



MIT Sloan School of Management

MIT Sloan School Working Paper 4678-08
11/18/2007

Decomposing the Congestion Effect and the Inference
Effect of Competition: A Field Experiment

Catherine Tucker, Juanjuan Zhang

© 2007 Catherine Tucker, Juanjuan Zhang

All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission, provided that full credit including © notice is given to the source.

This paper also can be downloaded without charge from the
Social Science Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=1084995>

Decomposing the Congestion Effect and the Inference Effect of Competition: A Field Experiment

November 18, 2007

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 of Competition: A Field Experiment

Abstract

Are firms more or less likely to enter a market if they observe that competitors have entered? This most basic question has received contradictory empirical answers. The normative recommendation to firms that make entry decisions is therefore ambiguous. We reconcile this controversy by introducing demand uncertainty as a moderator of how entrants respond to existing competition. We distinguish between two effects of competition on entry decisions: a negative “congestion effect,” where competition dissipates profit when demand is fixed and is known, and a positive “inference effect,” where firms infer high demand from a large number of competitors. To tease apart these two effects empirically, we use field experiment data from a website that brings together buyers and sellers of used goods. Before each potential 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 evidence for a positive inference effect: When the number of buyers is not disclosed, the overall effect of the number of sellers on entry is neutral; when the number of buyer is disclosed, however, a larger number of sellers lowers the entry propensity due to the congestion effect. We discuss how our results should affect the information disclosure strategies of two-sided platforms.

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

JEL Classifications: C93, D83, L11, M31

1 Introduction

Are firms more likely or less likely to enter a market if they observe that competitors have entered? Research in both marketing and industrial organization emphasizes that competition deters entry, both theoretically and empirically. However, this received wisdom has been questioned in a recent behavioral economics literature that finds “competition neglect” (e.g., Simonsohn (2006)), and a marketing literature that documents “competition contagion,” where firms are indifferent to or even more likely to enter heavily congested markets (e.g., Debruyne and Reibstein (2005)). Understanding entry decisions is a key part of successful market forecasting, so we aim to reconcile these contradictory findings.

Our key insight is that the nature of spillovers between entry decisions is crucially affected by firms’ uncertainty about market desirability. We distinguish between two types of spillovers that are governed by the availability of demand information: A “congestion effect,” where potential entrants avoid competitive and consequently unprofitable markets, and an “inference effect” where potential entrants infer the desirability of entering a market from existing competition. For example, there are 4,023 nail salons in New York.¹ The number could indicate that the local manicure market is saturated and unprofitable (a congestion effect); or that New Yorkers have a high demand for nail grooming (an inference effect). How someone contemplating opening a new nail salon reacts to the number may therefore depend on how well-informed she is about local demand and on whether she would use density of competition as an indicator of the level of local demand.

We develop a theoretical framework that incorporates both effects of competition and prescribes how market uncertainty affects the entry decisions of buyers and sellers who transact over a common platform. Our framework generalizes the classic platform model of

¹NAILS Magazine 2005-2006 Big Book, Regional Statistics.

Rochet and Tirole (2006) to emphasize the crucial role played by market uncertainty. Some recent research uses bounded rationality to explain a lack of negative correlation in entry decisions, but our model provides a starting point to interpret all possible scenarios of entry correlation - negative, neutral, or positive - within a rational framework.

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 direct data on the degree of uncertainty and on the variation of uncertainty across time or market. As a result, until now the two effects of competition have not been disentangled, and this has led to potentially biased academic and managerial interpretations. In particular, the two opposite forces might cancel out in aggregate, causing underestimation of either effect. However, historical data do not allow researchers to separate out these two effects, as either there is no information about whether firms are informed about demand, or such information acquisition is endogenous.

We circumvent these problems by using data from a field experiment where potential entrants were randomly informed about demand. The field experiment was conducted by a website that brings together sellers and buyers of used goods in the same metropolitan area. After incurring a fixed registration cost, a seller can post an item for sale on the website. Before each potential seller made the posting decision, 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 presented, a larger number of sellers reduces a seller's posting propensity. However, when information about the number of sellers is presented in isolation, it has a small and statistically insignificant effect on the seller's posting decision. The results support our contention that the effect of competition on entry can be either negative or positive, and that whether it is negative or positive depends on how much potential entrants rely on competition to resolve demand uncertainty. In our particular experimental setting, when

firms do not have access to demand information, the inference effect offsets the congestion effect, leaving the total effect of competition on entry neutral.

