

# The Salience of Entrepreneurship: Evidence from Online Business

Yi Huang, Chen Lin, Sibol Liu, Gustavo Manso\*

May 8, 2021

## Abstract

We study the psychological biases underlying people's decision of becoming an entrepreneur in the context of online business. Using the entire universe of entrepreneurs on the world's largest online shopping platform, we find that people are more likely to become an online entrepreneur when observing the emergence of successful stores located in the close neighborhood. We rely on the store rating system and detailed geographical information for identification. The rating upgrade of an online store led to a significant increase in the number of new stores in the 0.5km radius of the neighborhood across rural areas of China. The effect increases with the level of upgrades, decreases with distance and, is robust to a wide range of rigorous model specifications. The entrants motivated by observing these upgraded stores underperform relative to others in terms of lower sales and a higher probability of exit, suggesting the entry decision is suboptimal. Overall, the results are most consistent with salience theories of choice and cannot be explained by regional development or rational learning.

*Keywords:* entrepreneurship, peer effect, salience theory, availability heuristic.

---

\*Huang is with Graduate Institute of Geneva, email: [yi.huang@graduateinstitute.ch](mailto:yi.huang@graduateinstitute.ch); Lin is with the University of Hong Kong, email: [chenlin1@hku.hk](mailto:chenlin1@hku.hk); Liu is with Lingnan University, email: [siboliu@ln.hku.hk](mailto:siboliu@ln.hku.hk); Manso is with University of California, Berkeley, email: [manso@berkeley.edu](mailto:manso@berkeley.edu). We acknowledge the generous support from Ant Financial Services Group for the data and insightful discussions on the institutional details.

**Disclaimer:** The authors highlight that the data and analytics reported in this paper may contain errors and are not suited for the purpose of company valuation or to deduce conclusions about the business success and/or commercial strategy of Ant Financial. All statements made reflect the private opinions of the authors and do not express any official position of Ant Financial and its management. The analysis was undertaken in strict observance of the Chinese law on privacy.

## 1. Introduction

What induces people to become entrepreneurs? Previous studies have considered the factors that motivate entrepreneurial activities including geography (Michelacci and Silva, 2007; Glaeser, et al. 2010; Pistaferri, et al. 2021), parental background (Lindquist and Van Praag, 2015; Li and Goetz, 2019; Nicolaou, et al. 2008; Dohmen, et al. 2012), financial constraint (Evans and Jovanovic, 1989; Hurst and Lusardi, 2004; Andersen and Nielsen, 2012; Hvide and Møen, 2010), etc. And there is a growing literature that highlights the importance of behavioral reasons underlying entrepreneurial decisions (e.g., Bernardo and Welch, 2001). In this study, we investigate the psychological biases underlying people's decision of becoming an entrepreneur in the context of online business. Specifically, we examine whether the decision to become an online entrepreneur is influenced by peers in the same neighborhood and evaluate the subsequent performance of the induced entrants.

It is widely perceived that entrepreneurs are motivated by role models or the entrepreneurial risk culture they represent.<sup>1</sup> Yet, the empirical evidence is mixed on whether entrepreneurial decisions depend on the behavior of other individuals. Some studies document a positive effect. For example, using data from Sweden, Giannetti and Simonov (2009) find individuals from highly entrepreneurial neighborhoods are more likely to become entrepreneurs. Based on the population data of workers in Denmark, Nanda and Sorensen (2010) find that an individual is more likely to become an entrepreneur if their coworkers have been entrepreneurs before. However, some studies stress the negative peer effect on entrepreneurship. For example, Lerner and Malmendier (2013) exploit the exogenous peer assignments in business-school sections and find a higher share of entrepreneurial peers decreases entrepreneurship. In contrast to these studies, we aim to investigate not only the effect of peers but also the effect of the noticeably successful peers.

It is empirically challenging to identify the peer effects in entrepreneurship. First, red tapes

---

<sup>1</sup> It was revealed that the Hewlett-Packard Company had motivated Steve Jobs, who inspired many subsequent entrepreneurs all around the world.

and administrative barriers make the timing of the entry difficult to observe. Specifically, the time lag to register a company varies from region to region, generating difficulties in identifying concurrent factors that affect the entry. As we will further discuss below, we mitigate this issue by examining the entry decision on the online marketplace where the registration process is efficient. This precise information on the entry timing facilitates our analysis of the intertemporal variations in factors that motivate entrepreneurship. Second, it's usually difficult to measure the performance of the entrants due to the data limitations. We, in contrast, rely on the high-frequency performance measures, such as the timing of exit, provided by the online platform. The detailed information on the store's location also allows us to identify peers' impact in a clean fashion.

The business models based on the internet have exploded over the past decade. According to the estimate from the United Nations Conference on Trade and Development (UNCTD), in 2018, e-commerce sales hit \$25.6 trillion globally, and more than 1.4 billion people shopped online. E-commerce sales in the U.S. account for 42% of GDP.<sup>2</sup> Recent studies have evaluated the impact of online business in terms of price-setting (Cavallo, 2017, 2018; Goolsbee and Klenow, 2018), responses of the traditional business model (Seamans and Zhu, 2014), and welfare effect (Dolfen et al., 2019). It is less clear, however, whether incentives underlying online entrepreneurs are distinct from offline counterparts. To approach this issue, we conduct our research using the universe of online transactions and store information on Taobao Marketplace (Taobao.com), the largest online consumer-to-consumer (C2C) retail shopping platform in the world. As subsidiaries of Alibaba Group, Taobao Marketplace (C2C platform) and Tmall (B2C platform) combinedly were serving over 10 million sellers and more than 454 million active buyers all around the world per year (Ali Research, 2017). We focus on Taobao Marketplace as it serves individual online merchants<sup>3</sup> in China, making it an ideal laboratory for studies in online entrepreneurs.

---

<sup>2</sup> Source: <https://unctad.org/press-material/global-e-commerce-hits-256-trillion-latest-unctad-estimates>

<sup>3</sup> As the counterpart of Amazon, Tmall, is a platform that serves offline incorporated businesses selling products online.

We access the highly disaggregated data from Taobao Marketplace between August 2014 to August 2016 with detailed information on store performance, the precise location of the stores, demographic information on store owners, etc. The online business fluctuates in a very timely fashion. For example, only half of the newly registered stores could survive more than five months. Therefore, the 2-year sample still allows us to identify the dynamics of online entrepreneurial activities. We rely on the noticeable store rating upgrades to identify the peer effect. Specifically, we identify stores that experience rating upgrades during the sample period and investigate the emergence of the new stores in the neighborhoods with different distance cutoffs. We restrict the data to stores operating in China's rural areas, where people exchange information through social networks more frequently.

To render an initial support to our empirical design, we collaborated with a market research company and conducted a survey on online store owners among 12 provinces and 68 cities in China. We interviewed the sample store owners regarding their motivations in the decision of becoming online entrepreneurs. The findings suggest that people are largely affected by the entrepreneurial activities in their neighborhood. In particular, the performance of the store, measured by store ratings, indeed attracts the attention of the locals when considering to open an online store. People also tend to be disappointed and regret when initially motivated by peers.

The baseline estimation based on the archival data from Taobao.com suggests people are more likely to become an online entrepreneur when observing store upgrades in their neighborhood. The rating upgrade of an online store leads to a 3.5% increase in the number of new stores in the 0.5km radius of the neighborhood relative to the sample mean. The effect increases with the level of upgrades, decreases with distance, and is robust to a wide range of rigorous model specifications. In particular, we include region-by-year fixed effects and economic controls at the granular level to capture the local trends and mitigate the

omitted variable concerns. And the documented effect is concentrated among newly established stores selling the same product as the upgraded store, ones run by less educated owners, unmarried owners, and owners with low estimated income.

Additional tests suggest the entry decision is suboptimal. First, the dynamic analysis suggests the effect only exists in the very month of store upgrade. The short-lived effect implies people's attention is channeled towards those salient events temporarily. Second, the entrants motivated by observing these upgraded stores underperform relative to others in terms of lower sales and a higher probability of exit within 3-month or 6-month following the registration. By including store fixed effects and region trends, we compare the share of exit among stores registered in the upgrade month with other entrants. Third, we find entrants who are motivated by peers are more likely to ignore the current product market conditions, leading to inadequate economic outcomes. To sum up, the empirical findings suggest the excess entry of online stores induced by peer effects are driven by behavioral reasons.

We argue that the results are most consistent with salience theories of choice pioneered by recent studies (Gennaioli and Shleifer, 2010; Bordalo, Gennaioli, and Shleifer, 2012; Kőszegi and Szeidl, 2013). Studies in salience theory suggest people overweight information that their minds focus on instead of fully consider all available information. That is, the salient payoffs are overweighted, leading to risk-seeking activities when the upside payoff is salience and risk-averse activities when the downside is salient. As entrepreneurial activities are extremely risky,<sup>4</sup> the salience theory is consistent with the observed suboptimal entry decision as people tend to focus on high payoff outcomes in the local area instead of full information. In addition, salience theories assume that the specification of salience captures the ideas that people attend to differences rather than absolute values. This assumption is consistent with our identification of salience captured by changes in store ratings. In contrast, the findings in our analysis are not supportive of the alternative stories, including demand-driven

---

<sup>4</sup> According to U.S. Bureau of Labor Statistics, roughly 50% of new business establishments could survive more than 5 years. Source: <https://www.bls.gov/bdm/entrepreneurship/entrepreneurship.htm>

explanations, regional development, and rational learning.

We contribute to the literature in three aspects. First, we add to the literature on drivers for entrepreneurship. In the context of online business, we highlight the behavioral reasons in motivating entrepreneurship. Previous studies mainly investigate the correlation between entrepreneurial decisions among peers (Sweden, Giannetti and Simonov, 2009; Denmark, Nanda and Sorensen, 2010, Lerner and Malmendier, 2013). We, in contrast, focus on how the noticeable change in peer's performance affects the entry decision.

Second, our study joins previous works on salience theories. Recent studies documented how salience theories can explain patterns in consumption behavior (Bordalo, Gennaioli, and Shleifer, 2013a; Busse et al., 2015), asset prices (Bordalo, Gennaioli, and Shleifer, 2013b), tax effects (Chetty, Looney, and Kroft, 2009), insurance purchases (Chang et al., 2018), college major choice (Choi, Lou, and Mukherjee, 2016), and corporate policy (Dessaint and Matray, 2017). We extend the analysis to the realm of entrepreneurial activities. In particular, our identification is based on the intertemporal variation in the store's rating, which ties closely to the assumption of salience theories. The detailed measures for aftermarket performance also deliver additional support to the behavioral explanations for the excess entry.

Last, our paper relates to the emerging literature on business creations in the digital economy. The online business models usually feature two-sided markets (e.g., Seamans and Zhu, 2014; Rysman, 2009), utilize user-generated data (Mayzlin et al., 2014; Cohen et al., 2016), and demonstrate distinct business strategies (Cavallo, 2017, 2018). However, the origins and determinants of online entrepreneurship are less understood, mainly due to the data limitation. In a study, Fan et al. (2018) assess the welfare gains from e-commerce using region aggregated data from Alibaba. Two contemporaneous papers using the same dataset in our study analyze how the microcredit allocation shapes online stores' growth (Hau et al., 2019) and service quality (Huang et al., 2019). We, in contract, evaluate the behavioral reasons in

the online business creation. By examining the role of social interactions in online entrepreneurship, we also contribute to the growing literature in social networks and finance (e.g., Breza, 2012; Iyer and Puri, 2012; Banerjee et al., 2018; Breza and Chandrasekhar, 2019).

The remainder of the paper proceeds as follows. We, in the next section, introduce the institutional background of Taobao Marketplace and its store rating system. We develop our hypotheses in Section 3 and describe the data and our identification strategy in Section 4. Section 5 presents our findings. And Section 6 concludes.

