The Salience of Entrepreneurship: Evidence from Online Business

Yi Huang, Chen Lin, Sibo Liu, Gustavo Manso*

March 16, 2022

Abstract

We explore psychological biases underlying the decision to become an entrepreneur in the online business context. Using entrepreneurs affiliated with Taobao Marketplace, the world's largest online shopping platform, as our sample, we find that people who observe the emergence of successful stores in their neighborhood are more likely to become online entrepreneurs. Relying on the Taobao store rating system and detailed geographical information for identification, we find that in rural areas of China, an increase in the online rating (upgrade event) of a store leads to a significant increase in the number of new stores within a 0.5-km radius. This effect increases with the magnitude of the upgrade event, decreases with physical distance from the focal store and is robust to a wide range of rigorous model specifications. However, such decisions to enter the market may be suboptimal, as entrants whose entrepreneurs are motivated by local upgrade events underperform relative to other entrants in terms of sales, especially during negative overall market conditions, and have a higher probability of market exit. Our results lend support to salience theories of choice and cannot be fully explained by rational learning or other regional confounders.

Keywords: entrepreneurship, peer effect, salience theory, availability heuristic
JEL Codes: L26, D91, G41

*Huang is affiliated with FISF, Fudan University and CEPR, email: huang_y@fudan.edu.cn; Lin is affiliated with the University of Hong Kong, email: chenlin1@hku.hk; Liu is affiliated with Hong Kong Baptist University, email: siboliu@hkbu.edu.hk; Manso is affiliated with University of California, Berkeley, email: manso@berkeley.edu. We acknowledge the generous support of Ant Financial Services Group, which provided 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 or its management. The analysis was undertaken in strict observance of Chinese privacy law.

1. Introduction

What induces a person to become an entrepreneur? The literature considers various factors that motivate entrepreneurial activities, such as geography (Michelacci and Silva, 2007; Glaeser et al. 2010; Pistaferri et al. 2021), parental background (Lindquist and Van Praag, 2015; Li and Goetz, 2019; Dohmen et al. 2012), and financial constraints (Evans and Jovanovic, 1989; Hurst and Lusardi, 2004; Andersen and Nielsen, 2012; Hvide and Møen, 2010). Increasingly, the literature on entrepreneurial decisions highlights the importance of underlying behavioral reasons (e.g., Bernardo and Welch, 2001). In this study, we investigate the plausible psychological bias underlying the decision to become an online entrepreneur. Specifically, we examine whether this decision is influenced by neighborhood peers and evaluate the subsequent performance of the store whose owners are induced to enter the online market.

Entrepreneurs are perceived to be motivated by either role models per se or the entrepreneurial risk culture they represent.[1] However, the empirical evidence does not clarify whether entrepreneurial decisions depend on others' behavior. Some studies document a positive effect. Using data from Sweden, Giannetti and Simonov (2009) find that individuals from highly entrepreneurial neighborhoods are more likely to become entrepreneurs. Using 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 previously. Other studies emphasize the negative effect of peers on entrepreneurship. For example, Lerner and Malmendier (2013) exploit exogenous peer assignments in the sections at Harvard Business School and find that entrepreneurship decreases as the proportion of entrepreneurial peers increases. But both Giannetti and Simonov (2009) and Lerner and Malmendier (2013) find evidence of peer effects that are consistent with rational models of learning. In contrast to these studies, we aim to go beyond the overall effect of peers on entrepreneurial decisions and investigate the effect of

[1] For example, the Hewlett-Packard Company is known to have motivated Steve Jobs, who subsequently inspired entrepreneurs around the world.

noticeably successful peers. We also explore the possibility that a behavioral explanation rather than rational theories could explain our findings.

Empirically, it is challenging to identify peer effects in entrepreneurship. The timing of a company's entry into the market is difficult to observe because of red tape and other administrative barriers. Specifically, regional variations in the time lag to register a company make it difficult to identify concurrent factors that affect market entry. As discussed herein, we mitigate this issue by examining the decision to enter the online marketplace in the context of an efficient registration process. Precise information on entry timing facilitates our analysis of intertemporal variations in the factors that motivate entrepreneurship. Additionally, it is difficult to measure the performance of new entrants due to data limitations. To overcome this challenge, we rely on high-frequency performance measures, such as the timing of exit, which are provided by the online marketplace platform. To enhance identification, we use detailed physical store location data to elucidate the impact of local peers.

