for browsing postings. The website receives a total of 240,000 clicks per day.

Although a fee is not charged, a seller must register and log in to an individual user account at the website, and subsequently fill in a "posting form" to be able to list an item for sale. A seller will therefore post their item and "enter the market" if their expected return from posting exceeds the opportunity cost of time spent filling out the forms. After a posting is submitted, it is listed chronologically on the website. Buyers can view postings without signing up for the website. Other things being equal, the return on posting for a seller increases with the number of buyers, in part due to a higher chance of match.

There has been a trend for businesses to publicize information on peer presence. For example, YouTube is known to highlight the number of new videos posted that day. In response to this trend, the website conducted a field experiment to answer two questions: first, whether and how disclosing the number of users on either side of the platform affects posting behaviors; and second, whether and how it affects total site traffic. To answer these questions, the website randomly varied whether to display the number of sellers and/or buyers to potential sellers, and

[5] The website's name and location is protected due to confidentiality agreements.

if so, how many sellers/buyers to claim.[6]

The website employed a between-subject design. Right after a potential seller has chosen the product category they intend to post in, and before they continue to the next webpage to fill out the posting form, they were exposed to an "information page". The text content displayed on the information page was randomly drawn from the following four treatment conditions:

1. "Presently, there are S postings and B users viewing these postings in the [category name] category of [city name]."
2. "Presently, there are S postings in the [category name] category of [city name]."
3. "Presently, there are B users viewing these postings in the [category name] category of [city name]."
4. (blank)

The number of postings S and the number of viewers B, if shown, were randomly drawn for each potential seller regardless of the product category. Individual-level randomization ensures that the correlation between entry and S or B does not pick up market-specific unobservable factors. Based on the actual long-run site traffic, both S and B were drawn from a uniform distribution between 5 and 200. By using the opaque wording "presently" to describe the time frame, management avoided deceiving customers by the randomization procedure.

Prior to the experiment ran, there was no information displayed about the number of buyers. Meanwhile, the formatting of the website made the number of sellers obscure. The categories we study are almost uniformly for sellers with a single unit of a good for sale. These sellers, due to the lack of prior experiences, are likely to face an unfamiliar market with uncertain demand

[6] Although the buyer-side entry decision is equally interesting to two-sided networks, management focused on potential sellers as experimental subjects. Since sellers need to register to post while buyers do not need registration to view the posts, the seller-side entry decision is a more deliberate choice task.

for every item they post. This should increase the data responsiveness to the experimental manipulation.

For field experiments conducted by an outside party, it is important to ensure that randomization was implemented correctly so that there is no systematic variation across conditions other than the experimental treatment. For example, it would be problematic if the website had displayed different conditions at different times of the day, as the behaviors of nighttime posters might be different from morning posters. We ran a series of regressions to ensure random assignment. These results are reported in full in the appendix.[7]

After being presented with the information page, a potential seller could either quit posting or proceed to the next page, fill in the posting form and complete the posting process. Once the seller had submitted the posting form, their item appeared on the website immediately. It is possible that a seller chooses the content of the posting form, such as the asking price, based on the experimental treatment they receive. However, due to confidentiality concerns, we were denied access to posting content data, and consequently we do not model post-entry competition among sellers. Instead we take the "envelope theorem" approach, assuming that entry decisions already take into account post-entry profit maximization. This allows us to focus on the effects of pre-entry perception of supply and demand.

3 Data

The field experiment ran from November 29, 2006 to January 15, 2007 within in the largest city market, which accounts for 16% of the total site traffic. During the period of the experiment, the other city markets showed no traffic change on either the seller or the buyer side, reassuring us that there were no nationwide market shocks which could have contaminated the experimental

[7]The only marginally significant correlation we found was that sellers in the "tickets" and "general" categories were more likely to see a higher number of sellers. Conversations with the firms about these categories lead us to believe that this is merely a statistical accident. For robustness, however, we repeated our empirical analyses with and without the "tickets" and "general" categories and obtained qualitatively similar results. We also report all results with errors clustered at the category level, to adjust for any within-category correlation.

results. During the experiment, the website received 9,722 new posting requests in the test city markets. Figure 1 shows the distribution of seller postings across product categories during the experiment. Rentals, Commercial Property, and Office were the most active categories.