## **2. Institutional Background and Data**

### **2.1 Taobao Marketplace**

Our sample consists of online transactions on Taobao Marketplace, which is a subsidiary of Alibaba Group, a tech-service conglomerate listed in NYSE and HKEX. Within two years after its establishment in May 2003, Taobao Marketplace quickly gained its dominance. Taobao Marketplace mainly supports C2C retail business via its platform on which the stores are established, owned, and managed by the individual merchants. One noticeable feature is about 97% of the merchants on Alibaba platforms employ less than 5 workers, and 95% of them invest less than \$4,500 (Ali Research, 2017). As those merchants who run the online business are not employees in the companies or public sectors, they are consistent with the typical definition of entrepreneurs in the literature (e.g., Hurst et al., 2014).

Our study is also related to the phenomenon that "Taobao Villages" grew rapidly and spread across regions in China, as later we will restrict the sample to stores operating in rural areas for identification purposes. "Taobao Villages" refer to villages that are significantly engaged in e-commerce with a total annual e-commerce transaction volume of at least RMB 10 million and at least 100 active online shops. Online stores are geographically clustered in those villages, and e-commerce has become the village residents' primary business. According to joint research between the World Bank and the Alibaba Group, "Taobao

Villages" had increased from 20 in 2013 to 3,202 in 2018.<sup>5</sup> In 2020, there are over 5,425 "Taobao Villages" and over 1,700 "Taobao Townships" across 28 provinces, creating 8 million jobs.<sup>6</sup> However, Couture et al. (2020) find little evidence for income gains to rural producers and workers after adopting e-commerce in those areas.

## 2.2 Online Store Registration

Unlike starting a traditional offline business, to register an online store is efficient. Appendix 1 illustrates the registration procedures. The whole process will be through the Taobao platform, which requires the essential documentation, including name, personal identity card, registered phone number, and bank account. The platform will review the application. When all information is complete and verified, the application will be approved within 72 hours. The initial cost of store registration is RMB 1000 (about USD 144). This registration fee is refundable when the store is dissolved. The owner is required to pay a monthly technical support fee of RMB 50 (about USD 7.2). The precise date of the application is available to us, allowing us to identify the timing of the entry decision. The entry barrier is much lower for e-commerce on Taobao Marketplace comparing with the traditional start-ups. In what follows, we will discuss how we capture the entrepreneurial decisions in the context of online business in detail.

## 2.3 Raw Data Description

We access the online transaction data on Taobao Marketplace from August 2014 to August 2016. Figure 1 shows a summary of the raw data over the sample period. As shown in Panel A, the platform's total sales are about 25 billion RMB per month. The troughs in February mark the Chinese New Year when all business operations become inactive. On average, there are 2.5 million active stores in our sample. The number of stores remains stable over time. Panel B depicts the trajectory of new stores registered in the sample period. Later in our

---

<sup>5</sup> Source: <http://www.aliresearch.com/en/Research/Researchdetails?articleCode=21626&type=TaobaoVillages>

<sup>6</sup> Source: <http://www.aliresearch.com/ch/information/informationdetails?articleCode=126860487966199808&type=%E6%96%B0%E9%97%BB&adcode=&villageCode=&villageYear=&item=%E6%B7%98%E5%AE%9D%E6%9D%91>



analysis, we will impose reasonable restrictions on the sample for identification.

[Figure 1 about here]

The performance of online stores is highly skewed. We present the distribution of sales in Figure 2. Panel A shows the probability density function of monthly sales for each store, while Panel B presents the log-log scale plot, indicating a power-law distribution. The Pareto tail is consistent with theories that explain the inequality patterns in the economy (Gabaix, 2016; Gabaix et al., 2016).

[Figure 2 about here]

There is a considerable variation across regions in the number of stores on the Taobao Marketplace. As shown in Figure 3, more stores are clustered in the coastal areas, partly due to better infrastructures in terms of information technology, logistics, and supply chains. We build our analysis of the time-varying differences across regions.

[Figure 3 about here]

## **2.4 Store Rating System**

Taobao Marketplace establishes the store rating system to improve the store quality's transparency through feedbacks from customers. It has been shown in the literature that the reputation system in the online platform mitigates concerns regarding informational asymmetries. It informs future buyers about the outcomes of a seller's past behavior, turning a one-shot game into a repeated game. Luca (2016) finds a restaurant rating on Yelp has a causal impact on consumer demand. Klein, Lambertz, and Stahl (2016) document eBay's adoption of the Feedback Mechanism improves online stores' service quality.

Taobao's rating system records a cumulative rating for a merchant since the start of its business. For every completed transaction, the customer is invited to evaluate his shopping experience with one of the three feedbacks, "Positive", "Neutral", and "Negative". Each feedback bears a score that is to be added to the total feedback score of the merchant. A "Positive", "Neural", and "Negative" feedback contributes one point, zero point, and

negative one point respectively to the internal total score. The internal total score is then converted to a 20-level icon rank for the merchant as the cumulative store rating. We illustrate the rating system in Figure 4. The first column in Panel A shows the cutoffs in the internal score. The second column shows the symbol of the rating, the lowest rank marked by one heart, and the highest rank marked by five gold crowns. We assign a number, as shown in column 4 to each rating. The internal score in column 1 is not available to us or the public. As shown in Panel B, however, the overall rating of the store is publicly available. We exploit the visible rating for each store for identification, which will be discussed in Section 4.

[Figure 4 about here]

### 3. Theoretical Framework and Hypotheses

We investigate whether an individual's entrepreneurial decision is motivated by the performance of his/her peers. We exploit the discrete change in rating to identify the visible shift in store performance. As we will discuss later, the rating is an informative measure for the past performance and a forward-looking indicator that predicts a store's sales and online traffic volume.

Previous studies document both positive and negative peer effects in the traditional offline business (Sweden, Giannetti and Simonov, 2009; Nanda and Sorensen, 2010; Lerner and Malmendier, 2013). In particular, Lerner and Malmendier (2013) document a negative effect of peers, which is confined to unsuccessful entrepreneurs. In contrast, we look at whether the intertemporal change in peer's performance plays a role. We start off with the following hypothesis.

To motivate our analysis, we propose a simple model as follows. The market has incumbent online stores in  $N$  different neighborhoods. A store rating is either  $1$  with probability  $p$  or  $-1$  with probability  $1 - p$ . There are two types of risk-neutral entrants, one who observes all incumbents' ratings and the other who only observes ratings in their own

neighborhoods. After observing the ratings of incumbents, entrants decide whether to enter the market. Profits of each entrant are equal to its realized rating minus a cost  $c$  to enter the market.

Entrants do not know the probability  $p$  of a positive rating but have a prior  $1/2$  that market conditions are virtuous and the probability of a positive rating is  $q$  and  $1/2$  that market conditions are poor and the probability of a positive rating is  $1 - q$  with  $q > 1/2$ . Store ratings are conditionally independent. From Bayes rule, after observing the incumbents' ratings, the odds of virtuous market conditions versus poor are:

$$o(n) = \frac{P(p=q|n)}{P(p=1-q|n)} = \left(\frac{q}{1-q}\right)^n,$$

where  $n$  is the sum of the observed incumbent ratings. Given odds ratio  $o(n)$ , the profit from entering is:

$$\begin{aligned} \frac{o(n)}{1+o(n)}(q - (1-q)) + \frac{1}{1+o(n)}((1-q) - q) &= \\ &= \frac{o(n) - 1}{1+o(n)}(2q - 1) \end{aligned}$$

We assume that entry is expected to be profitable with some positive evidence ( $n \geq k$ ) from incumbent ratings:

$$c < \frac{o(k)-1}{1+o(k)}(2q - 1).$$

This implies that given available evidence  $n$  expected profits,

$$\frac{o(n)-1}{1+o(n)}(2q - 1) - c,$$

are positive if and only if  $n \geq k$ .

Incumbents observing all ratings enter when  $n \geq k$ . Their expected profits will be

$$\frac{o(n)-1}{1+o(n)}(2q - 1) - c > 0.$$

On the other hand, incumbents who only observe ratings in their neighborhood have to rely on that subset of signals to make a decision. Incumbents in neighborhood  $i \in N$  will enter the market when  $n_i \geq k$ , where  $n_i$  is the sum of the incumbent ratings in neighborhood

*i.* If it turns out that when they enter the overall market is doing well, which happens with probability  $P(n \geq k | n_i \geq k)$ , then their expected profits will be positive and equal to the profits of incumbents observing all ratings. However, if it turns out that when they enter the overall market is doing poorly, which happens with probability  $P(n < k | n_i \geq k)$ , then their expected profits will be negative and less than the profits of incumbents observing all ratings.

In the model, many entrants rely only on signals from their own neighborhood. These entrants will on average underperform other entrants, some of which rely on all signals. The underperformance will be especially severe when their neighborhood is doing well, which will trigger their entry, but the overall market is doing poorly. We formulate our two hypotheses as follows.

**H1:** People are more likely to become an online entrepreneur when observing store rating upgrades in the neighborhood.

Empirical evidence supporting this hypothesis is consistent with a rational learning story, regional economic development, product demand, or behavioral explanations, such as salience theories. Specifically, a rational learning story suggests that people acquire useful information and identify business opportunities by observing the store performance in the neighborhood. The positive correlation between peer's performance and entry decision may be due to some common factors, such as improved infrastructure, demand shocks on some specific products that a particular region has a comparative advantage. The entry decision can also be because of behavioral reasons that people ignore the full information and make suboptimal choices. To further disentangle the theories that can explain the hypothesis, we explore the economic outcome of the entry decision.

**H2:** Stores motivated by store rating upgrades in their immediate neighborhood

underperform others.

Rational learning and economic confounders suggest that the entry decision tends to be rational and value-enhancing, therefore contradicting **H2**. Behavioral reasons related to salience, however, are consistent with **H2**, suggesting the entry decision is suboptimal.

#### **4. Sample, Variables, and Empirical Design**

We present in this section the sample and the empirical design. In particular, we discuss how we construct the sample, the definition of the entry, and the model specifications we employ.

##### **4.1 Identification and Sample Construction**

Our identification of the impact of peers' performance on the online entrepreneurial decision is built upon several essential assumptions. First, people pay attention to their peers in the neighborhood. Second, the signal sent to observers is informative about the performance of a store. Third, when people make the entry decision, the registration process shall be efficient enough so that we can link the decision timing to the concurrent factors of interest. As we discussed above, the efficient registration process on Taobao Marketplace fulfills the third assumption. We thus focus on the first two assumptions in the following.

Regarding the first assumption, we cannot provide direct evidence that people pay close attention to other people in the same neighborhood. But some studies highlight the importance of word of mouth and social networks in shaping high stake decisions (Morris, 2000; Bailey et al., 2018; Breza and Chandrasekhar, 2019). For example, Hong et al. (2005) document a mutual fund manager's portfolio is correlated with other managers' in the same city due to the information shared through social networks. And social networks are critical in areas without well-functioning formal institutions (Chandrasekhar et al., 2018). Therefore, we confine our analysis to stores located in rural areas of China, where residents rely more on social interactions in cooperation so that peer effect could reasonably manifest.

We illustrate the procedures to identify stores located in rural areas in Appendix 3. Specifically, we access the store's precise coordinates and match with administrative divisions of China using geographic information system (GIS) techniques. We follow the official definition of rural area in China using the urban-rural classification code. Appendix 3 Panel B shows in the sample of stores with precise location information, 59.5% of stores are located in the urban areas, and 40.5% are in the rural areas. Comparing with rural stores, urban stores have larger sales per month and a better rating. However, as shown in Appendix 3 Panel C, the product categories are not significantly different between urban stores and rural stores.