Internet-based business models have exploded over the past decade. According to the United Nations Conference on Trade and Development (UNCTD), in 2018, global e-commerce sales reached $25.6 trillion, and more than 1.4 billion people shopped online. In the U.S., e-commerce sales account for 42% of GDP.[2] Recent studies evaluate the impact of online business in terms of price-setting (Cavallo, 2017; Goolsbee and Klenow, 2018), traditional business model responses (Seamans and Zhu, 2014), and the welfare effect (Dolfen et al., 2019). However, it remains unclear whether the incentives driving online entrepreneurs differ from those driving their offline counterparts. To approach this issue, we conduct 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

[2] Source: https://unctad.org/press-material/global-e-commerce-hits-256-trillion-latest-unctad-estimates

(B2C platform) together serve more than 10 million sellers and more than 454 million active buyers worldwide each year (Ali Research, 2017). We focus on Taobao Marketplace because it serves individual online merchants[3] in China, making it an ideal laboratory for studies on online entrepreneurs.

We access highly disaggregated data from Taobao Marketplace during the August 2014–August 2016 period. The dataset contains detailed information on store performance, precise store locations and store owners' demographic information. Online businesses fluctuate rapidly; for example, only half of newly registered stores survive longer than 5 months. Therefore, our 2-year sample allows us to identify the dynamics of online entrepreneurial activities. We rely on evident increases in store ratings (rating upgrades) to identify the peer effect. Specifically, we identify stores that experience rating increases during the sampling period and investigate the concomitant emergence of new stores in the surrounding neighborhoods, using different distance cutoffs. We restrict the data to stores operating in rural areas of China, where people more frequently exchange information through social networks in comparison with urban counterparts. Note that rating uptick events might not be perfectly exogeneous but associated with individual characteristics and local trends. We detail in what follows how those concerns affect our casual interpretations.

To support our empirical design, we initially collaborated with a market research company and conducted a survey of online store owners in 12 provinces and 68 cities across China. We interviewed store owners in our sample to determine the factors motivating their decisions to become online entrepreneurs. The findings suggest that people are largely affected by entrepreneurial activities in their neighborhoods. In particular, the performance of a store, as measured by store ratings, attracts attention locally and influences others' decisions to open online stores. However, we find that peer-influenced motivation tends to lead to disappointment and regret.

[3] Tmall, a counterpart of Amazon, is a platform enabling offline incorporated businesses to sell products online.

The results of baseline estimation using archival data from Taobao Marketplace suggest that increased store ratings tend to encourage people in the neighborhood to become online entrepreneurs. Specifically, an increased rating (upgrade event by at least one level in the store rating) of an online store leads to a 3.5% increase (relative to the sample mean) in the number of new stores within a 0.5 km radius. This effect increases with the magnitude of the upgrade event, decreases with distance, and is robust to a wide range of rigorous model specifications. We include region-by-year fixed effects and economic controls at the granular level to capture local trends and mitigate concerns about omitted variables. We find that the documented effect is concentrated among newly established stores selling the same product as the upgraded store and those run by owners who have a low level of education or estimated income and are unmarried.

Additional tests suggest that a peer-influenced decision to enter the market might not be optimal. First, dynamic analysis suggests that the effect exists only in the month corresponding to the upgrade event. This short-lived effect implies that salient events have only a temporary effect on attention. Second, we investigate the aftermarket performance of the entrants as the entry decision relying on local information is not necessarily behavioral. We find new entrants whose owners are motivated by their observation of these upgraded stores tend to underperform relative to their peers in terms of lower sales and have a higher probability of exit within 3 or 6 months after registration. We construct a sample that consists of only new entrants and excluding any existing stores. Therefore, we compare a cohort of entrants motivated by successful peers with other cohorts not triggered by peers. To this end, we estimate a model including regional trends and the fixed effects of the motivating stores. Third, we find that new entrants whose owners are motivated by peers tend to ignore current product market conditions, leading to inadequate economic outcomes. The findings on underperformance of the entrants suggest the effect of peer effect might deviate from explanations based on rational learning.