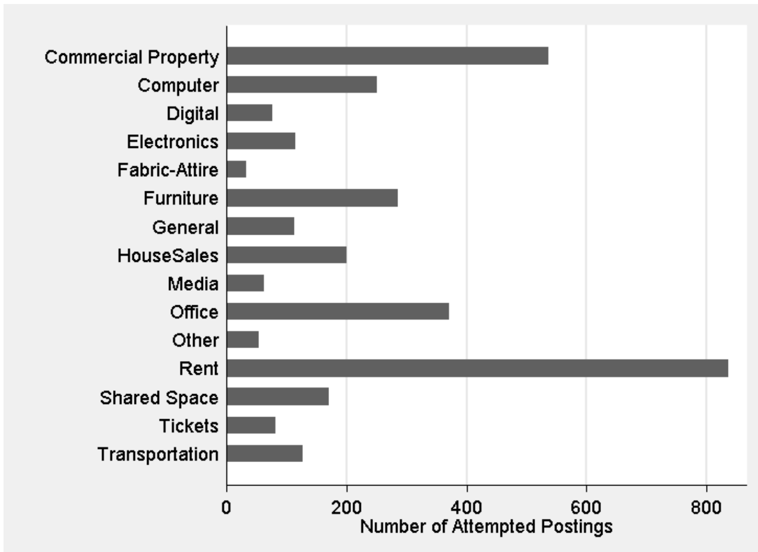

Figure 1: Distribution of Seller Postings across Categories

Two separate datasets were collected: a click-stream dataset, and a treatment dataset. First, using its Apache web server, the website captures the precise sequence of webpages requested by each user, identified by an IP address. Each entry in this click-stream data consists of a time stamp, the user's IP address, a record of all webpage requests, an error code, and the web browser used. This click-stream allows us to track whether a potential seller did actually make a posting, and the browsing sequence of a buyer. Second, during the experiment the website also complied a treatment dataset that recorded the "information page" each potential seller was exposed to. Each entry in the treatment data contains an IP address, a time-stamp, the product category the potential seller intended to post in, whether information on the number of buyers and/or sellers was displayed, and the actual number of buyers and/or sellers drawn if applicable. These treatment data spanned all potential sellers, including those who decided not to continue posting after receiving the treatment information. Out of all potential sellers we study, 88.5%

actually submitted a posting after receiving the treatment.[8]

A major challenge in interpreting the data is from the large number of repeat postings. The majority of repeat postings came from spammers, who employed automated posting tools that produce a large number of repeated posts. For example, one user (or bot) made 735 postings during the experiment, most of which were in the used computer equipment category. Since spammers would enter the market regardless of the information page content, including them would bias the estimates of entry responsiveness to the treatment. We therefore exclude spammers from the analyses. We defined a spammer as a seller who had submitted over 10 postings within the same category during the experiment, and removed 1,509 postings as a result. Other repeat postings were made by sellers who either accidentally posted twice in one day (for example, by refreshing the posting page or double-clicking the submit button), or deliberately posted their items in different categories. 83% of these repeat postings were made within the same category or in closely related categories (such as computers and electronics). Accidental repeat posts would inflate the statistical weight of the corresponding data points; while deliberate re-posters might have been exposed to contradictory information pages due to the full randomization protocol. Therefore, we retained data on the first posting, but removed subsequent postings from the same IP address on the same day.[9]

Among the remaining 3,315 potential sellers, 808 were given a blank information page, 872 only saw information about the number of buyers, 823 only saw information about the number of sellers, and 812 saw information about both buyers and sellers. Again, the data exclusion criteria are not significantly correlated with the experimental treatments.

[8]We match the treatment data with the browsing data using the IP address and the time stamp. We are unable to match 128 observations that contain errors, generally caused by time-outs or web-browser incompatibility. We exclude these 128 observations from our empirical analyses. There was no statistically significant relationship between our ability to match the data and the treatment condition.

[9]Since IP addresses do not uniquely identify users, we may have deleted observations where different users shared the same public computer. We report the empirical results including all potential sellers in the Appendix, which are qualitatively similar to the results excluding repeat postings.

4 Model and Results

4.1 How does the Number of Sellers and Buyers Affect Entry?

We want to assess how information on the number of buyers and sellers affects potential sellers' entry probabilities. Let N^{S*} denote the existing number of sellers in the market, N^{B*} the existing number of buyers, and $U^S(N^{S*}, N^{B*})$ a potential seller's utility from entering this market. If this potential seller observes N^{S*} but not N^{B*}, her utility from entry is affected by the number of sellers in the following way:[10]

$$\frac{dU^S(N^{S*}, N^{B*}(N^{S*}))}{dN^{S*}} = \frac{\partial U^S}{\partial N^{S*}} + \frac{\partial U^S}{\partial N^{B*}} \cdot \frac{\partial N^{B*}}{\partial N^{S*}} \tag{1}$$

We label the first component in the above equation, $\frac{\partial U^S}{\partial N^{S*}}$, the "congestion effect" of competition. We label the second component, $\frac{\partial U^S}{\partial N^{B*}} \cdot \frac{\partial N^{B*}}{\partial N^{S*}}$, the "inference effect" of competition, where a potential seller infers the number of buyers from the number of sellers. When $\frac{\partial U^B}{\partial N^S} \geq 0$ and $\frac{\partial U^B}{\partial N^B} \leq 0$, where U^B is a buyer's entry utility, the inference effect is always nonnegative. (Please see the Appendix for the proof.)