Regarding the second assumption, we employ the rating upgrade as a signal that will influence people's decisions in entrepreneurship. Store rating is publicly visible, while the store's sales are not directly observable to other people in the neighborhood. By construction, store rating reflects the retrospective sales and service quality of a store. But the rating is also informative in predicting the future success of a store. As shown in Figure 5, the rating is monotonically correlated with monthly sales and pageviews of a store. The number of pageviews represents consumers' attention to this store, which will translate to product demand and revenue. Therefore, we argue that the abrupt change in the rating is a valid signal for a peer's performance.

[Figure 5 about here]

Table 1 Panel A further demonstrates the summary of store performance across store ratings for the full sample at the monthly frequency. In the first column, we assign a value, "Rating," to each rating category. To capture the performance change, we calculate the difference between Rating in month  $t$  and Rating in month  $t-1$  and present the summary in Panel B. The store rating upgrades are rare events, accounting for less than 4% of total store-month observations. It suggests rating upgrade marks the millstone in the life cycle of a store. We observe very few downgrade events. We first identify stores that experience upgrade events in the sample period and then retrieve the monthly observations for this group of stores.

The final sample contains 3,052,391 store-month observations with 133,583 distinct rural stores that experience rating upgrades during the sample period. Appendix 2 summarizes the sample construction process.

[Table 1 about here]

## 4.2 Variables

We gauge the entrepreneurial entry decisions in the neighborhood centering at the upgraded store. First, we define the registration month as the entry month. Second, we restrict the entrants to ones with at least one-month of non-zero sales in our sample period. This restriction mitigates the concern that people may register the store without additional entrepreneurial activities.<sup>7</sup> We construct the number of entrants according to different distance cutoffs. For example,  $\text{Entrant [0.5 km]}_{i,t}$  is the number of stores registered operating in 0.5 km radius of the upgraded sample store  $i$  in month  $t$ .<sup>8</sup>  $\text{Upgrade}_{i,t}$  is an indicator variable set to one if the sample store  $i$  experiences positive changes in the rating in month  $t$ , and zero otherwise. When a store experience multiple upgrade events, we only consider the first event. Table 2 shows the summary statistics of the baseline panel.

[Table 2 about here]

We construct some regional time-varying variables to pick up economic confounders. County Store Number  $_{c,t}$  is the number of active stores in a given county  $c$  in month  $t$ . County Sales  $_{c,t}$  is the average monthly sales for active stores in a given county  $c$  in month  $t$ . As a measure of economic growth, we construct County Light  $_{c,t}$  as the log value of the night light in a given county  $c$  in month  $t$  following Henderson et al. (2012). We capture store performance using exit probability. Specifically,  $\text{Entrant Exit Share (3m/6m)}_{i,t}$  is the percentage of stores that exit in the 3-month/6-month period after the registration for stores that registered in 0.5 km radius of the upgraded sample store. These variables can only be defined for store-

---

<sup>7</sup> In the table available upon request, we show that the results are qualitatively similar when this restriction is removed.

<sup>8</sup> There are rare cases that newly registered stores with the exact same location as the upgraded stores. We remove those cases to mitigate the concern that the motivated entrants are closely connected to the upgraded store owners.

month observations where at least one store is registered in the neighborhood and missing otherwise. All variables are defined in Appendix 4.

### 4.3 Model Specifications

We estimate the following ordinary least square (OLS) regression model.

$$Entry_{it}^D = \alpha_0 + \alpha_1 Upgrade_{it} + X_{it} + \theta_i + \gamma_{ct} + \mu_{pt} + \varepsilon_{it} \quad (1)$$

where the sample is at the store-month level.  $Entry_{it}^D$  is the number of entrants in  $D$  km radius of store  $i$  in month  $t$ .  $Upgrade_{it}$  is the main variable of interest, an indicator for a positive change in the rating for store  $i$  in month  $t$ . A significantly positive  $\alpha_1$  indicates the existence of the peer effect, supporting **H1**.  $X_{it}$  is regional characteristics.  $\theta_i$  denotes store fixed effects.  $\gamma_{ct}$  denotes city-by-month fixed effects.  $\varepsilon_{it}$  is error term. The standard errors are two-way clustered at the store and month level. In the subsequent analysis regarding **H2**, the dependent variables will take on the value of performance measures.

To assess the dynamic effect of peer performance, we estimate the following regression model with lagged independent variables.

$$Entry_{it}^D = \alpha_0 + \sum_{j=1}^5 \alpha_{1,-j} Upgrade_{i,t-j} + X_{it} + \theta_i + \gamma_{ct} + \mu_{pt} + \varepsilon_{it} \quad (2)$$

where  $j$  denotes the number of lag periods. We include a total of 5 lags. Other model specifications follow the convention in Equation (1).

## 5. Results

### 5.1 A Survey on Rural Taobao Merchants

Before we investigate the archival data for online store at Taobao.com, we start off our analysis by conducting a field survey on Taobao owners (merchants). We collaborated with a market research company based in mainland China and interviewed 305 individuals each managing a Taobao store. Specifically, we restrict our online store to be ones normally running and located in rural areas of China. The sample stores are from 12 provinces and 68 cities. The sampling approach roughly mirrors the geographical distribution of online



stores across different Chinese provinces. We present the statistics on the answers to 8 questions regarding the motivations for online entrepreneurship in Appendix 5. We summarize our findings as follows.

First, the survey results suggest that people are motivated by peers in the neighborhood in the decision of becoming an online entrepreneur. The answers to Question 1 reveal that 46.2% of the interviewees who list “observing other locals running Taobao store” as the motivation for online entrepreneurship. This reason is the third most mentioned reasons right after “gaining wealth” (86.9%) and “becoming your own boss” (55.4%). Other reasons, such as flexible schedule and government support, are less important comparing with the motivations mentioned above. The evidence is supportive to the assumptions in our theoretical model that people tend to overweight signals in their neighborhood.

Second, we also find that peers in the neighborhood also affect how merchants run the business through social interactions. Specifically, the survey results indicate online merchants often communicate with other other (Question 2) and are affected by peers in terms of business model and scope (Question 3) as well as product choice (Question 4), while a considerable portion of merchants do not analyze the market demand of different products when you choose the type of products to operate (Question 5). The findings is consistent with our research design that focus on people in the rural area where social interactions are perceived to be essential in day-to-day life. We also build our heterogeneous tests upon the survey results regarding the inattention to the product demand.

Last, some findings in the survey support our research design detailed in Section 4. Most of merchants (76.7% in Question 6) pay attention to the performance of nearby stores, such as store ratings, thereby supporting our identification strategy based on the rating upgrade events. Merchants who are motivated by the performance of peers are more likely to be disappointed (51.3% in Question 7) or even exit the market (39.2% in Question 8), which

mirrors our tests on the aftermarket performance and exit decisions.

## 5.2 Baseline Estimation

We present the baseline estimation results in Table 3. In column 1, we estimate the standalone effect by only including store and month fixed effects. The coefficient is significantly positive, indicating that the upgrade event leads to more newly registered stores in the 0.5km radius neighborhood. We compare the coefficient with the mean of the dependent variable to assess the economic magnitude. The effect in column 1 accounts for a 7.6% increase relative to the sample mean. We include regional control variables in column 2, and the results remain similar. In columns 3 and 4, we consider city-by-month fixed effects to pick up any city-level trends. The positive peer effect accounts for 3.5% of the sample mean. The evidence in Table 3 is supportive of **H1**.

[Table 3 about here]

We conduct a battery of robustness checks and present the results in Appendix 6. First, we report the results estimated using standard errors clustered at the township level or at both township and month level. The results in Columns (1) and (2) of Appendix 6 shows the results are robust to this alteration. We then replace the dependant variable with the inverse hyperbolic sine transformation of the number of entrants to correct for the potential bias due to the distribution of the original count variable. The result in Column (3) shows the results remain qualitatively similar. Another potential concern is that our findings might be driven by changes at the product level either on the demand-side or supply-side. We mitigate this concern by constructing a store-product-month level sample and re-estimate the effect of the peers. Our findings still hold when we consider the product-level trends (Column (4)) or the trends at the city-product level (Column (5)).

We estimate the dynamic effect using Equation (2) and present the results in Table 4. Our baseline model with regional controls in column 2 indicates the effect only exists in the contemporaneous month, and all lagged variables are insignificant. The short-lived effect is

inconsistent with explanations based on certain long-term economic confounders and more suggestive arising from behavioral reasons about individual attention (see, e.g., Busse et al., 2015; Chang et al., 2018).

[Table 4 about here]

### **5.3 Dosage of Salience**

Our definition of rating upgrade is compatible with salience theories that emphasize that people attend to differences rather than absolute values (Gennaioli and Shleifer, 2010). We explore the heterogeneous effect along two dimensions that capture the dosage of salience.

Panel A of Table 5 presents the peer effect on the entry decision by the upgrade level. As shown in Panel B of Table 1, there are few cases that a store upgrade by more than one level, reflecting a more salient signal to observers in the neighborhood. We split the original treatment variables into two. Upgrade [1 level] pinpoints the timing of the upgrade with one level. Upgrade [2+ level] marks the upgrade events with more than one level. As shown in Table 5, Panel A, Upgrade [2+ level] delivers a large impact, accounting for 9.2% of the sample mean.

We next dig into the results according to the distance. Specifically, we construct the dependent variable using alternative distance cutoffs. The results presented in Table 5 show that the effect within a 0.1km radius is similar to that in regions in the radius between 0.1km and 0.5km. But the effect quickly weakens beyond 0.5km and becomes insignificant beyond 1km. The pattern that the effect decays by distance is consistent with both rational learning and behavior explanations. A rational learning story implies the acquisition of better information by the decision-makers leads to better decisions as information quality declines with distance (Bernstein et al., 2016; Huang et al., 2017). Among behavioral explanations, the availability heuristic suggests that people tend to rely on immediate examples that come to a given person's mind when evaluating a specific topic, concept, method, or decision (Folkes,

1988).

[Table 5 about here]

#### **5.4 Types of Entrants**

The detailed information on the store's product category and owner demographics enables us to investigate the entrants by type. We first explore the heterogeneous effect by product categories. In the data provided by Taobao Marketplace, we can observe the main product category assigned to each store. We count the number of entrants with the same product category as the upgraded store. Conversely, we define entrants with product categories different from the upgraded store. As shown in Panel A of Table 6, the upgrade events mainly lead to more stores selling products in the same category as the upgraded stores. We argue that the finding is consistent with both behavioral reasons and rational learning. Entrants may mimic the behavior of the upgraded stores due to behavioral biases. Alternatively, entrants may learn from the upgraded store about profitable products to sell. It is also possible that in some regions, only certain types of products can be produced.

We next examine the heterogeneous effect on the store owner's demographics. Figure 6 presents the percentage of entrants by age group, education, gender, and marital status in our baseline sample. For example, most of the store owners do not hold a college degree. Panel B and Panel C of Table 6 indicate that people who are over the age of 30, who are less educated, who are female or not married are more subject to the effect of peer's performance. Panel D shows the results based on the owner's estimated income level provided by Taobao platform. Income level is highly positively correlated with age. Therefore we define the age-adjusted income group by categorizing the newly registered stores into seven groups according to the estimated monthly income in each age group. The age group is illustrated in Figure 6. Income Group (1) thus indicates the income group with the lowest income level. And Income Group (5) indicates the income group with the highest income level. Panel D suggests the peer effect decreases with the level of income. It is consistent with more

sophisticated people proxied by income level are less prone to behavioral biases.