In the rest of the section, we first review previous research on how competition affects entry and discuss the contribution of this study. We then discuss the managerial implications of our findings.

1.1 Previous Literature on Entry

Past studies on competition and entry have found that competition discourages entry, is neglected by potential entrants, or induces entry. We discuss these three categories of results in turn.

Classic economic models of firm behavior with homogenous goods, such as Bertrand, suggest that competition tends to discourage further entry. Similarly, in Salop (1979)'s model of differentiated markets, the equilibrium profit margin and market share both decline with the number of firms in operation, and entry ceases when the equilibrium profit of entering firms decreases to zero. Empirical evidence also abounds that industry profit declines with the number of firms (e.g., Bresnahan and Reiss (1991), Berry (1992)). The most recent research 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)). There are also a few studies that have taken the opposite approach and estimated 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 model and interpret entry correlations in light of such assumptions. It is very unusual, however, for researchers to directly observe the exact information structure of the potential entrants. Therefore, the validity of the informational

assumptions key to this stream of research is hard to test, and it is therefore also hard to draw conclusions on the relationship between competition and entry. Even if researchers have precise information on the state of firm's knowledge about market conditions, there is a deeper concern that that information acquisition itself may be an endogenous variable (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. Through a field experiment approach, we are able to address both questions by exogenously controlling the level of market uncertainty and tracing the causal effect of information. Our result is the first explicit empirical evidence that demand uncertainty moderates the impact of competition on entry.

This traditional body of work on how competition deters entry has been challenged by a behavioral literature which argues that there is "competition neglect" due to bounded rationality. The foundation of this research is a series of lab experiments that test theoretical Nash-entry games. All the participants in these experiments are fully informed about capacity (or demand) and consequent payoffs. The early literature has documented that even in large-scale games, "as if by magic," participants' entry decisions correspond well to a standard Nash entry equilibrium (e.g., Kahneman (1988) and Rapoport, Seale, Erev, and Sundali (1998)). The only exception is the work of Camerer and Lovo (1999), who find that when the entry decision depends on skills potential entrants are likely to exhibit "reference group neglect", failing to recognize that other entrants will be as skilled as they are.² Though lab evidence of excess entry has been confined in these games involving participants' skills, the potential of excess entry and the work by Camerer, Ho, and Chong (2004)

²It does not require substantial skills for firms to enter the market in our field experimental setting, which helps to minimize the impact of reference group neglect. Nevertheless, through a full between-subject experimental design, we are able to randomize out any reference group neglect effect and isolate the treatment effects.

on limited iterative thinking capacity has inspired many empirical researchers to document excess entry using historical data.³ One example is Simonsohn (2006) who finds that more eBay auctions end on Sunday afternoons than are warranted by high demand, and suggests that sellers overemphasize high demand and underemphasize competition. However, without precise information on whether sellers are indeed fully informed about the demand level on Sunday afternoons, competition neglect may not be the only way to explain positively correlated entry decisions. For this same reason, it is difficult to replicate in the field the laboratory findings of competition neglect where full information is imposed on participants.

Our paper contributes to the understanding of competition neglect in two ways. First, the seeming lack of correlation between competition and subsequent entry could be an artifact of the congestion and learning effects of competition offsetting each other. In other words, it could be the case that potential entrants do take competition into serious consideration, but that the total effect of competition is neutralized. Indeed, when there is full information about demand we find scant evidence of competition neglect, consistent with the earlier lab findings of Kahneman (1988) and Rapoport, Seale, Erev, and Sundali (1998). Second, while many studies use bounded rationality to explain over-entry and high incidence of post-entry failure,⁴ our conceptualization of competition provides a starting point 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. (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.)

³Sometimes Einav (2007)'s work on the timing of releases of blockbuster films is cited as an example of "excess entry" in the field. However, the seasonality in entry can be rationalized in equilibrium by the rigidities of cinema prices.

⁴Urban and Hauser (1993) report that across industries an average of 35% of new products fail after launch.