The model yields two empirical predictions. First, when the number of buyers is disclosed together with the number of sellers, removing the necessity for inference, a larger number of sellers hinders entry through a pure congestion effect. Second, when only the number of sellers is disclosed, its total effect on entry depends on how the congestion effect and inference effect play out, but should be more positive than the effect when numbers on both sides are disclosed.

To explore these predictions, we estimate a logit model where the dependent variable is whether a potential seller makes a posting after being exposed to the experimental manipulation. The independent variables include the number of sellers if shown, the number of buyers if shown,

[10] We do not examine the potential for forward-looking behavior or dynamics. See Dube, Hitsch, and Chintagunta (2008) for an example of research that explicitly incorporates dynamics into a two-sided network model.

and a series of category dummies, week dummies and day-of-week dummies. Table 1 reports the results.[11]

Table 1: Number of Sellers, Number of Buyers, and Entry Probabilities

	Only #Sellers Displayed	**#Sellers and #Buyers both Displayed**	**Only #Buyers Displayed**	**No Information Displayed**
#Sellers	-0.0003***	-0.0013***		
	(0.000)	(0.000)		
#Buyers		0.0022***	0.0001	
		(0.000)	(0.000)	
Constant	1.0119***	1.4161***	0.8013***	1.2657***
	(0.056)	(0.143)	(0.252)	(0.104)
Category Dummies	Yes	Yes	Yes	Yes
Time Dummies	Yes	Yes	Yes	Yes
Observations	808	768	857	783
Log-Likelihood	-318.45	-290.18	-325.86	-292.11
Pseudo-R2	0.10	0.13	0.11	0.17

Sample: Potential sellers considering posting
Dependent Variable: Indicator of whether a seller posts
* p<0.10, ** p<0.05, *** p<0.01
Probit Estimates. Standard errors clustered by category

The results are consistent with our predictions. First, when both the number of sellers and the number of buyers are displayed, entry declines with the number of sellers shown. That is, when there is no demand uncertainty, rival presence poses a traditional congestion effect. Second, seller presence has a less negative impact on entry when only the number of sellers is displayed than

[11] We report the results where a separate logit regression is run for each experimental condition. We have also pooled all data into one regression, identifying conditions with dummy variables. The pooled regression, allowed us to test whether the coefficient on the number of sellers differs across conditions. The estimation results of these two specifications are close. Also, for the results reported in Table 1 and Table 2, standard errors are clustered at the category (i.e., market) level to allow for unobservable category-specific common shocks. We have also estimated the model using either robust standard errors, or standard errors clustered by other potential sources of inter-group correlation (such as day of week). These different specifications of the error term lead to similar estimation results, as expected in a randomized field experiment.

when numbers on both sides are displayed ($-0.0003 > -0.0013$, $p = 0.000$). This is indicative of a positive inference effect that partially offsets congestion concerns.

The number of buyers, when shown in conjunction with the number of sellers, has a positive effect on entry. However, when only the number of buyers is displayed, its effect on entry seems neutral. The reason is that potential sellers may infer competition from demand, similarly to the way they infer demand from competition. In fact, we can derive a dual formula of equation 1 as $\frac{dU^S(N^{B*})}{dN^{B*}} = \frac{\partial U^S}{\partial N^{B*}} + \frac{\partial U^S}{\partial N^{S*}} \cdot \frac{\partial N^{S*}}{\partial N^{B*}}$, where the first component on the right-hand side represents a positive "surplus-extraction effect" of higher demand, whereas the second component represents a negative "competition inference effect". The intuition is familiar. A new textbook promoter, for instance, should be cautious in entering a large college market, as the readily observed high demand for textbooks might have attracted a number of veteran sellers.

4.2 How does Buyer Search Intensity Affect Entry?

Buyer search intensity can vary significantly across product categories. Possible factors that affect buyer search intensity are the nature of the goods (e.g., search product vs. experience product), product similarity within the category, and product substitutability. Therefore, we treat search intensity as inherent to the category, and measure it as the number of seller listings browsed by buyers divided by the number of all listings in that category. It is plausible that a seller would have a sense of relative search intensity in a category (for example, attire vendors would be aware that attire customers liked to browse broadly) without knowing the precise number of participants in this category. Figure 2 shows the cross-category variation in buyer search intensity. "Fabric-Attire" and "Other Categories" received the most intensive browsing per user.[12] Users browsed fewer computer postings, however, before terminating their search.