[Table 6 about here]

## 5.5 Performance

We next test **H2** to differentiate further the theories that can better explain our findings. We measure the performance of entrants in the subsequent 3-month and 6-month, respectively. The results are reported in Table 7. The dependent variable here is calculated using the newly registered stores' sales instead of the upgraded store. Average Entrant Sales (3m) is the average sales for the entrants over a 3-month period after the registration. Average Entrant Sales (6m) is defined similarly for a 6-month period. If there are no entrants in a given month, the dependent variable cannot be defined and dropped accordingly. Table 7 suggests stores motivated by upgrade events underperform by a range between 7.9% to 8.7% of the sample mean. The results are supportive of **H2**.

[Table 7 about here]

When store owners realize the entrepreneurial activities are not rational, they may reverse the decision and exit. We calculate the percentage of the entrants that exit after the registration in the 3-month period and in the 6-month period, respectively. Table 8 suggests stores motivated by the upgraded stores are more likely to exit, in a range between 1.5% to 1.8% of the sample mean, suggesting the original entry decision tends to be suboptimal

[Table 8 about here]

The model proposed in Section 3 predicts that entrepreneurs are more likely to rely only on signals from their own neighborhood, leading to an inadequate economic consequence. We test this prediction by constructing the product level sales growth in a given city from month  $t-1$  to month  $t$ . In Table 8, we interact this variable with the main treatment variable. Columns (1) and (2) indicate that the underperformance of the entrants are exaggerated when the market condition in a specific product is experiencing a downturn. We also define a binary variable, *Low Product Sales Growth*, which is set to one if the aggregated sales growth for product  $k$  from month  $t-1$  to month  $t$  for stores located in city  $c$  is below median and zero

otherwise. Columns (3) and (4) of Table 8 shows consistent results that confirms the model predictions that store owners motivated by peers that underperform are more likely to ignore the overall market condition and overweight the signals in the close neighborhood.

[Table 9 about here]

To sum up, the results on the exit decision support **H2** and thus contradict with explanations for our main findings such as rational learning and economic confounders. The results are mostly consistent with behavioral reasons, in particular, the salience theories.

### **5.6 Downgrade Events**

We also explore the impact of the downgrade events. As shown in Panel B of Table 1, stores rarely experience downgrades. In our sample, we only identify 520 stores that experience rating downgrade events with precise location information and other essential variables. In Appendix 7, we present the results using an indicator, Downgrade, which is set to one for the month of rating downgrade for a given store. We do not find significant effect of these events on the number of entrants in the neighborhood, suggesting that the entrepreneurial decisions are inclined to be influenced by positive surprises instead of negative ones.

## **6. Conclusion**

Online businesses take a large portion of the whole economy. The incentives underlying the online business creation, however, is less understood. In this study, we leverage a unique data set from the largest C2C platform in the world to explore whether and how social interactions motivate (excess) online entry.

We find people are more likely to become online entrepreneurs when observing an upgraded store in the neighborhood. The effect is statistically and economically significant and is robust to a range of rigorous model specifications. The impact of peer's performance increases with the level of saliency and decreases with distance. Importantly, additional tests suggest the entry decision is suboptimal as the entrants motivated by upgrade events

underperform other entrants in terms of lower sales and a high probability of exit in the aftermarket.

The results are mostly consistent with the salience theories of choice, which argue that people tend to overweight salient payoffs in decision making. And we rule out alternative explanations such as regional economic development and rational learning. Our study contributes to the literature on the drivers for entrepreneurship in the context of online business. By exploiting the platform's nuanced measures, we add to the debate on the peer effect in entrepreneurial decisions.

## References

- Andersen, S. and Nielsen, K.M., 2012. Ability or finances as constraints on entrepreneurship? evidence from survival rates in a natural experiment. *The Review of Financial Studies*, 25(12), pp.3684-3710.
- Bailey, M., Cao, R., Kuchler, T. and Stroebel, J., 2018. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6), pp.2224-2276.
- Banerjee, A.V., Chandrasekhar, A.G., Duflo, E. and Jackson, M.O., 2018. Changes in social network structure in response to exposure to formal credit markets. *Available at SSRN 3245656*.
- Bernardo, A.E. and Welch, I., 2001. On the evolution of overconfidence and entrepreneurs. *Journal of Economics & Management Strategy*, 10(3), pp.301-330.
- Bernstein, S., Giroud, X. and Townsend, R.R., 2016. The impact of venture capital monitoring. *The Journal of Finance*, 71(4), pp.1591-1622.
- Bordalo, P., Gennaioli, N. and Shleifer, A., 2012. Saliency theory of choice under risk. *The Quarterly Journal of Economics*, 127(3), pp.1243-1285.
- Bordalo, P., Gennaioli, N. and Shleifer, A., 2013. Saliency and consumer choice. *Journal of Political Economy*, 121(5), pp.803-843.
- Bordalo, P., Gennaioli, N. and Shleifer, A., 2013. Saliency and asset prices. *American Economic Review*, 103(3), pp.623-28.
- Breza, E., 2012. Peer effects and loan repayment: Evidence from the krishna default crisis. *Job Market Paper MIT*, 1.
- Breza, E. and Chandrasekhar, A.G., 2019. Social networks, reputation, and commitment: evidence from a savings monitors experiment. *Econometrica*, 87(1), pp.175-216.
- Busse, M.R., Pope, D.G., Pope, J.C. and Silva-Risso, J., 2015. The psychological effect of weather on car purchases. *The Quarterly Journal of Economics*, 130(1), pp.371-414.
- Cavallo, A., 2017. Are online and offline prices similar? Evidence from large multi-channel retailers. *American Economic Review*, 107(1), pp.283-303.
- Cavallo, A., 2018. Scraped data and sticky prices. *Review of Economics and Statistics*, 100(1), pp.105-119.
- Couture, V., Faber, B., Gu, Y. and Liu, L., 2020. *Connecting the countryside via E-commerce: evidence from China*. Forthcoming in *American Economic Review: Insights*.
- Chandrasekhar, A.G., Kinnan, C. and Larreguy, H., 2018. Social networks as contract enforcement: Evidence from a lab experiment in the field. *American Economic Journal: Applied Economics*, 10(4), pp.43-78.
- Chang, T.Y., Huang, W. and Wang, Y., 2018. Something in the air: Pollution and the demand for health insurance. *The Review of Economic Studies*, 85(3), pp.1609-1634.
- Chetty, R., Looney, A. and Kroft, K., 2009. Saliency and taxation: Theory and evidence. *American Economic Review*, 99(4), pp.1145-77.
- Choi, D., Lou, D. and Mukherjee, A., 2016. Superstar Firms and College Major Choice. Working



Paper.

- Cohen, P., Hahn, R., Hall, J., Levitt, S. and Metcalfe, R., 2016. *Using big data to estimate consumer surplus: The case of uber* (No. w22627). National Bureau of Economic Research.
- Dessaint, O. and Matray, A., 2017. Do managers overreact to salient risks? Evidence from hurricane strikes. *Journal of Financial Economics*, 126(1), pp.97-121.
- Dohmen, T., Falk, A., Huffman, D. and Sunde, U., 2012. The intergenerational transmission of risk and trust attitudes. *The Review of Economic Studies*, 79(2), pp.645-677.
- Dolfen, P., Einav, L., Klenow, P.J., Klopach, B., Levin, J.D., Levin, L. and Best, W., 2019. *Assessing the gains from e-commerce* (No. w25610). National Bureau of Economic Research.
- Evans, D.S. and Jovanovic, B., 1989. An estimated model of entrepreneurial choice under liquidity constraints. *Journal of Political Economy*, 97(4), pp.808-827.
- Fan, J., Tang, L., Zhu, W. and Zou, B., 2018. The Alibaba effect: Spatial consumption inequality and the welfare gains from e-commerce. *Journal of International Economics*, 114, pp.203-220.
- Folkes, V.S., 1988. The availability heuristic and perceived risk. *Journal of Consumer Research*, 15(1), pp.13-23.
- Gabaix, X., 2016. Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30(1), pp.185-206.
- Gabaix, X., Lasry, J.M., Lions, P.L. and Moll, B., 2016. The dynamics of inequality. *Econometrica*, 84(6), pp.2071-2111.
- Gennaioli, N. and Shleifer, A., 2010. What comes to mind. *The Quarterly Journal of Economics*, 125(4), pp.1399-1433.
- Glaeser, E.L., Kerr, W.R. and Ponzetto, G.A., 2010. Clusters of entrepreneurship. *Journal of Urban Economics*, 67(1), pp.150-168.
- Goolsbee, A.D. and Klenow, P.J., 2018, May. Internet rising, prices falling: Measuring inflation in a world of e-commerce. In *AEA Papers and Proceedings* (Vol. 108, pp. 488-92).
- Hau, H., Huang, Y., Shan, S., and Sheng, Z., 2019. TechFin at Ant Financial: Credit Market Completion and its Growth Effect. Working Paper.
- Henderson, J.V., Storeygard, A. and Weil, D.N., 2012. Measuring economic growth from outer space. *American Economic Review*, 102(2), pp.994-1028.
- Hong, H., Kubik, J.D. and Stein, J.C., 2005. Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *The Journal of Finance*, 60(6), pp.2801-2824.
- Huang, Y., Chen, L., Sheng, Z., and Wei, L., 2019. FinTech Credit and Service Quality: Evidence from Alibaba Online Merchants. Working Paper.
- Huang, Z., Li, L., Ma, G. and Xu, L.C., 2017. Hayek, local information, and commanding heights: Decentralizing state-owned enterprises in China. *American Economic Review*, 107(8), pp.2455-78.
- Hurst, E., Li, G. and Pugsley, B., 2014. Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of Economics and Statistics*, 96(1), pp.19-33.

- Hurst, E. and Lusardi, A., 2004. Liquidity constraints, household wealth, and entrepreneurship. *Journal of Political Economy*, 112(2), pp.319-347.
- Hvide, H.K. and Møen, J., 2010. Lean and hungry or fat and content? Entrepreneurs' wealth and start-up performance. *Management Science*, 56(8), pp.1242-1258.
- Iyer, R. and Puri, M., 2012. Understanding bank runs: The importance of depositor-bank relationships and networks. *American Economic Review*, 102(4), pp.1414-45.
- Jannati, S., Kumar, A., Niessen-Ruenzi, A. and Wolfers, J., 2020. In-group bias in financial markets. *Available at SSRN 2884218*.
- Klein, T.J., Lambertz, C. and Stahl, K.O., 2016. Market transparency, adverse selection, and moral hazard. *Journal of Political Economy*, 124(6), pp.1677-1713.
- Kőszegi, B. and Szeidl, A., 2013. A model of focusing in economic choice. *The Quarterly Journal of Economics*, 128(1), pp.53-104.
- Li, M. and Goetz, S.J., 2019. The Intergenerational Persistence of Self-Employment across China's Planned Economy Era. *Journal of Labor Economics*, 37(4), pp.1301-1330.
- Lindquist, M.J., Sol, J. and Van Praag, M., 2015. Why do entrepreneurial parents have entrepreneurial children?. *Journal of Labor Economics*, 33(2), pp.269-296.
- Luca, M., 2016. Reviews, reputation, and revenue: The case of Yelp. com. *Com (March 15, 2016)*. *Harvard Business School NOM Unit Working Paper*, (12-016).
- Mayzlin, D., Dover, Y. and Chevalier, J., 2014. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8), pp.2421-55.
- Michelacci, C. and Silva, O., 2007. Why so many local entrepreneurs?. *The Review of Economics and Statistics*, 89(4), pp.615-633.
- Morris, S., 2000. Contagion. *The Review of Economic Studies*, 67(1), pp.57-78.
- Nicolaou, N., Shane, S., Cherkas, L., Hunkin, J. and Spector, T.D., 2008. Is the tendency to engage in entrepreneurship genetic?. *Management Science*, 54(1), pp.167-179.
- Pistaferri, L., Guiso, L. and Schivardi, F., 2021. Learning Entrepreneurship From Other Entrepreneurs?. *Journal of Labor Economics*, 39(1), pp.135-191.
- Rysman, M., 2009. The economics of two-sided markets. *Journal of Economic Perspectives*, 23(3), pp.125-43.
- Seamans, R. and Zhu, F., 2014. Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers. *Management Science*, 60(2), pp.476-493.