In this sense, our approach echoes the work by Wernerfelt (1995) who reinterprets the compromise effect, another well-known bounded-rationality phenomenon, as rational behavior governed by limited information.

Last, a sociological literature has examined the possibility of mimetic adoption, and has documented positive correlations in entry decisions or “competitive contagion” (e.g., Haveman (1993), Greve (1998)). The work of Debruyne and Reibstein (2005) introduces the concept of competitive contagion to a marketing setting. They document positive correlations in entry in the online brokerage industry. It is not a main focus of this literature to test between different behavioral mechanisms that lie behind competitive contagion. For example, contagious entry can either be a result of resource-based competition (Narasimhan and Zhang (2000)) or because the competitor’s actions are believed to contain information about the desirability of the market. More generally, any measured effect could be driven by firm inference of demand from competition, or by market-specific time-variant unobservables that impact both the incumbents and the entrants. We contribute to this literature by using randomized exogenous variation in information availability to pinpoint the precise inference mechanism at work.

In sum, our study enriches the entry literature by isolating two opposite effects of competition. By doing so, we provide a unifying framework that reconciles the mixed findings on the relation between competition and entry.

1.2 Managerial Contribution

Our results not only address an ongoing academic debate, but also offer crucial insights for managers. In particular, our results highlight the importance of strategic information disclosure for firms whose business model depends on bringing together buyers and sellers

and profiting from their interaction. Though there is a growing theoretical literature on two-sided networks (see Armstrong (2006) and Rochet and Tirole (2006) for an overview), there is less work that sheds light on managerial strategies for expanding these businesses.⁵ So far empirical research along this path has been limited, due to the problems of ascribing causality to endogenous firm strategy using historical data. We circumvent this endogeneity problem by using field experiment data where platform strategies were explicitly randomized, and by empirically evaluating different information provision strategies. Our results suggest that in almost all cases, providing information enhances the likelihood of sellers joining the platform. In the long run, we find that the best strategy for retaining sellers is to highlight the coexistence of a high number of buyers and a low number of sellers. We also find evidence that while advertising a low number of sellers attracts sellers, advertising a high number of sellers does not have a significant negative effect.

The rest of the paper is organized as follows. Section 2 introduces a general framework with an analytical representation of the congestion effect and the inference effect. Section 3 describes the field experiment and Section 4 presents the data. Section 5 first discusses the main results where sellers react differently to information about the level of competition, depending on whether it is presented in isolation or in combination with demand information. We then present an extension, where we explore the optimal information disclosure strategies for platforms interested in attracting entry. Section 6 summarizes the paper and discusses potential directions for future research.

⁵Exceptions are Chen and Xie (2007), who theoretically study the role of cross-market asymmetry in strategies to build customer loyalty, and Tucker and Zhang (2007), who find that publicizing vendor popularity information increases visits to an internet based listing service

2 Theoretical Model

In this section, we build a general model to analyze how information availability changes the effect of competition on entry. Towards this goal, we decompose the overall effect of competition on entry into a negative congestion component and a positive inference component. 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 two competition 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:

$$U^i = U^i(N^i, N^j) - c^i \tag{1}$$

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

$$\frac{\partial U^i}{\partial N^j} \geq 0, \quad \frac{\partial U^i}{\partial N^i} \leq 0 \tag{2}$$

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 Tirole (2006), 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:

$$\begin{aligned} N^{B*} &= M^B \cdot F^B(U^B(N^{B*}, N^{S*})) \\ N^{S*} &= M^S \cdot F^S(U^S(N^{S*}, N^{B*})) \end{aligned} \tag{3}$$

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.