It is not obvious how buyer search intensity affects potential sellers' entry decisions in our setting. Superficially, intensive buyer search may aggravate congestion concerns as a firm may

[12] "Other Categories" contain disparate products such as medical equipment and beauty salon supplies.

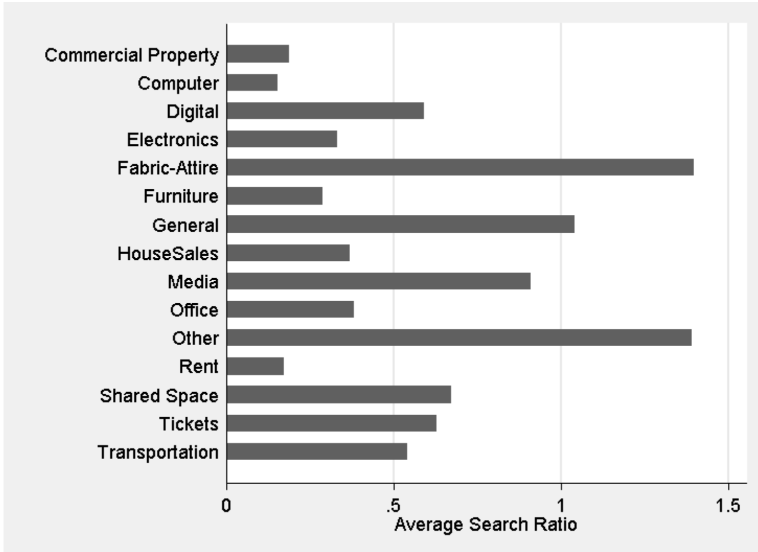

Figure 2: Distribution of Search Intensity across Categories

be deterred from posting if they know that customers are more likely to browse others' listings. However, customers who need to browse a lot will be attracted to two-sided networks with more sellers, making competition a stronger sign of demand. Therefore, greater buyer search intensity is expected to amplify the impact of demand uncertainty on entry.

We augment the previous specification by adding two interactive terms to the right-hand-side variables: $\#Sellers{*}SearchIntensity$, and $\#Buyers{*}SearchIntensity$. To avoid contaminating our inferences with endogenous buyer search, we use pre-experiment data to calibrate search intensity. The correlation in buyer search intensity between the pre-experiment period and the during-experiment period is 0.99. Figure 2 presents the distribution of search intensity across categories. Since we measure search intensity as a category-specific attribute, its main effect on entry is captured by the category dummies.

Table 2 reports the results. Consistent with our prediction that search intensity amplifies the impact of demand uncertainty, a larger number of sellers discourages entry in search-intensive categories when both demand and supply information is displayed, but increases entry likelihood in search-intensive categories when only the number of sellers is displayed. Meanwhile, a larger number of buyers encourages entry in search-intensity categories, but the effect is smaller when only the number of buyers is displayed than when the numbers on both sides are displayed. This can again be explained by the competition inference effect, where a potential seller infers dense rivalry when there are many buyers who prefer search.

Substantively, the results help identify search-intensive categories as the most susceptible to market size information. Methodologically, the results further confirm the central role of demand uncertainty in moderating the impact of competition on entry. In particular, besides demand uncertainty, the different amount of information across conditions could induce different levels of affect (e.g., a more informative page might make the website appear more professional), and different information evaluability (e.g., the information on the number of sellers is harder to evaluate when presented in isolation, and is therefore discounted. See for example Hsee (2000)). While affect difference is captured by the condition-specific constants, evaluability could be a competing explanation of demand uncertainty. However, the results in Table 2 helps rule out evaluability since it cannot explain the interactive effects of the number of sellers/buyers and search intensity.

5 Conclusion

This paper examines how information on the number of users affects user participation in two-sided networks. We identify demand uncertainty as a key moderator of competitiveness of a market impacts entry decisions. Specifically, we decompose the effect of competition into two components: a negative congestion effect that comes from post-entry competition, and a positive inference effect where a potential entrant deduces high market potential from heavy rival presence. Using field experiment data from a website that brings together buyers and sellers of used equipment and real estate opportunities, we are able to tease apart these two effects empirically. In particular, when the number of buyers is displayed together with the number of sellers, which renders inference unnecessary, a higher number of sellers reduces seller posting propensity. However, when the number of sellers is displayed in isolation, it has a significantly