**Table 1. Store Rating System on Taobao Marketplace**

Notes: This table presents a summary of the store rating system. Panel A shows the number of stores across different rating categories. We also present the average monthly sales and average monthly page views for stores with different ratings. Panel A is based on the full sample of stores on Taobao.com. Panel B shows a summary of the events of rating changes. The first column shows the change in the rating level from month  $t-1$  to month  $t$ . Stores with the upgrade events constitute our baseline sample.

**Panel A. Summary of Performance across Store Rating**

Rating	Rating Name	Number of Stores	Average Monthly Sales	Average Monthly Page Views
0	no rating	5545977	436	116
1	1 heart	1136325	660	278
2	2 hearts	1880862	1317	419
3	3 hearts	1278553	2135	684
4	4 hearts	784573	3182	942
5	5 hearts	740805	4289	1243
6	1 crystal	1031379	6037	1627
7	2 crystals	857647	10115	2501
8	3 crystals	689709	16260	3831
9	4 crystals	641062	28032	6222
10	5 crystals	292391	44315	10466
11	1 crown	185064	66502	16696
12	2 crowns	117906	112366	31684
13	3 crowns	37038	193000	61692
14	4 crowns	16031	307926	117786
15	5 crowns	7211	586484	276163
16	1 gold crown	1660	1156974	639958
17	2 gold crowns	548	2309458	1529410
18	3 gold crowns	214	4402325	3292865
19	4 gold crowns	48	6465623	4211521
20	5 gold crowns	24	8180417	10129738

**Table 1. Store Rating System on Taobao Marketplace (Continued)****Panel B. Summary of Rating Changes**

Rating (t) - Rating (t-1)	Freq.	Percent
-11	5	0.00
-10	8	0.00
-9	19	0.00
-8	34	0.00
-7	59	0.00
-6	126	0.00
-5	158	0.00
-4	278	0.00
-3	252	0.00
-2	466	0.00
-1	1,306	0.01
0	18,183,384	96.19
1	658,725	3.48
2	45,268	0.24
3	9,308	0.05
4	2,821	0.01
5	1,100	0.01
6	631	0.00
7	248	0.00
8	124	0.00
9	65	0.00
10	29	0.00
11	29	0.00
12	1	0.00

**Table 2. Summary Statistics of the Baseline Sample**

Notes: This table presents a descriptive summary of the baseline sample. The sample is at the store-month level and consists of 133,583 unique online stores located in China's rural areas that experience rating updates in the sample period from August 2014 to August 2016. The panel sample contains 3,052,391 store-month observations. The sample construction process is detailed in Appendix 2. Appendix 4 shows the variable definitions.

	N	Mean	SD	Median
<b>Entry Variables</b>				
Entry [<0.5km]	3048315	1.298	4.542	0.000
Entry [0.1km]	3048315	0.622	2.744	0.000
Entry [0.1km - 0.5km]	3048315	0.676	2.656	0.000
Entry [0.5km - 1km]	3048315	1.069	2.910	0.000
Entry [1km - 2km]	3048315	2.370	5.213	0.000
<b>Treatment Variables</b>				
Upgrade	3048315	0.044	0.204	0.000
Upgrade [1 level]	3048315	0.040	0.196	0.000
Upgrade [2+ level]	3048315	0.004	0.062	0.000
<b>Control Variables</b>				
County Store Number	3048315	8077.838	9568.852	5085.000
County Sales	3048315	10.318	7.819	8.580
County Light	3048315	10.698	1.080	10.717
<b>Performance</b>				
Entrant Exit Share (3m)	1014379	0.267	0.381	0.000
Entrant Exit Share (6m)	677607	0.513	0.442	0.500

**Table 3. Baseline Estimation: Rating Upgrades and Online Entrepreneurship**

Notes: This table presents the baseline estimation of the spillover effect of rating upgrade events on online stores' entry in the same neighborhood. The sample is at the store-month level and consists of stores that experience rating upgrades in the sample period. The dependent variable, *Entry* [ $<0.5km$ ], is the number of entrants in the 0.5 km radius of the given store  $i$  in month  $t$ . Independent variable, *Upgrade*, is an indicator set to one if the given store  $i$  is upgraded in month  $t$ . We include different combinations of store fixed effects, month fixed effects, and city-by-month fixed effects. Control variables include the number of active stores in a county, the average monthly sales (in thousand) for active stores in a county, and the log value of night light in a county. The  $t$ -statistics are reported in the brackets and are based on standard errors clustered at both the store level and month level. \*\*\*, \*\*, and \* indicate significance at 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)
	Entry [ $<0.5km$ ], t			
Upgrade, t	0.099** [2.68]	0.102*** [2.81]	0.046*** [3.49]	0.046*** [3.53]
County Store Number		-0.001*** [-3.50]		-0.001*** [-3.31]
County Sales		0.017*** [3.20]		0.005 [1.18]
County Light		0.251** [2.32]		0.288* [2.00]
N	3048315	3048315	3048315	3048315
R2	0.717	0.725	0.791	0.792
Store FE	X	X	X	X
Month FE	X	X		
City-Month FE			X	X
Beta/Mean (%)	7.63	7.86	3.54	3.54

**Table 4. The Dynamic Effect**

Notes: This table presents the dynamic effects of rating upgrades by including lagged variables for upgrading events. We include different combinations of store fixed effects, month fixed effects, and city-by-month fixed effects. Control variables include the number of active stores in a county, the average monthly sales (in thousand) for active stores in a county, and the log value of night light in a county. The *t*-statistics are reported in the brackets and are based on standard errors clustered at both the store level and month level.

	(1)	(2)
	Entry [ $<0.5\text{km}$ ], <i>t</i>	
Upgrade, <i>t</i>	0.043***	0.044***
	[3.85]	[3.87]
Upgrade, <i>t</i> -1	-0.007	-0.006
	[-0.50]	[-0.43]
Upgrade, <i>t</i> -2	-0.025	-0.024
	[-1.42]	[-1.39]
Upgrade, <i>t</i> -3	-0.014	-0.013
	[-1.15]	[-1.09]
Upgrade, <i>t</i> -4	-0.006	-0.005
	[-0.64]	[-0.56]
Upgrade, <i>t</i> -5	0.003	0.004
	[0.34]	[0.45]
N	3048315	3048315
R2	0.791	0.792
Controls		X
Store FE	X	X
City-Month FE	X	X

**Table 5. The Dosage of Salience**

Notes: This table presents the dosage effect of upgrading events. Panel A shows the results by the level of upgrades. Panel B presents the results for the number of entrants by different distances to the given upgraded store. We include different combinations of store fixed effects, month fixed effects, and city-by-month fixed effects. Control variables include the number of active stores in a county, the average monthly sales (in thousand) for active stores in a county, and the log value of night light in a county. The *t*-statistics are reported in the brackets and are based on standard errors clustered at both the store level and month level.

**Panel A. Upgrade Level**

	(1)	(2)
	Entry [ $<0.5\text{km}$ ]	Beta/Mean (%)
Upgrade [1 level]	0.039*** [3.38]	3.00
Upgrade [2+ level]	0.119*** [3.42]	9.17
N	3048315	
R2	0.792	
Controls	X	
Store FE	X	
City-Month FE	X	

**Panel B. Distance**

	(1)	(2)	(3)	(4)
	Entry [ $<0.1\text{km}$ ]	Entry [ $0.1\text{km}-0.5\text{km}$ ]	Entry [ $0.5\text{km}-1\text{km}$ ]	Entry [ $1\text{km}-2\text{km}$ ]
Upgrade	0.022*** [4.06]	0.024*** [2.89]	0.020*** [3.03]	0.012 [1.39]
N	3048315	3048315	3048315	3048315
R2	0.768	0.750	0.698	0.775
Controls	X	X	X	X
Store FE	X	X	X	X
City-Month FE	X	X	X	X
Mean of Dep. Var.	0.622	0.676	1.069	2.370
Beta/Mean (%)	3.54	3.55	1.87	0.5



**Table 6. Types of Entrants**

Notes: This table presents the number of entrants by store characteristics. Panel A shows the number of entrants considering whether the product category is the same as the incumbent store. Panel B presents the number of entrants by the age and education of the owner. "Non-college" labels the entrants' owners do not hold a college degree, while "college" indicates the owners do have a college degree. In Panel C, we show the results by the gender and marital status of the owner. Panel D presents the effect on each age-adjusted income group. We categorize the newly registered stores into five groups according to the estimated monthly income in each age group, as defined in Figure 6. Income Group (1) indicates the income group with the lowest income level. And Income Group (5) indicates the income group with the highest income level. We include store fixed effects and city-by-month fixed effects. Control variables include the number of active stores in a county, the average monthly sales (in thousand) for active stores in a county, and the log value of night light in a county. The *t*-statistics are reported in the brackets and are based on standard errors clustered at both the store level and month level.

**Panel A. Product Category**

	(1)	(2)
	Entry [ $<0.5\text{km}$ ]	
	Same Product	Different Product
Upgrade	0.040**	0.006
	[2.36]	[0.88]
N	3048315	3048315
R2	0.805	0.675
Controls	X	X
Store FE	X	X
City-Month FE	X	X
H0: $\beta(1)=\beta(2)$	p<0.01	
Mean of Dep. Var.	0.541	0.793
Beta/Mean (%)	7.39	0.79

**Panel B. Age and Education**

	(1)	(2)	(3)	(4)
	Entry [ $<0.5\text{km}$ ]			
	Age [ $\leq 30$ ]	Age [ $>30$ ]	Non-College	College
Upgrade	0.020***	0.026***	0.046***	0.000
	[3.58]	[3.13]	[3.42]	[0.20]
N	3048315	3048315	3048315	3048315
R2	0.737	0.752	0.792	0.467
Controls	X	X	X	X
Store FE	X	X	X	X
City-Month FE	X	X	X	X
H0: $\beta(1)=\beta(2)$	p=0.01			
H0: $\beta(3)=\beta(4)$			p<0.01	
Mean of Dep. Var.	0.760	0.488	1.128	0.170
Beta/Mean (%)	2.63	5.32	4.07	0.00

**Table 6. Types of Entrants (Continued)**

**Panel C. Gender and Marital Status**

	(1)	(2)	(3)	(4)
	Entry [<0.5km]			
	Male	Female	Married	Not Married
Upgrade	0.021***	0.025***	0.016***	0.021***
	[3.60]	[3.23]	[3.16]	[3.68]
N	3048315	3048315	3048315	3048315
R2	0.745	0.736	0.685	0.748
Controls	X	X	X	X
Store FE	X	X	X	X
City-Month FE	X	X	X	X
H0: beta(1)=beta(2)	p=0.08			
H0: beta(3)=beta(4)			p=0.02	
Mean of Dep. Var.	0.698	0.550	0.532	0.467
Beta/Mean (%)	3.01	4.55	3.01	4.50

**Panel D. Income Groups**

	(1)	(2)	(3)	(4)	(5)
	Entry [<0.5km]				
	Income Group (1)	Income Group (2)	Income Group (3)	Income Group (4)	Income Group (5)
Upgrade	0.018***	0.009***	0.008***	0.005*	0.001
	[2.94]	[4.32]	[3.28]	[1.87]	[0.86]
N	3048105	3048105	3048105	3048105	3048105
R2	0.771	0.675	0.590	0.525	0.389
Controls	X	X	X	X	X
Store FE	X	X	X	X	X
City-Month FE	X	X	X	X	X
Mean of Dep. Var.	0.297	0.246	0.229	0.207	0.151
Beta/Mean (%)	6.06	3.66	3.49	2.42	0.66

**Table 7. Performance of Entrants**

Notes: This table presents the aftermarket performance of the entrants around the upgraded stores. Average Entrant Sales (3m) is the average monthly sales (in thousand) in the 3-month period after the registration for stores that registered in 0.5 km radius of the upgraded sample store  $i$  in month  $t$ . Average Entrant Sales (6m) is the average monthly sales (in thousand) in the 6-month period after the registration for stores that registered in 0.5 km radius of the upgraded sample store  $i$  in month  $t$ . Control variables include the number of active stores in a county, the average monthly sales (in thousand) for active stores in a county, and the log value of night light in a county, value of sales and the sales growth. The  $t$ -statistics are reported in the brackets and are based on standard errors clustered at both the store level and month level.