The above aggregate market equilibrium is reached as a result of individual traders making equilibrium entry decisions after observing their private draw of entry costs. Individual traders' entry decisions are straightforward with complete demand information. For example, if a seller knows both the equilibrium number of buyers and the equilibrium number of sellers (this is equivalent to knowing M^B and M^S), it will enter the market if and only if $U^S(N^{S*}, N^{B*}) \geq c^S$. Since $\frac{\partial U^S}{\partial N^{S*}} \leq 0$, this seller's possibility of entry weakly decreases in the number of sellers on the market due to pure "congestion" concerns. These results change if the same 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 $N^{B*}(N^{S*})$. Her utility from entry is therefore

$$U^S(N^{S*}) = U^S(N^{S*}, N^{B*}(N^{S*})) - c^S \quad (4)$$

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, $\frac{\partial N^{B*}}{\partial N^{S*}} = -\frac{\partial \phi}{\partial N^{S*}} / \frac{\partial \phi}{\partial N^{B*}} \geq 0$. Finally, by the chain rule,

$$\frac{dU^S(N^{S*})}{dN^{S*}} = \frac{\partial U^S}{\partial N^{S*}} + \frac{\partial U^S}{\partial N^{B*}} \cdot \frac{\partial N^{B*}}{\partial N^{S*}} \quad (5)$$

We call the first component in equation 5, $\frac{\partial U^S}{\partial N^{S*}} \leq 0$, the "congestion effect" of competition. We call the second component, $\frac{\partial U^S}{\partial N^{B*}} \cdot \frac{\partial N^{B*}}{\partial N^{S*}} \geq 0$, the "inference effect" of competition. The implication of the degree of competition to a potential entrant is thus two-fold: although having more rival sellers hurts the potential seller's market power, the existence of intense competition is a sign of great market potential. Whether a potential entrant should respond positively to competition depends on how the congestion effect and the inference effect play

out against each other given the industry structure and the set of environmental parameters. In the rest of the paper, we use data from a field experiment to tease apart the two competition effects empirically.

3 Business Context and Field Experiment

We obtained our field experiment data from a website similar to craigslist.org.⁶ The website provides a common platform for sellers and buyers of used goods to advertise these goods and to read the advertisements. Figure 1 presents the span and size of product categories. More than 40 major metropolitan areas are served. Because the goods are immobile (being computers, cars, couches, etc.), 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 a monetary amount 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 to be able to post 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 registering at the website.⁷ Buyers can view postings without signing up for the website. Other things being equal, the return on posting for a seller would presumably increase with the number of buyers, due to a higher chance of match and greater bargaining power.

There has been a trend for merchants and networks to publicize popularity information

⁶We cannot disclose the name of the website due to confidentiality agreements.

⁷After a posting is made, it is listed chronologically on the website. This listing format minimizes resource-based competition that has strategic effects on entry (Narasimhan and Zhang (2000))

in such formats as “518 people have seen this movie review.” In response to this widespread practice, the website conducted a field experiment in order to answer two questions. Management wanted to find out 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, and if so, how many sellers/buyers to claim.

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 have been S postings in the [category name] category of [city name].”
3. “Presently, there have been 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. Individual-level randomization ensures that the correlation between entry and S or B does not pick up market-specific unobservable factors, and that the only systematic variation comes from the experimental treatment. 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 ensured that they did not deceive their customers by the randomization procedure.

Before the experiment ran, there was no information displayed about the number of buyers. In addition, the formatting of the website made the number of sellers obscure. Our empirical analysis focuses on sellers with a single unit of a good for sale. The advantage of focusing on single-unit sellers is that such sellers, due to the lack of prior experiences, are likely to face an unfamiliar market with uncertain demand for every item they post. This should increase 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 the 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. Table 1 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 1. 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.

After being presented with the information page, a potential seller could choose either

to quit posting or to 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 not given access to posting content data, and consequently we do not model post-entry competition among sellers. Instead we remain focused on the entry decision based on pre-entry perception about supply and demand.

4 Data

The field experiment ran from November 29, 2006 to January 15, 2007. Although the website operates in over 40 metropolitan areas, the experiment was only conducted 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 results. During the length of the experiment, the website received 4,152 new posting requests in the city market where the test was conducted. Figure 1 shows the distribution of seller postings across product categories during the time of the experiment. Computer, Furniture, and Automobile were the most active categories.

Two separate datasets were collected: a click-stream dataset, and a treatment dataset. 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 that the surfer used to make the webpage request. This click-stream allows us to track whether a potential seller did actually make a posting. During the experiment the

website also collected additional treatment data that recorded the “information page” each potential seller was exposed to. Specifically, 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 such information was displayed. These treatment data spanned all potential sellers, including those who decided not to continue posting after receiving the treatment information. Out of the potential sellers we study, 88 percent 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 time of the experiment, most of which were in the used computer equipment category. Since spammers would enter the market regardless of the information page content, the inclusion of spammers in the empirical analyses would give biased estimates of the sensitivity of entry decisions to the treatment. We remove spammers from the data. We defined a spammer as a seller who has submitted over 10 postings in the course of the experiment and removed 2,799 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.