Table 2: Search Intensity and Entry Probabilities

	Only #Sellers Displayed	**#Sellers and #Buyers both Displayed**	**Only #Buyers Displayed**	**No Information Displayed**
#Sellers	-0.0026***	0.0004		
	(0.0003)	(0.0004)		
#Sellers * Search Intensity	0.0073***	-0.0058***		
	(0.0001)	(0.0012)		
#Buyers		-0.0010***	-0.0004**	
		(0.0003)	(0.0002)	
#Buyers * Search Intensity		0.0109***	0.0015***	
		(0.0009)	(0.0001)	
Constant	-2.3945***	-2.5540***	0.6336**	-9.1481***
	(0.2248)	(0.1189)	(0.2715)	(0.3056)
Category Dummies	Yes	Yes	Yes	Yes
Time Dummies	Yes	Yes	Yes	Yes
Observations	808	767	857	784
Log-Likelihood	-316.06	-279.15	-325.59	-291.49
Pseudo-R2	0.11	0.16	0.12	0.18

Sample: Potential sellers considering posting
Dependent Variable: Indicator of whether a seller posts
* $p<0.10$, ** $p<0.05$, *** $p<0.01$
Probit Estimates. Standard errors clustered by category

less negative effect on posting decisions, indicating a positive inference effect. In a similar vein, potential sellers react more positively to high demand when competition density information is also provided than when high demand information is presented in isolation. Furthermore, we find that buyer search intensity amplifies the effect of demand uncertainty: a larger number of sellers discourages entry in search intensive categories when information on both sides of the market is displayed, but encourages entry in search intensive categories when only the number of sellers is displayed.

Our results suggest ways for two-sided networks to attract more traffic. The number of existing users can be a simple yet powerful network growth tool, but should be applied wisely. For example, networks should target the advertisement of high (low) seller concentration to potential sellers who know less (more) about demand. Similarly, if it is easily found out that sellers are few in volume, to attract more sellers the network should make demand information transparent too. Whether potentially sellers know about demand is especially important when buyers tend to engage in intensive search. Our framework may also help reconcile the opposing findings in the literature on how competition affects entry. The results suggest that the direction and magnitude of the total effect of competition crucially depends on what market information is available to potential entrants. We contribute to the entry literature by experimentally identifying market information availability as a driver of entry decisions.

One direction of future research is to explicitly integrate the impact of information on both sides of a two-sided network. While this current research examines how competitor information affects seller behavior in a two-sided network and Tucker and Zhang (2008) examine how popularity information affects buyer choices, there has been no work that investigates the effect of information on both sides of the market simultaneously. This direction of research is important in understanding the positive feedback mechanism between the two sides that drives network growth. Future research could also investigate other strategic variables such as post-entry price (as discussed in (Chen, Iyer, and Padmanabhan 2002)), which were not available to us in this study. Another possibility is to incorporate the intricacies of the meanings conferred by numbers. For example, it has been found that having more options may lead to fewer choices (e.g., Iyengar and Lepper (2000), Kuksov and Villas-Boas (2005)). It would be interesting to explore how buyers' mixed reaction towards the number of sellers modifies the inference effect of competition.

References

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

Banerjee, A. V. (1992, August). A simple model of herd behavior. *The Quarterly Journal of Economics 107*(3), 797–817.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *The Journal of Political Economy 60*(4), 889–917.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural changes as information cascades. *The Journal of Political Economy 100*(5), 992–1026.

Bradlow, E. T. and Y.-H. Park (2007). Bayesian estimation of bid sequences in internet auctions using a generalized record-breaking model. *Marketing Science 26*(2), 218–229.

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

Camerer, C. F., T.-H. Ho, and J.-K. Chong (2004, August). A cognitive hierarchy model of games. *The Quarterly Journal of Economics 119*(3), 861–898.

Camerer, C. F. and D. Lovallo (1999, mar). Overconfidence and excess entry: An experimental approach. *The American Economic Review 89*(1), 306–318.

Chen, Y., G. Iyer, and V. Padmanabhan (2002). Referral infomediaries. *Marketing Science 21*(4), 412–434.

Chen, Y. and J. Xie (2007). Cross-market network effect with asymmetric customer loyalty: Implications for competitive advantage. *Marketing Science 26*(1), 52–66.

Debruyne, M. and D. J. Reibstein (2005). Competitor see, competitor do: Incumbent entry in new market niches. *Marketing Science 24*(1), 55–66.

Dube, J.-P. H., G. J. Hitsch, and P. K. Chintagunta (2008). Tipping and Concentration in Markets with Indirect Network Effects. Mimeo, Chicago GSB.

Dudey, M. (1990). Competition by choice: The effect of consumer search on firm location decisions. *The American Economic Review 80*(5), 1092–1104.

Fath, G. and M. Sarvary (2003). Adoption dynamics in buyer-side exchanges. *Quantitative Marketing and Economics 1*, 305335.

Gould, E. D., B. P. Pashigian, and C. J. Prendergast (2005). Contracts, externalities, and incentives in shopping malls. *The Review of Economics and Statistics 87*(3), 441–422.