	(1)	(2)	(3)	(4)
	Average Entrant Sales (3m)		Average Entrant Sales (6m)	
Upgrade	-0.378***	-0.390***	-0.800**	-0.797**
	[-2.92]	[-3.02]	[-2.55]	[-2.55]
N	1014705	1014705	677820	677820
R2	0.185	0.185	0.210	0.210
Controls		X		X
Store FE	X	X	X	X
City-Month FE	X	X	X	X
Beta/Mean (%)	7.72	7.97	8.27	8.24

**Table 8. Exit Decisions of Entrants**

Notes: This table presents the exit decisions of the entrants around the upgraded stores. Entrant Exit Share (3m) is the percentage of the entrant that exit in the 3-month period after the registration. Entrant Exit Share (6m) is the percentage of the entrant that exit in the 6-month period after the registration. We include store fixed effects and city-by-month fixed effects. Control variables include the number of active stores in a county, the average monthly sales (in thousand) for active stores in a county, and the log value of night light in a county. value of sales and the sales growth. The *t*-statistics are reported in the brackets and are based on standard errors clustered at both the store level and month level.

	(1)	(2)	(3)	(4)
	Entrant Exit Share (3m)		Entrant Exit Share (6m)	
Upgrade	0.004***	0.004***	0.009***	0.009***
	[2.92]	[2.90]	[3.49]	[3.49]
N	1014705	1014705	677820	677820
R2	0.488	0.488	0.627	0.627
Controls		X		X
Store FE	X	X	X	X
City-Month FE	X	X	X	X
Beta/Mean (%)	1.50	1.50	1.75	1.75

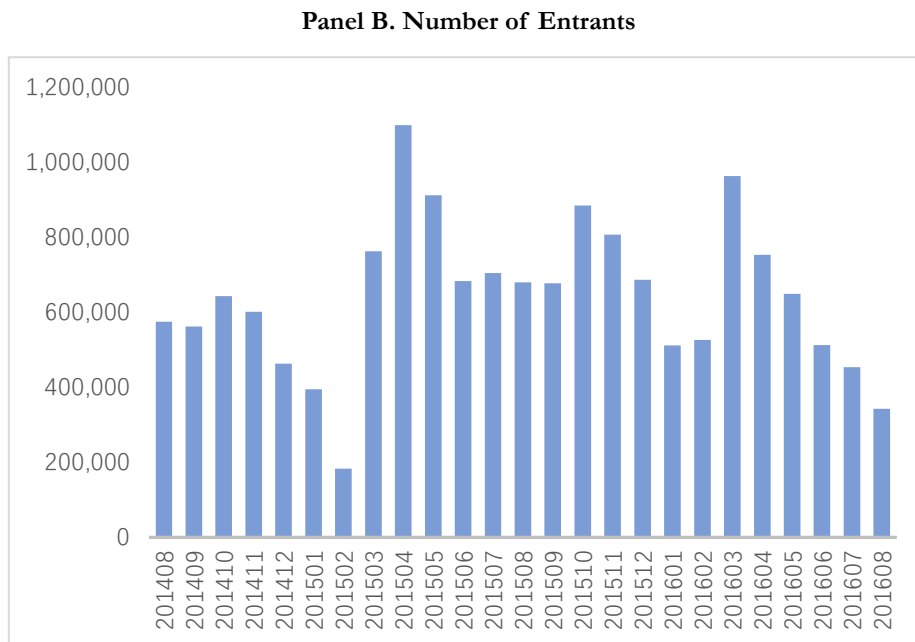
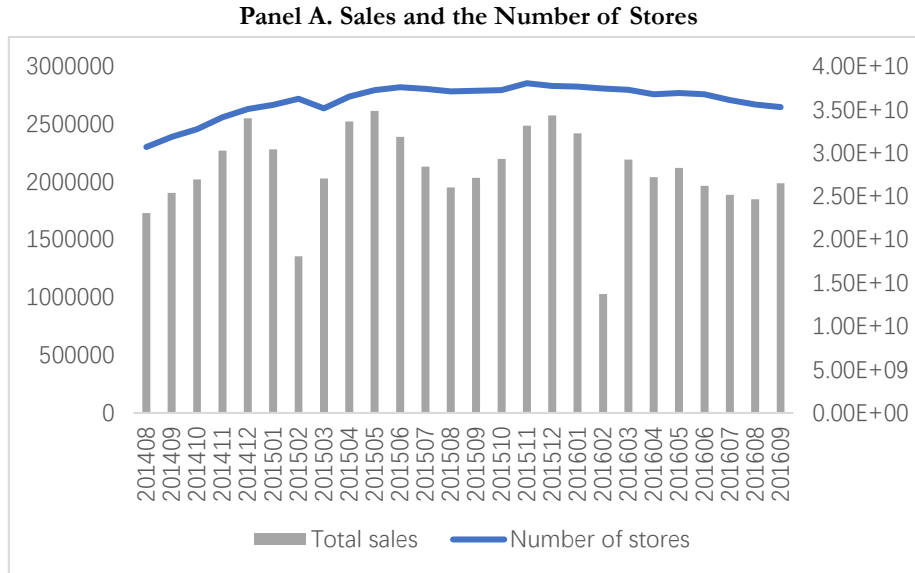
**Table 9. Entrant Performance and Product Market Performance**

Notes: This table presents results regarding the aftermarket performance of the entrants and product market performance. We use the aggregated sales growth in a given product category in a city to measure the product market performance. The sample is at the store-product-month level. Dependant variable *Average Entrant Sales (3m)* is the average monthly sales in product category  $k$  in the 3 month period after the registration for stores that registered in 0.5 km radius of the upgraded sample store  $i$  in month  $t$ . *Product Sales Growth* is the aggregated sales growth for product  $k$  from month  $t-1$  to month  $t$  for stores located in city  $c$ . *Low Product Sales Growth* is an indicator variable set to one if the aggregated sales growth for product  $k$  from month  $t-1$  to month  $t$  for stores located in city  $c$  is below median and zero otherwise. We include store fixed effects, city-by-month fixed effects, and product-by-month fixed effects. Control variables include the number of active stores in a county, the average monthly sales (in thousand) for active stores in a county, and the log value of night light in a county. The  $t$ -statistics are reported in the brackets and are based on standard errors clustered at both the store level and month level.

	(1)	(2)	(3)	(4)
	Average Entrant Sales (3m)			
Upgrade	-0.222**	-0.227***	0.147	0.143
	[-2.77]	[-2.85]	[1.17]	[1.12]
Upgrade × Product Sales Growth	0.406**	0.403**		
	[2.63]	[2.66]		
Upgrade × Low Product Sales Growth			-1.178**	-1.182**
			[-2.40]	[-2.40]
N	2058898	2058898	2058898	2058898
R2	0.156	0.156	0.156	0.156
Controls		X		X
Store FE	X	X	X	X
City-Month FE	X	X	X	X
Product-Month FE	X	X	X	X

### Figure 1. Sales and Entrants over Time

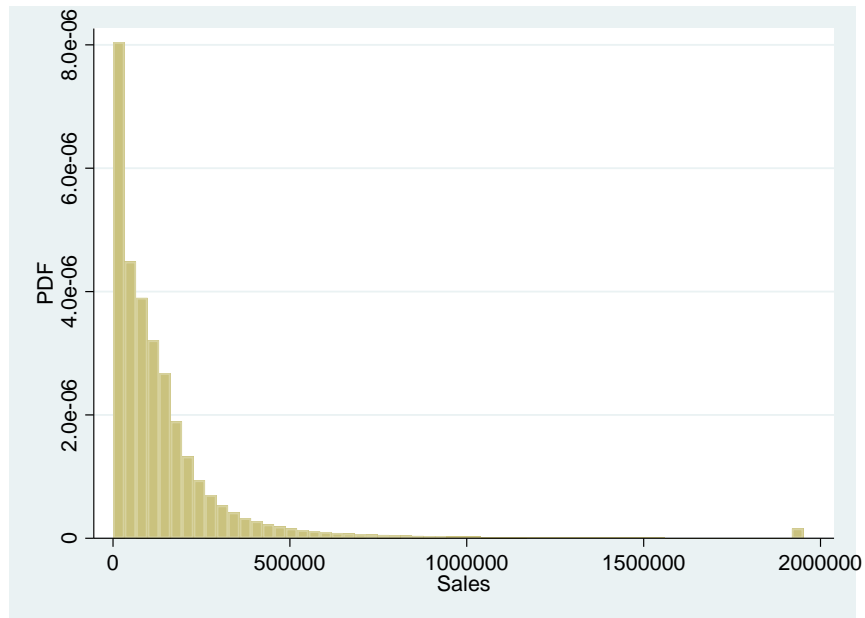
Notes: This figure presents the number and the performance of the online store on Taobao.com. The figure is based on the full sample of Taobao stores labeled as "normal" status. Panel A shows the total sales (right scale) and the number of stores (left scale) over time. The number of entrants in the sample period is presented in Panel B.



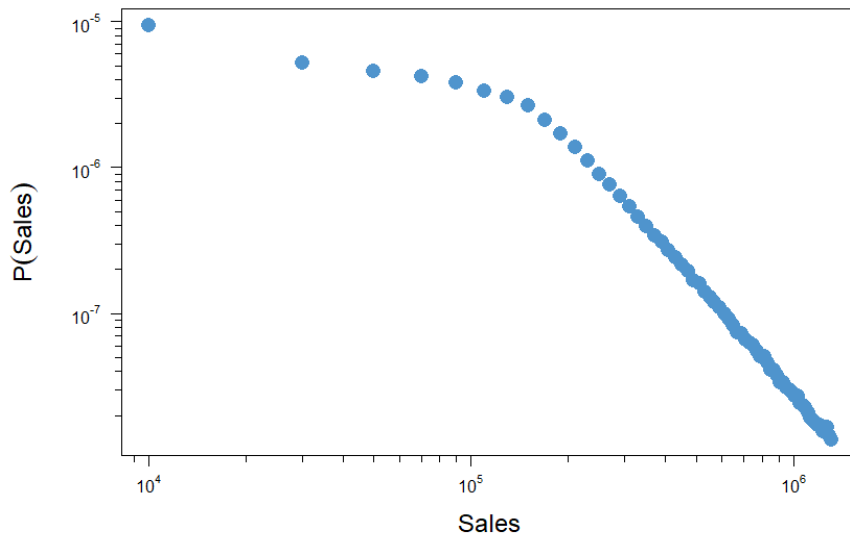
## Figure 2. Sales Distribution

Notes: This figure shows the distribution of sales across online stores on Taobao.com. Panel A presents the probability density function of the monthly sales across stores. We plot the density function on the log-log scale in Panel B.