⁸We 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.

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 832 potential sellers, 197 were given a blank information page, 227 only saw information about the number of buyers, 199 only saw information about the number of sellers, and 209 saw information about both buyers and sellers. Again, the data exclusion criteria are not significantly correlated with the experimental treatments.

5 Results

The field experiment was designed to test two things. First, it was designed to test how the nature of information presented to potential sellers affects their entry decisions. Second, and more broadly, it was designed to test whether a website should reveal information about how many users it has. In other words, we first look at the effect of the within-treatment variation in information, and then look at the effect of between-treatment variations in Section 5.2.

5.1 How does the nature of information affect entry decisions?

We want to assess how information about the number of buyers and sellers moderates the impact of competition on entry probabilities. Following equation 5, we run a regression of the posting/entry decisions on the number of buyers and/or sellers and uncover the sign of the partial derivatives. We use a logit specification for the binary entry decision. The dependent variable is an indicator variable that equals 1 if the potential seller makes a post, and 0 otherwise. The independent variables include a constant term, the number of sellers if

⁹Since IP addresses do not uniquely identify users, we may also delete observations where different users used the same public computer.

shown, and the number of buyers if shown. We stratify our results to compare the affect of information in each of the four possible information conditions. Table 2 reports the results.

5.1.1 The Effect of the Number of Sellers

When only the number of sellers is displayed, the actual number of sellers shown has only an insignificant effect on the likelihood of posting. This result is consistent with the findings in the competition neglect literature, where potential entrants seem to ignore existing competition. However, in the condition where 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 and hence no need to infer demand from competition, the existence of more rivals leads to the more traditional negative congestion effect.

When we interpret these results in light of equation 5, in the condition where only the number of sellers is displayed, the coefficient of -0.0049 corresponds to the total derivative of entry utility with respect to the number of sellers. The coefficient of -0.0070 on the number of sellers when demand information is presented corresponds to the partial derivative that measures the pure congestion effect. The difference between the two coefficients captures the inference effect of competition.

For the results reported in 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.

5.1.2 The Effect of the Number of Buyers

The actual number of buyers shown, when shown in conjunction with supply information, has a positive effect on entry, consistent with expectations.

In the condition where only the number of buyers is displayed, the positive effect of demand information becomes smaller and less significant than in the condition where the numbers of sellers and buyers are displayed in combination. This is because, similarly to the way a potential entrant infers demand from competition, she may also infer competition from demand. In fact, because of the symmetry of equation 3, we can derive a dual formula of equation 5:

$$\frac{dU^S(N^{B*})}{dN^{B*}} = \frac{\partial U^S}{\partial N^{B*}} + \frac{\partial U^S}{\partial N^{S*}} \cdot \frac{\partial N^{S*}}{\partial N^{B*}} \quad (6)$$

We can show that the first component on the right-hand side is positive, which corresponds to the pure surplus-extraction effect of higher demand, whereas the second component represents a negative competition-inference effect. The intuition is not unfamiliar. 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.

5.2 When Should Firms Disclose Information?

The previous estimates show how market information, when presented, affects the relative size of congestion and inference effects on entry. In this section, we ask what information disclosure strategy attracts the most postings to this website.

Table 3 displays the results of a new logit specification. The independent variables include

a constant term and a series of dummy variables that indicate the treatment condition to which a potential seller was assigned. We differentiate high-number conditions and low-number conditions, where a high number of sellers or buyers is defined as a number higher than the median number displayed. These dummy variables are designed to span a typical set of information disclosure alternatives that a two-sided platform operator may be facing. All these dummy variables capture the effects of information provision relative to the baseline condition of no information provided.