Hitsch, G. J. (2006). An empirical model of optimal dynamic product launch and exit under demand uncertainty. *Marketing Science 25*(1), 25–50.

Hsee, C. K. (2000). *Choices, Values and Frames*, Chapter Attribute evaluability and its implications for joint-separate evaluation reversals and beyond, pp. 543–563. Cambridge, England: Cambridge University Press.

Iyengar, S. S. and M. R. Lepper (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology 79*(6), 995–1006.

Kuksov, D. and J. M. Villas-Boas (2005). When more alternatives lead to less choice. Working paper, Washington University, St. Louis, University of California, Berkeley.

Narasimhan, C. and Z. J. Zhang (2000). Market entry strategy under firm heterogeneity and asymmetric payoffs. *Marketing Science 19*(4), 313–327.

Orhun, Y. (2007). Spatial differentiation in the supermarket industry. Mimeo, Chicago GSB.

Rochet, J.-C. and J. Tirole (2006). Two-sided markets : A progress report. *Rand Journal of Economics 37*(3), 645–667.

Salop, S. C. (1979). Monopolistic competition with outside goods. *The Bell Journal of Economics 10*(1), 141–156.

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

Simonsohn, U. (2006). Excess Entry into High-Demand Markets: Evidence from the Timing of Online Auctions. Mimeo, Wharton.

Toivanen, O. and M. Waterson (2005). Market structure and entry: where's the beef? *RAND Journal of Economics 36*(3), 680–699.

Tucker, C. (2008, January). Two-sided network effects in the adoption of electronic payments.

Tucker, C. and J. Zhang (2008). How does popularity information affect choices? theory and a field experiment. MIT Sloan Research Paper No. 4655-07.

Urban, G. L. and J. R. Hauser (1993, May). *Design and Marketing Of New Products*. Prentice Hall; 2nd edition.

Vitorino, M. A. (2007). Empirical entry games with complementarities: An application to the shopping center industry. Mimeo, Chicago GSB.

Wernerfelt, B. (1995, March). A rational reconstruction of the compromise effect: Using market data to infer utilities. *Journal of Consumer Research: An Interdisciplinary Quarterly 21*(4), 627–33.

Zhu, T., V. Singh, and A. J. Dukes (2005, May). Local competition and impact of entry by a dominant retailer. CIE Discussion Papers 2005-05.

6 Appendix

6.1 Modeling the Demand Inference Effect of Competition

In this section, we build an analytical model to show how potential entrants can infer demand from competition. To stay general, we abstract where possible from parameterizing the firm objective function and focus on directional conclusions. However, once we specify a functional form for a given market, our model yields point predictions of the magnitude of the inference effects.

Let there be two groups of traders on the market: buyers and sellers. For example, on websites such as craigslist.com, the buyers are the viewers of the posts, and the sellers are the posters. Let N^i denote the number of traders on side i, where $i \in \{B, S\}$ stands for buyers or sellers. The utility for a trader on side i to enter the market is:

$$U^i = U^i(N^i, N^j) - c^i \tag{2}$$

where $j \in \{B, S\}, j \neq i$ denotes the other side of the market. The functional form of $U^i(\ )$ is common knowledge. Without loss of generality, we assume that

$$\frac{\partial\ U^i}{\partial\ N^j} \geq 0, \quad \frac{\partial\ U^i}{\partial\ N^i} \leq 0 \tag{3}$$

The above utility specification captures the dynamics of most markets, where a trader (weakly) benefits from an increased number of traders on the other side of the market, and is (weakly) hurt by a larger number of traders on its own side. For example, compared to a monopolistic market, a market with more firms dissipates firm profits and increases consumer surplus. The literature on two-sided platforms has focused on positive feedback mechanisms in markets such as video-games, and has therefore assumed away congestion effects (e.g., Rochet and Tirole2006, Armstrong 2006). Our model nests the classic specification of two-sided network utilities that does not consider congestion effects (i.e., $\frac{\partial U^i}{\partial N^i} = 0$), where a trader's gain from participation is written as $U^i = a^i \cdot N^j - c^i$, where $a^i > 0$.