We find that the results of our study are most consistent with salience theories of choice as introduced by recent studies (Gennaioli and Shleifer, 2010; Bordalo, Gennaioli, and Shleifer, 2012; Kőszegi and Szeidl, 2013). Studies on salience theory suggest that people overweight information that captures their attention and do not fully consider all available information. That is, salient payoffs are overweighed, leading to risk-seeking activities when the upside payoff is salient and risk-averse activities when the downside is salient. As entrepreneurial activities are extremely risky,[4] the salience theory is consistent with the observed suboptimal nature of entry decisions, as people tend to focus on high payoff outcomes among local peers instead of considering all available information. In addition, salience theories assume that people's attention is drawn to differences rather than absolute values. This assumption is consistent with our identification of changes in store ratings as salient events. In contrast, the findings of our analysis do not support alternative explanations, such as market demand, regional development, and rational learning.

With this paper, we contribute to the literature in three aspects. First, we add to the literature on the drivers of entrepreneurship by highlighting the potential behavioral aspects of entrepreneurship in the context of online business. Whereas most studies investigate correlations between entrepreneurial decisions among peers (Sweden, Giannetti, and Simonov, 2009; Denmark, Nanda, and Sorensen, 2010; Lerner and Malmendier, 2013), we focus on how a noticeable change in a peer's performance affects an entrepreneur's decision to enter the market. The analysis on the aftermarket performance of entrants provides additional evidence consistent with behavioral motivators of the entry decision.

Second, our study lends support to previous work on salience theories. Recent studies document how salience theories can explain patterns in consumption behavior (Bordalo, Gennaioli, and Shleifer, 2013a; Busse et al., 2015), asset prices (Hirshleifer and Teoh, 2003;

[4] According to the U.S. Bureau of Labor Statistics, roughly 50% of new business establishments survive longer than 5 years. Source: https://www.bls.gov/bdm/entrepreneurship/entrepreneurship.htm

Della Vigna and Pollett, 2009; Hirshleifer, Lim, and Teoh, 2011; Bordalo, Gennaioli, and Shleifer, 2013b), and tax effects (Chetty, Looney, and Kroft, 2009). We extend analysis in this field to the realm of entrepreneurial activities. Particularly, we base our approach on intertemporal variations in store ratings, which are consistent with the assumption underlying salience theories. Our detailed measures of aftermarket performance also lend additional support to behavioral explanations of excess market entry. In particular, our research is also related to studies on how peer information affects people's choices under risk.

Third, our work is related to the emerging literature on business creation in the digital economy. Online business models usually feature two-sided markets (e.g., Seamans and Zhu, 2014), utilize user-generated data (Mayzlin et al., 2014; Cohen et al., 2016), and apply distinct business strategies (Cavallo, 2017, 2018). However, the origins and determinants of online entrepreneurship are less well understood due to data limitations. Fan et al. (2018) assess welfare gains from e-commerce using region-aggregated data from Alibaba, and two contemporaneous papers use the same dataset applied in our study to analyze how microcredit allocation shapes online store growth (Hau et al., 2019) and service quality (Huang et al., 2019). In contrast, we evaluate the behavioral factors influencing the decision to create an online business. By examining the role of social interaction in online entrepreneurship, we also contribute to the growing literature on social networks and finance (e.g., Banerjee et al., 2018).

The remainder of the paper proceeds as follows. In Section 2, we introduce the institutional background of Taobao Marketplace and describe its store rating system. We develop our hypotheses in Section 3 and describe the data and our identification strategy in Section 4. We present our findings in Section 5 and our conclusions in Section 6.

2. Institutional Background and Data

2.1 Taobao Marketplace

Our sample comprises online transactions on Taobao Marketplace, which is a subsidiary of Alibaba Group, a tech-service conglomerate listed on the New York and Hong Kong stock exchanges. It was established in May 2003 and achieved market dominance within 2 years. Taobao Marketplace mainly supports C2C retail business through its platform on which stores are established, owned, and managed by individual merchants. Approximately 97% of merchants affiliated with Alibaba platforms employ fewer than 5 workers, and 95% invest less than $4,500 (Ali Research, 2017). Most of the stores on Taobao Marketplace only operate through the online channel, especially those located in the remote areas. And merchants who run online business are not employees in commercial or public sectors and therefore meet the typical definitions of entrepreneurs in the literature (e.g., Hurst et al., 2014).