We ran two sets of regressions. In the first regression (the first column of Table 3), the dependent variable is a dummy variable of a potential seller’s posting decision. To shed light on how information provision affects the long term growth of platforms, we ran a second regression (the second column of Table 3) to examine the treatment effect on sellers’ decisions to return to the website for repeat business. The dependent variable for the second regression is a dummy variable that equals 1 if a seller returned to the website and initiated the process of making another posting.¹⁰ The rate of repeat postings was small (around 3.5 %). This is not surprising given the market we study, where sellers tend to have few multiple occasions to sell used equipment within a short time.¹¹

In the first regression, the effect of information provision is positive or at worse neutral relative to the baseline of no information provision, although most point estimates were insignificant. This could be because potential posters take information provision as a signal of a higher quality website, or because they derive extra utility from having ready access to

¹⁰We abstract from modeling whether these repeat sellers actually *completed* the posting process because the mere fact that they were exposed to contradictory information conditions might confound the results.

¹¹We were able to add 31 extra observations to this set of regressions, as we did not require data from the Apache log file and consequently we did not lose observations due to the inability to match time-stamps. However, we automatically lost 30 observations because of right censoring, as we could not study the sellers who initially posted on the last day of our experiment because there was no subsequent potential day in our data for them to repost. There were five users who did not post initially but returned on a later date to post. All of these chose to make their post in the category they initially had considered posting in. This group is too small for separate empirical analysis and their inclusion or exclusion in our regression analysis does not change our results.

information. In the short run, the “High Buyer High Seller” condition has a positive and statistically significant effect on posting propensity. Interestingly, potential sellers also react positively at their initial posting stage to the news that there is a small number of sellers, although the news about high seller density does not encourage entry. This asymmetry could be driven by the fact that overall, congestion concerns dominate inference concerns in the market we study.

In the regression that investigates what drives the re-posting decisions, the intercept is significantly lower than in the initial-posting regression. This is not surprising, given the low re-posting rate. The other coefficients are in general more negative than in the initial-posting regression, although they are again insignificant. This could be due to the diminishing role that website design plays in attracting site traffic. The one condition that has a positive significant effect on re-posting is when the potential seller is informed for certain of high demand and low competition. One possible explanation is that by the time of the re-posting decision, a seller has already acquired some experience from the first transaction, and therefore has obtained a better understanding of demand. As a result, competition is encoded more as a sign of congestion, and a less competitive market becomes more likely to encourage entry.

In sum, it appears that advertising “High Buyer High Seller Info” is more effective at generating short run growth than attracting repeat customers; while “High Buyer Low Seller” has more profound long-run effects. The results suggest that in general information disclosure encourages entry in the short run. However, in the long run the most effective strategy to retain sellers on the platform is to convince them that there are more buyers than sellers.

6 Conclusion

This paper examines how demand uncertainty moderates potential entrants' response to the level of competition in the target market. We derive a general theoretical framework in which the effect of competition is decomposed 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. Using field experiment data from a website that brings together buyers and sellers of used goods, we are able to tease apart these two effects empirically. In particular, the number of sellers has no effect on the potential seller's posting decision when that information is displayed in isolation. However, when the number of buyers is also displayed, which renders inference unnecessary, a higher number of sellers reduces posting propensity. 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. Our model and results help to reconcile mixed past findings about the effect of competition on entry. The direction and magnitude of the total effect of competition crucially depends on what market information is available to potential entrants. Previous research has been limited in identifying market information availability as a driver of entry decisions, due to both unobservability and endogeneity challenges.

As well as reconciling opposing findings in the literature on how competition affects entry, our results also yield ready managerial implications to businesses, such as two-sided networks who profit from the total transaction volume that increases with the number of sellers. We find that a good short-run strategy of network growth is to publicize information on both the number of sellers and the number of buyers when demand is high. However, to attract repeat sellers it is most effective to present information that suggests a low number of sellers and a high number of buyers.

A profitable step for further research is to incorporate dynamics in our framework. For example, one possible direction is to explore how an industry evolves while each firm makes strategic entry timing decisions anticipating the information value of their entry to later entrants. It would be also useful to study whether and how repeated inference about demand in a multi-period setting leads to over entry into an industry.