Suppose a trader incurs a fixed cost in order to enter the market. Let c^i denote such a cost for a trader on side i. We allow traders to be heterogeneous with respect to their entry costs. Let c^i be randomly distributed across side-i traders following a cumulative distribution function $F^i(\cdot)$, which is common knowledge. In other words, although a trader does not directly observe the entry cost of a particular competitor, she knows the distribution of entry costs across all traders. Last, let M^i denote the total number of *potential* traders on side i. The value of M^i is exogenous to the model. Among these M^i potential traders, those with $U^i(N^i, N^j) \geq c^i$ will choose to enter the market. While the potential market size M^i is exogenous, in equilibrium the actual number of entrants on both sides of the market N^i is endogenously determined in the following way:

$$\begin{aligned} N^{B*} &= M^B \cdot F^B(U^B(N^{B*}, N^{S*})) \\ N^{S*} &= M^S \cdot F^S(U^S(N^{S*}, N^{B*})) \end{aligned} \tag{4}$$

From this simultaneous equation system, we can derive the equilibrium number of traders on both sides of the market, once we know the functional form of the utilities and of the entry cost distribution. For example, if the trade utility is $U^i = \frac{N^j}{N^i} - c^i$ for a two-sided network that allows congestion within the same side, and if entry costs on side i are uniformly distributed over $[0, \bar{c}^i]$, it can be shown that in equilibrium $N^{i*} = \sqrt[3]{\frac{M^{i2} M^j}{\bar{c}^{i2} \bar{c}^j}}$. Note that the number of traders on one side of the market increases in the market potential (M^j) and decreases in the entry costs on the other side.

Now suppose that the market has evolved to an equilibrium, and that another agent (a seller without loss of generality) has newly arrived at the market and is contemplating entry. Her entry decision is straightforward if she observes both the equilibrium number of buyers and the equilibrium number of sellers, which is equivalent to knowing M^B and M^S. In a more interesting case, assume that this seller knows the equilibrium number of sellers N^{S*} but does not know N^{B*}, and has no information on M^B or M^S. This potential seller then bases her actions on the knowledge that in equilibrium the number of buyers is related to the number of sellers through the function $N^{B*}(N^{S*})$. Below we derive how N^{B*} changes with N^{S*}.

Let $\phi = N^{B*} - M^B \cdot F^B(U^B(N^{B*}, N^{S*})) = 0$. We know $\frac{\partial \phi}{\partial N^{S*}} = -M^B \cdot f^B(U^B(N^{B*}, N^{S*})) \cdot \frac{\partial U^B}{\partial N^{S*}}$, where $f(\cdot) \geq 0$ is the density function of entry cost c^B. Since $\frac{\partial U^B}{\partial N^{S*}} \geq 0$, $\frac{\partial \phi}{\partial N^{S*}} \leq 0$. Similarly, $\frac{\partial \phi}{\partial N^{B*}} = 1 - M^B \cdot f^B(U^B(N^{B*}, N^{S*})) \cdot \frac{\partial U^B}{\partial N^{B*}} \geq 0$. By the Implicit Function Theorem, $\frac{\partial N^{B*}}{\partial N^{S*}} = -\frac{\partial \phi}{\partial N^{S*}} / \frac{\partial \phi}{\partial N^{B*}} \geq 0$. That is, given the rather mild assumption stated in Equation 3, a potential entrant can infer a (weakly) larger number of buyers from a larger number of sellers.

6.2 Censoring

The empirical analyses in the main text have excluded postings which contain problems discussed in the Data section. Table 3 and Table 4 report the results where all potential sellers are included. The two sets of results are qualitatively similar.

6.3 Check Randomization of Experimental Manipulation

We ran a series of regressions to ensure that the website had correctly implemented randomization. Table 5 reports the regression results. The first three columns investigate whether the assignment into the four treatment conditions was correlated with category, time of posting, or day of the week. We also regressed the numbers of sellers and buyers displayed on category, time of posting, and day of the week. These results are reported in the last two columns of table 5. The only marginally significant correlation we found was that sellers in the "tickets" and "general" categories were more likely to see a higher number of sellers. Conversations with the firms about these categories lead us to believe that this is merely a statistical accident. None,

Table 3: Number of Sellers, Number of Buyers, and Entry Probabilities

	Only #Sellers Displayed	**#Sellers and #Buyers both Displayed**	**Only #Buyers Displayed**	**No Information Displayed**
#Sellers	-0.0006***	-0.0010***		
	(0.000)	(0.000)		
#Buyers		0.0023***	0.0001	
		(0.001)	(0.000)	
Category Dummies	Yes	Yes	Yes	Yes
Time Dummies	Yes	Yes	Yes	Yes
Observations	792	751	834	759
Log-Likelihood	-318.83	-285.70	-315.48	-286.79
Pseudo-R2	0.10	0.13	0.12	0.16

Sample: All potential sellers considering posting
Dependent Variable: Indicator of whether a seller posts
* $p<0.10$, ** $p<0.05$, *** $p<0.01$
Standard errors clustered by category

the less we have clustered standard errors at the category level and also include category specific dummies in our specifications as a precaution.