Panel A. Probability Density Function



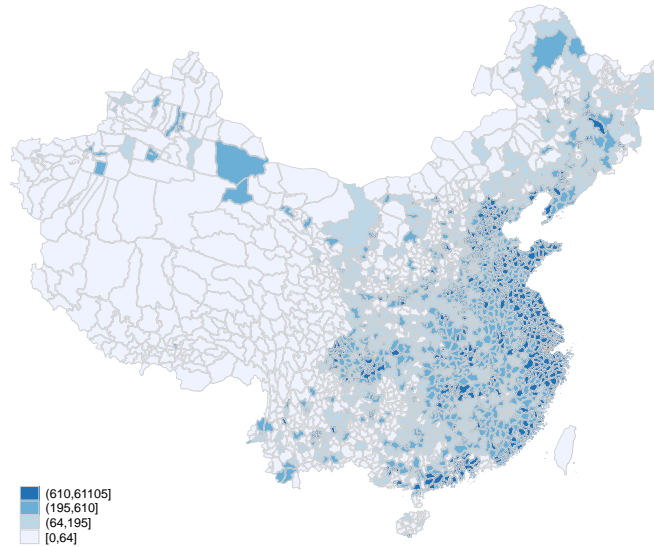
Panel B. Log-log Scale Plot



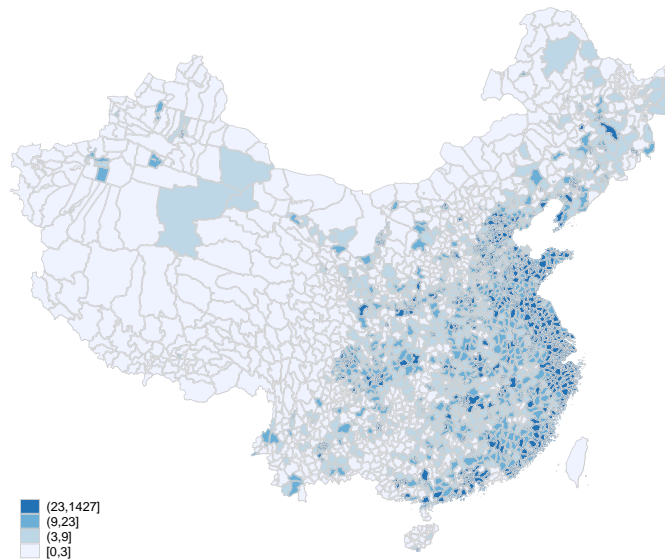
### Figure 3. Spatial Distribution of Online Stores

Notes: This figure shows the geographical distribution of online stores across Chinese counties. Panel A presents the number of active stores across counties in our sample period. Panel B shows the number of entrants across counties in our sample period.

**Panel A. Number of Stores across Counties**



**Panel B. Number of Entrants across Counties**





## Figure 4. Store Rating System

Notes: Panel A illustrates the store rating system on Taobao.com. The first column presents the internal score for rating categories. Columns 2-3 present the symbols and names of the store rating, respectively. The last column shows the value we assigned for each rating category. We provide a store example in Panel B. The rating is publicly available to all customers.

Panel A. The Rating System

所积分数	等级图标	信誉等级	Rating
4分-10分 <b>Score</b>	♥ <b>Symbol</b>	一星 <b>Name</b>	1
11分-40分	♥♥	二星	2
41分-90分	♥♥♥	三星	3
91分-150分	♥♥♥♥	四星	4
151分-250分	♥♥♥♥♥	五星	5
251分-500分	👑	一钻	6
500分-1000分	👑👑	二钻	7
1001分-2000分	👑👑👑	三钻	8
2001分-5000分	👑👑👑👑	四钻	9
5001分-10000分	👑👑👑👑👑	五钻	10
10001分-20000分	👑👑	一皇冠	11
20001分-50000分	👑👑👑	二皇冠	12
50001分-100000分	👑👑👑👑	三皇冠	13
100001分-200000分	👑👑👑👑👑	四皇冠	14
200001分-500000分	👑👑👑👑👑👑	五皇冠	15
500001分-1000000分	👑	一金冠	16
1000001分-2000000分	👑👑	二金冠	17
2000001分-5000000分	👑👑👑	三金冠	18
5000001分-10000000分	👑👑👑👑	四金冠	19
10000001分以上	👑👑👑👑👑	五金冠	20

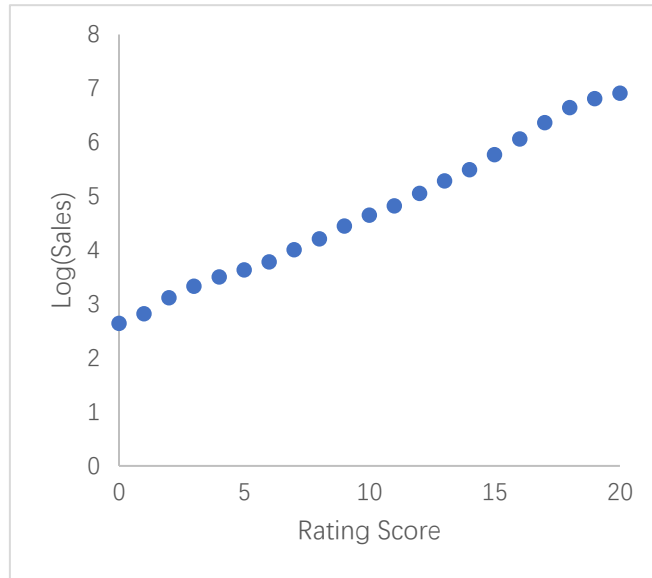
Panel B. An Example



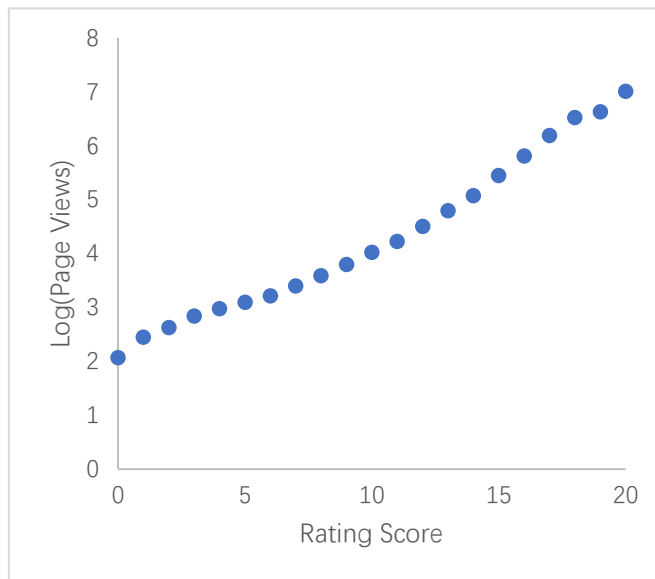
### Figure 5. Store Rating and Performance

Notes: This figure shows the relationship between store rating and performance. Panel A plots the average log value of monthly sales across store ratings. Panel B shows the average monthly page views for stores in different rating bins.

**Panel A. Store Rating and Sales**

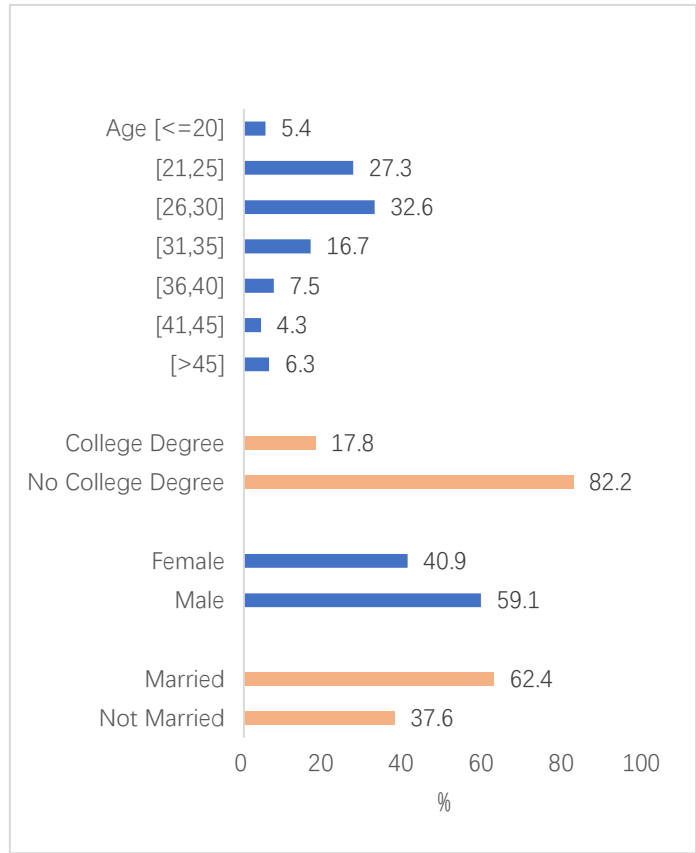


**Panel B. Store Rating and Page Views**



### Figure 6. Entrants Breakdown

Notes: This figure shows the composition of entrants by the owner's characteristics.



## Appendix 1. Registration of an Online Store on Taobao.com

Notes: This appendix presents the process of registering an online store. The initial cost of store registration is RMB 1000 (about USD 144). This registration fee is refundable when the store is dissolved. The owner is required to pay a monthly technical support fee of RMB 50 (about USD 7.2). Documentations required in registration include personal ID, certified phone number, and bank account. The process is illustrated below.



## Appendix 1. Registration of an Online Store on Taobao.com (Continued)

开店条件检测

恭喜您满足开店条件，请继续完成下面的开店认证后才能创建店铺！认证帮助

申请开店认证

选择所在地： 中国大陆  香港/澳门  台湾  海外

认证：

状态	认证名称	操作
未通过	支付宝实名认证	重新认证
未开始	淘宝开店认证	立即申请

**Under Review**

设置支付方式 你正在进行身份验证，请输入一张本人银行卡验证并绑定。

承诺你的用卡安全

- 信息加密 确保隐私不泄露
- 风险识别 陌生环境要安检
- 资金保障 众安保险全额赔

银行卡号

持卡人姓名

证件

手机号码

[《快捷签约协议》](#)

**Bank Account**

**Personal ID**

**Phone Number**

## **Appendix 2. Sample Construction**

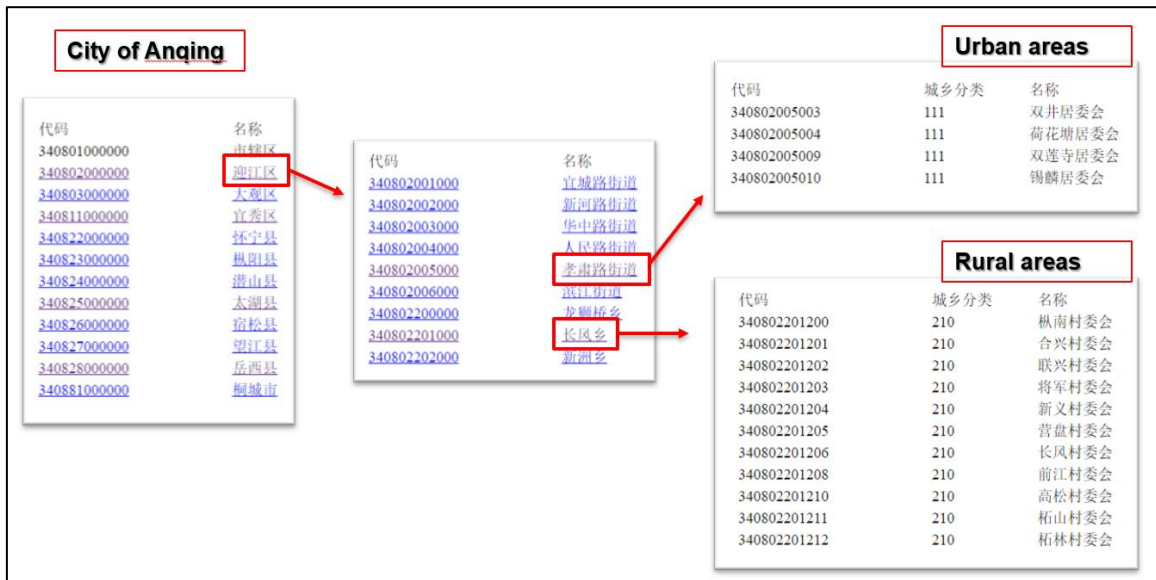
Notes: The baseline sample is constructed as follows.