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.
- 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. Lovo (1999, mar). Overconfidence and excess entry: An experimental approach. *The American Economic Review* 89(1), 306–318.
- 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.
- Einav, L. (2007, Spring). Seasonality in the U.S. Motion Picture Industry. *Rand Journal of Economics* 38(1).
- Greve, H. R. (1998). Managerial cognition and the mimetic adoption of market positions: What you see is what you do. *Strategic Management Journal* 19, 967–988.

- Haveman, H. A. (1993, dec). Follow the leader: Mimetic isomorphism and entry into new markets. *Administrative Science Quarterly* 38(4), 593–627.
- Hitsch, G. J. (2006). An empirical model of optimal dynamic product launch and exit under demand uncertainty. *Marketing Science* 25(1), 25–50.
- Kahneman, D. (1988). *Bounded rational behavior in experimental games and markets*, Chapter Experimental Economics: A Psychological Perspective. Springer.
- 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.
- Rapoport, A., D. A. Seale, I. Erev, and J. A. Sundali (1998). Equilibrium play in large group market entry games. *Manage. Sci.* 44(1), 119–141.
- 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. *SSRN eLibrary*.
- Toivanen, O. and M. Waterson (2005). Market structure and entry: where’s the beef? *RAND Journal of Economics* 36(3), 680–699.
- Tucker, C. and J. Zhang (2007). Long tail or steep tail? a field investigation into how online popularity information affects the distribution of customer choices. 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. Technical report.