Our study is also related to the phenomenon of rapid growth and spread of "Taobao villages" across regions in China. This term is applied to villages that are significantly engaged in e-commerce, defined as a total annual e-commerce transaction volume of at least RMB 10 million and at least 100 active online shops. Taobao villages create geographical clusters of online stores, where e-commerce is the primary business of the residents. According to joint research by the World Bank and Alibaba Group, the number of Taobao villages increased from 20 in 2013 to 3,202 in 2018.[5] In 2020, more than 5,425 Taobao villages and 1,700 Taobao townships were distributed across 28 provinces, creating 8 million jobs.[6] However, Couture et al. (2020) find little evidence of income gains to rural producers and workers in areas that adopt e-commerce. Later, we will restrict our sample to stores operating in rural areas for identification purposes.

2.2 Online Store Registration

Registration of a new online store for e-commerce on Taobao Marketplace is more efficient and has a lower barrier to entry than the start-up process of a traditional offline business.

[5] Source: http://www.aliresearch.com/en/Research/Researchdetails?articleCode=21626&type=TaobaoVillages

[6] 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

Appendix 1 illustrates the registration procedures required for an online store. The entire process is conducted through the Taobao platform, which requires essential documentation such as the entrepreneur's name, personal identity card, registered phone number, and bank account information. The completed application is then reviewed to verify the information and approved within 72 hours. The initial store registration fee is RMB 1000 (approximately USD 144), and this fee is refundable when the store is dissolved. The owner is required to pay a monthly technical support fee of RMB 50 (approximately USD 7.2). Our dataset contains the precise dates of application, allowing us to identify the timing of entry decisions. In the following sections, we discuss in detail how we capture entrepreneurial decisions related to online business.

2.3 Raw Data Description

We access online transaction data on Taobao Marketplace during the August 2014–August 2016 period. Figure 1 summarizes the raw data over the sample period. As shown in Panel A, the total monthly sales on the platform are approximately RMB 25 billion. The troughs in February mark the Chinese New Year, when all business operations become inactive. The average number of active stores in our sample is 2.5 million, and this value remains stable over time. Panel B of Figure 1 depicts the trajectory of new stores registered in the sample period. Later, we impose reasonable restrictions on the sample for identification.

[Figure 1 about here]

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

[Figure 2 about here]

We also observe considerable regional variation in the number of stores on Taobao Marketplace. As shown in Figure 3, stores are clustered in the coastal areas, partly due to better access to information technology, logistics, and supply chain infrastructure. We therefore build an analysis of time-varying differences across regions.

[Figure 3 about here]

2.4 Store Rating System

Taobao Marketplace uses a store rating system based on feedback from customers to improve transparency about store quality. The literature indicates that the use of a reputation system mitigates concerns about informational asymmetries in online platforms. Such a system informs future buyers about the outcomes of a seller's past behavior, thus turning a one-shot game into a repeated game. Luca (2016) finds a causal impact of restaurant ratings posted on Yelp on consumer demand. Klein, Lambertz, and Stahl (2016) demonstrate that eBay's adoption of a feedback mechanism has improved the quality of online stores' service.

The system used by Taobao records a cumulative rating for each merchant since the start of the business. For each completed transaction, the customer is invited to select one of three feedback responses—"positive," "neutral," or "negative"—to evaluate their shopping experience. Each type of response bears a score (1, 0, and -1 points, respectively) that is added to the merchant's internal total feedback score. The internal total score is then converted to a 20-level icon rank, which is used as the cumulative store rating. We illustrate this rating system in Figure 4. The first column in Panel A shows the cutoffs applied to the internal score. The second column shows the following rating symbols: the lowest rank is marked by one heart and the highest by five gold crowns. We assign a number to each rating, as shown in column 4. The internal score in column 1 is not available to us or the public. However, the overall rating of the store is publicly available, as shown in Panel B. We exploit each store's visible rating for identification, as 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 their peers. We also exploit discrete changes in store ratings to identify visible shifts in stores' performance. As we discuss later, a store rating both is an informative measure of past performance and a predictor of a store's future sales and online traffic volume.

Studies on traditional offline business document both positive and negative peer effects (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 that is restricted to unsuccessful entrepreneurs. In contrast, we explore whether an intertemporal change in a peer's performance affects an individual's entrepreneurial decision. Below, we develop our hypotheses.