1. The status of the store is "normal."
2. Restrict to stores with location information, and the stores are located in rural areas.
3. Restrict to stores that experience rating upgrading during the sample period.
4. Final sample contains 3,052,391 store-month observations with 133,583 distinct stores.

### Appendix 3. Rural vs. Urban Areas

Notes: This table illustrates the identification of online stores located in rural areas and urban areas. Panel A shows the administrative divisions of China and how we identify urban areas vs. rural areas. Panel B shows the differences in performance between firms located in urban areas and ones in rural areas. Panel C presents the top product categories in terms of the number of stores located in urban areas vs. rural areas.

#### Panel A. Example: Identifying Rural Areas



#### Panel B. Rural vs. Urban: Performance

	Number of Stores	Percentage of Stores	Average Monthly Sales	Average Rating
Urban	2,513,396	59.5%	11337.0	3.9
Rural	1,709,807	40.5%	9229.6	3.6

### Appendix 3. Rural vs. Urban Areas (Continued)

#### Panel C. Top Product Categories

Rank	Top Urban Product Categories	Top Rural Product Categories
1	Women's clothing	Women's clothing
2	Unused items for sale	Men's clothing
3	Beauty skincare	Unused items for sale
4	Men's clothing	Children's clothing
5	Digital Accessories	Flowers
6	Children's clothing	Beauty skincare
7	Women's shoes	Women's shoes
8	Fashion Accessories	Luggage and Bags
9	Luggage and Bags	Fashion Accessories
10	Underwear	Underwear
11	Flowers	Tools
12	Cars	Cars
13	Snacks	Decoration Supplies
14	Toys	Snacks
15	Customized products	Bedding
16	Tools	Fresh fruits and vegetable
17	Antiques	Furniture
18	Toiletries	Digital Accessories
19	Diapers	Toys
20	Household necessities	Men's shoes



## Appendix 4. Variable Definitions

Variables	Definitions
Entrant [X km] $_{i,t}$	The number of newly registered stores operating in X km radius of the upgraded sample store $i$ in month $t$ . The variant, Entrant [X km] $_{i,q,t}$ , is defined at the store, month and product level.
Upgrade $_{i,t}$	Indicator variable set to one if the sample store $i$ experiences positive changes in the rating in month $t$ , and zero otherwise.
Upgrade [Z level] $_{i,t}$	Indicator variable set to one if the sample store $i$ is upgraded by Z level in rating in month $t$ , and zero otherwise. For example, Upgrade [2+ level] indicates the sample store is upgraded by two levels or above in month $t$ .
County Store Number $_{c,t}$	The number of active stores in a given county $c$ in month $t$ .
County Sales $_{c,t}$	The average monthly sales for active stores in a given county $c$ in month $t$ .
County Light $_{c,t}$	The log value of the night light in a given county $c$ in month $t$ .
Rating	The rating of store $i$ in month $t$ . The rating has 20 scales defined in Table 1.
Entrant Average Sales (3m/6m) $_{i,t}$	The average monthly sales in the 3-month/6-month period after the registration for stores that registered in 0.5 km radius of the upgraded sample store $i$ in month $t$ .
Entrant Exit Share (3m/6m) $_{i,t}$	The percentage of stores that exit in the 3-month/6-month period after the registration for stores that registered in 0.5 km radius of the upgraded sample store $i$ in month $t$ .
Product Sales Growth $_{c,k,t}$	The aggregated sales growth for product $k$ from month $t-1$ to month $t$ for stores located in city $c$ .
Low Product Sales Growth $_{c,k,t}$	Indicator variable set to one if the aggregated sales growth for product $k$ from month $t-1$ to month $t$ for stores located in city $c$ is below median and zero otherwise.

## Appendix 5. A Survey on Rural Taobao Merchants

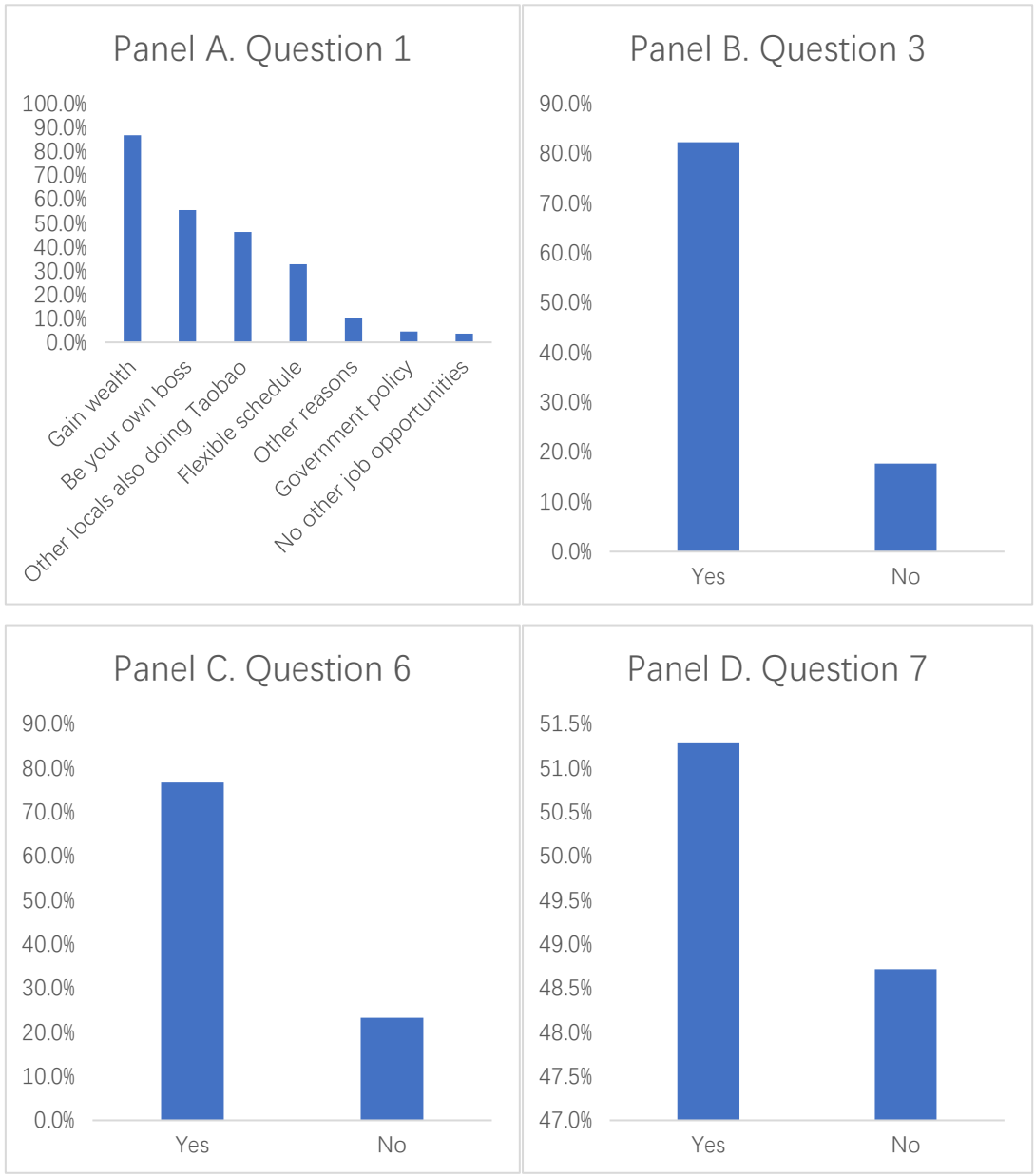
Notes: We conduct a field survey among Taobao merchants across 12 provinces and 68 cities and conduct a sample of 305 reliable observations, each corresponding to a owner of a Taobao store. The survey results regarding 8 questions are reported as follows.

---

<b>Q1. What are the main factors that triggered you to open a Taobao store?</b>	
A. Gain wealth	86.9%
B. No other job opportunities	3.6%
C. I saw other locals also doing Taobao	46.2%
E. Be your own boss	55.4%
F. Flexible schedule	32.8%
G. Government encouragement policy	4.6%
H. Other reasons	10.2%
<b>Q2. Do you often ask your friends about e-commerce issues?</b>	
A. Never	0.0%
B. Occasionally	26.9%
C. Normally	28.2%
D. Often	34.4%
E. Frequently	10.5%
<b>Q3. Do you learn from others in terms of online store's business model and business scope?</b>	
A. Yes	82.3%
B. No	17.7%
<b>Q4. Do you consider the product of nearby residents' Taobao stores as one of the considerations for in choosing the product to sell when opening a Taobao store?</b>	
A. Yes	74.4%
B. No	25.6%
<b>Q5. Do you analyze the market demand of different products when you choose the type of products to operate?</b>	
A. Yes	71.5%
B. No	28.5%
<b>Q6. Do you consider the performance of the nearby residents' Taobao stores (such as store ratings) as one of the considerations for opening a Taobao store?</b>	
A. Yes (Continue with Q7)	76.7%
B. No	23.3%
<b>Q7. If you take the performance of nearby residents' Taobao stores (such as store ratings) as one of the considerations for opening a Taobao store, do you feel that it is not in line with your expectations in your future operations?</b>	
A. Yes (Continue with Q8)	51.3%
B. No	48.7%
<b>Q8. If you take the performance of nearby residents' Taobao stores (such as store ratings) as one of the considerations for opening a Taobao store, and if you feel that your business is not in line with your expectations, will you terminate the store operation?</b>	
A. Yes	39.2%
B. No	60.8%

---

**Appendix 5. A Survey on Rural Taobao Merchants (Continued)**



## Appendix 6. Robustness Checks

Notes: This table shows the robustness checks regarding the baseline estimation in Table 3. Columns (1) and (2) report the results based on standard errors clustered at the township level, and at the both township and month level, respectively. Column (3) uses the inverse hyperbolic sine (IHS) transformation of the number of entrants as the dependant variable. Columns (4) and (5) are based on a sample at the store-product-month level. In Column (4), we consider product-month fixed effects. In Column (5), we include city-product-month fixed effects. Other model specifications follow the conventions in Table 3. The *t*-statistics are reported in the brackets and are based on standard errors clustered at both the store level and month level. \*\*\*, \*\*, and \* indicate significance at 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)	(5)
	Entry [<0.5km]				
	SE cluster: town	SE cluster: town,month	IHS Transformation	Store-Product-Month Sample	
Upgrade	0.047*** [3.54]	0.047*** [2.97]	0.004** [2.34]	0.003** [2.65]	0.004*** [2.86]
N	3048614	3048614	3048614	24270264	24270264
R2	0.792	0.792	0.729	0.073	0.493
Controls	X	X	X	X	X
Store FE	X	X	X	X	X
City-Month FE	X	X	X	X	X
Product-Month FE				X	
City-Product-Month FE					X

## Appendix 7. Downgrade Events

Notes: This table presents the effect of rating downgrades on entry decisions. We in our sample period identify 520 stores that experience rating downgrade events with precise location information and other essential variables. Downgrade is an indicator set to one for the month of rating downgrade. Other model specification follows the convention in Table 3.

	(1)	(2)
	Entry [<0.5km]	
Downgrade	0.074	0.056
	[0.58]	[0.39]
N	8997	8074
R2	0.843	0.851
Controls		X
Store FE	X	X
City-Month FE	X	X