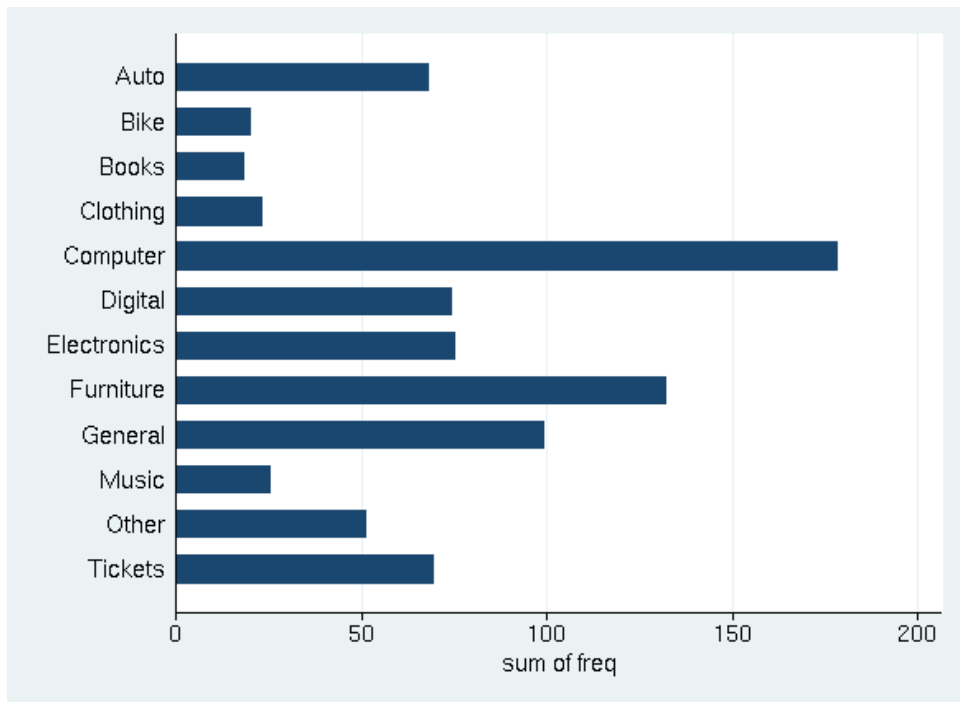


Figure 1: Distribution of Categories

Table 1: Empirical Check of Randomization

	<u>Multinomial Logit Regression</u>			<u>Linear Regression</u>	
	Sellers & Buyers Displayed	Buyers Displayed	Sellers Displayed	Number of Sellers Displayed	Number of Buyers Displayed
Day of Week	-0.1874 (0.1886)	-0.0735 (0.1901)	-0.0885 (0.1854)	2.6258 (4.9979)	-2.8704 (5.3550)
Day of Week SQ	0.0491 (0.0304)	0.0194 (0.0310)	0.0296 (0.0299)	-0.3667 (0.7933)	0.6469 (0.8506)
Time	1.4587 (1.7921)	0.9628 (1.8056)	1.9089 (1.7765)	33.3078 (48.6516)	-8.5331 (52.2237)
Time SQ	-1.5356 (1.5140)	-0.9621 (1.5187)	-1.9359 (1.4957)	-5.8887 (41.2392)	11.3066 (43.9791)
Bike	0.9663 (0.7581)	-0.3067 (0.8953)	-0.0937 (0.8451)	-3.2885 (17.9480)	8.8704 (18.5039)
Books	-0.0777 (0.6775)	-0.9847 (0.8013)	-1.4513 (0.8974)	-10.8669 (19.8739)	10.7204 (21.8778)
Clothing	0.2610 (0.7304)	0.2201 (0.7258)	-0.2428 (0.7620)	-4.2535 (17.4282)	18.0121 (19.6304)
Computer	-0.2511 (0.4369)	-0.1366 (0.4279)	-0.0159 (0.4165)	-7.2079 (11.1086)	8.9923 (11.3956)
Digital	-0.4642 (0.5109)	-0.1214 (0.4845)	-0.3781 (0.4873)	-0.9037 (13.0259)	-1.7071 (13.9858)
Electronics	0.1514 (0.5241)	0.26764 (0.5123)	0.1393 4 (0.5079)	-5.0175 (12.6882)	2.2994 (13.3481)
Furniture	-0.3187 (0.4408)	-0.5814 (0.4437)	-0.5328 (0.4320)	-9.0117 (11.6351)	3.3124 (12.0747)
General	-0.1902 (0.4596)	-0.3369 (0.4589)	-0.8251* (0.4721)	-18.0681 (11.9163)	21.9499* (12.9177)
Music	-0.8745 (0.7333)	-0.8906 (0.7278)	-0.0972 (0.6136)	-12.0588 (21.6051)	2.8962 (18.5756)
Other	-0.2898 (0.5416)	-0.6981 (0.5675)	-0.2425 (0.5187)	-3.5403 (14.9809)	15.0882 (14.6167)
Tickets	-0.7301 (0.5065)	-0.9264* (0.5133)	-0.4625 (0.4723)	8.1021 (14.3043)	26.7844* (13.9181)
Constant	-0.0237 (0.6676)	0.1270 (0.6693)	-0.0205 (0.6600)	88.0694*** (17.7019)	94.2610*** (18.5388)
Observations	832	832	832	408	436
Log-Likelihood	-1,130	-1,130	-1,130	-2,206	-2,386

Table 2: Market Information and Entry Probabilities

	Only #Sellers Displayed	#Sellers and #Buyers Displayed	Only #Buyers Displayed	No Information Displayed
Number Sellers	-0.0049 (0.0041)	-0.0070** (0.0035)		
Number Buyers		0.0065** (0.0031)	0.0029 (0.0024)	
Constant	2.2947*** (0.4574)	2.1460*** (0.4354)	1.4892*** (0.1961)	1.5674*** (0.1559)
Observations	199	209	227	197
Log-Likelihood	-80.12	-73.59	-93.72	-92.36

Sample: Used goods 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

Table 3: Presence of Information and Customer Behavior

	Whether to Post	Whether to Post Again
High Buyer Info Only	0.2993 (0.3420)	0.0193 (0.5629)
High Seller Info Only	-0.0930 (0.2663)	-0.3223 (0.7831)
High Buyer Low Seller Info	0.8530 (0.6071)	0.9521*** (0.3478)
High Buyer High Seller Info	0.5529* (0.3123)	-0.1183 (0.8333)
Low Buyer Info Only	0.1169 (0.3049)	-0.4833 (0.5063)
Low Seller Info Only	0.5959** (0.2334)	0.0358 (0.4761)
Low Buyer High Seller Info	0.0196 (0.4184)	-0.7935 (1.1539)
Low Buyer Low Seller Info	0.4475 (0.3221)	
Constant	1.5674*** (0.2839)	-3.2139*** (0.3735)
Observations	832	831
Log-Likelihood	-341.50	-129.79

Sample: Potential sellers
* p<0.10, ** p<0.05, *** p<0.01
Standard Errors clustered by category