Table 4: Search Intensity and Entry Probabilities

	Only #Sellers Displayed	**#Sellers and #Buyers both Displayed**	**Only #Buyers Displayed**	**No Information Displayed**
#Sellers	-0.0028***	0.0007***		
	(0.0003)	(0.0001)		
#Sellers * Search Intensity	0.0071***	-0.0060***		
	(0.0001)	(0.0009)		
#Buyers		-0.0009***	-0.0008***	
		(0.0002)	(0.0000)	
#Buyers * Search Intensity		0.0103***	0.0023***	
		(0.0007)	(0.0004)	
Category Dummies	Yes	Yes	Yes	Yes
Time Dummies	Yes	Yes	Yes	Yes
Observations	792	751	834	759
Log-Likelihood	-314.79	-276.63	-315.06	-285.31
Pseudo-R2	0.11	0.15	0.12	0.16

Sample: All potential sellers considering posting
Dependent Variable: Indicator of whether a seller posts
* $p<0.10$, ** $p<0.05$, *** $p<0.01$
Probit Estimates. Standard errors clustered by category

Table 5: Empirical Check of Randomization

	Multinomial Logit Regression			Linear Regression	
Dependent Variable:	Sellers & Buyers Displayed	Only Buyers Displayed	Only Sellers Displayed	#Sellers Displayed	#Buyers Displayed
Day of Week	-0.0443	0.0392	-0.0987	7.0978***	-1.6292
	(0.0946)	(0.0946)	(0.0925)	(2.6564)	(2.6947)
Day of Week Sq	0.0129	-0.0074	0.0228	-1.1088***	0.1024
	(0.0151)	(0.0153)	(0.0148)	(0.4258)	(0.4265)
Time	0.6734	2.0303*	0.8570	28.5447	-14.4693
	(1.0515)	(1.1149)	(1.0410)	(31.2777)	(30.2691)
Time Sq	-0.6146	-1.6084*	-0.5837	-23.2414	21.5615
	(0.8802)	(0.9212)	(0.8657)	(25.9803)	(25.2270)
Computer	0.1850	0.1602	0.3580	-4.0974	1.8440
	(0.2337)	(0.2227)	(0.2183)	(6.2501)	(6.1470)
Digital	-0.1924	-0.0793	-0.0456	3.7913	-10.5662
	(0.3649)	(0.3352)	(0.3357)	(10.1564)	(10.2675)
Electronics	-0.0188	-0.1091	-0.0636	5.4123	-3.2366
	(0.2990)	(0.2873)	(0.2878)	(8.3947)	(8.5046)
Fabric-Attire	0.2120	0.2889	-0.2731	6.9713	1.2659
	(0.5336)	(0.4987)	(0.5702)	(13.5575)	(15.8440)
Furniture	0.1126	-0.3324	0.0996	-0.4064	-1.0208
	(0.2113)	(0.2149)	(0.2022)	(6.1287)	(5.8282)
General	0.1297	-0.1000	-0.4369	-10.0497	14.1735
	(0.2868)	(0.2838)	(0.3101)	(8.0605)	(8.7652)
HouseSales	0.3401	-0.2158	-0.0819	-2.9314	-3.9186
	(0.2340)	(0.2424)	(0.2402)	(6.5974)	(6.6142)
Media	-0.2423	-0.5263	-0.0728	14.1885	1.5678
	(0.3864)	(0.3924)	(0.3517)	(11.9638)	(10.8358)
Office	-0.0769	-0.2863	-0.1516	0.3984	1.8786
	(0.1965)	(0.1908)	(0.1897)	(5.6045)	(5.6232)
Other	0.1645	-0.5130	0.1730	1.6370	4.9431
	(0.4061)	(0.4505)	(0.3886)	(12.4067)	(10.9627)
Rent	0.1312	0.0578	00.0678	2.7513	3.2216
	(0.1627)	(0.1564)	(0.1582)	(4.4921)	(4.5924)
Shared Space	0.2433	-0.3006	-0.0535	-15.9461**	-3.4092
	(0.2463)	(0.2574)	(0.2480)	(7.1188)	(6.9704)
Tickets	-0.0924	-0.6141*	0.0136	10.9550	18.3150**
	(0.3359)	(0.3604)	(0.3118)	(10.5577)	(9.3267)
Transportation	0.334	-0.112	-0.0274	4.8382	-10.4943
	(0.2794)	(0.2874)	(0.2861)	(7.8064)	(7.8730)
Constant	-0.2783	-0.4543	-0.1736	89.8177***	105.9281***
	(0.3427)	(0.3597)	(0.3387)	(10.0699)	(9.9069)
Observations	3314	3314	3314	1634	1686
Log-Likelihood	-4566	-4566	-4566	-8917	-9218

Sample: Customers included in the field experiment
* $p<0.10$, ** $p<0.05$, *** $p<0.01$