To facilitate our analysis, we propose the following simple model. A given market contains incumbent online stores in N different neighborhoods. Each store has a rating of either 1 with probability p or -1 with probability $1-p$. The market also contains two types of risk-neutral entrepreneurs: those who observe the ratings of all incumbents in the market and those who only observe the ratings of incumbents in their own neighborhoods. After observing these ratings, the entrepreneur decides whether to enter the market. For each entrant, the profits are equal to the realized rating minus the cost c to enter the market.

Entrants do not know the probability p of a positive rating. Under virtuous conditions, the probability p of a positive rating is $q > 1/2$, while under poor conditions it is $1-q$. For simplicity, the odds of virtuous or poor market conditions are assumed equal and store ratings are conditionally independent. According to Bayes' rule, after observing the ratings of incumbents, the odds of virtuous (versus poor) market conditions are calculated as follows:

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

where n is the sum of the observed incumbent ratings. Given the odds ratio $o(n)$, the profit from market entry is calculated as follows:

$$\frac{o(n)}{1+o(n)}\big(q-(1-q)\big)+\frac{1}{1+o(n)}\big((1-q)-q\big)=$$
$$=\frac{o(n)-1}{1+o(n)}(2q-1).$$

We base our assumption that entry is expected to be profitable on positive evidence ($n \geq k$) from incumbent ratings:

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

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

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

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

Entrants whose owners observe all of the ratings enter the market when $n \geq k$. Their expected profits are

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

However, owners who only observe ratings in their neighborhood must base their decisions only on that subset of signals. In this scenario, entrants in neighborhood $i \in N$ enter the market when $n_i \geq k$, where n_i is the sum of the incumbent ratings in neighborhood i. If the entrant enters a virtuous overall market, indicated by probability $P(n \geq k \mid n_i \geq k)$, then their expected profits are positive and equal to the profits of entrants who observe all ratings. However, if the entrant enters a negative overall market, indicated by probability $P(n < k \mid n_i \geq k)$, then their expected profits are negative and less than the profits of entrants who observe all ratings.

In the model, the owners of many entrants rely only on signals from their own neighborhood. On average, these entrants underperform their peers, some of whom rely on signals from the entire market. Underperformance is especially severe if new entrants are motivated by the good performance of neighborhood incumbents but the overall market performance is poor. We formulate our two hypotheses as follows.

> **H1**: People who observe upgrade events affecting stores in their neighborhood are more likely to become online entrepreneurs.

Empirical evidence supporting **H1** is consistent with a rational learning story, regional economic development, product demand, or behavioral explanations (e.g., salience theories). Specifically, a rational learning story suggests that people acquire useful information and identify business opportunities by observing the performance of stores in the neighborhood. The positive correlation between the performance of peers and the decision to enter the market may be due to common factors, such as improved infrastructure and demand shocks on specific products that confer a comparative advantage on a region. A decision to enter may also be attributable to behavioral factors that lead people to ignore the full set of information and make suboptimal choices. To further disentangle the theories that may support **H1**, we explore the economic outcome of the entry decision.

> **H2**: Entrants whose owners are motivated by upgrade events affecting stores in their immediate neighborhood underperform their peers, especially during negative overall market conditions.

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

4. Sample, Variables, and Empirical Design

We present in this section the study sample and the empirical design. In particular, we discuss the sample construction, define the term "entry," and present the model specifications.

4.1 Identification and Sample Construction

Our identification of the impact of peer performance on a new entrepreneur's decision to enter the online market is built on several essential assumptions. First, people pay attention to their neighbors. Second, observers receive an informative signal about the performance of a store. Third, after an entry decision is made, the registration process is efficient enough to allow us to link decision timing to the concurrent factors of interest. As discussed above, the efficient registration process used by Taobao Marketplace fulfills the third assumption. We focus on the first two assumptions in the following paragraphs.

Regarding the first assumption, we cannot provide direct evidence that people pay close attention to their neighbors. However, studies highlight the importance of word of mouth and social networks in shaping high-stakes decisions among individuals and firms (Robinson and Stuart, 2007; Bailey et al., 2018; Bailey et al., 2019; Kuchler and Stroebel, 2020). For example, Hong et al. (2005) demonstrate that a mutual fund manager's portfolio is correlated with the portfolios of other managers in the same city due to the sharing of information through social networks. Such networks are crucial channels in areas without well-functioning formal institutions (Chandrasekhar et al., 2018). Therefore, to reasonably ensure manifestation of the peer effect, we confine our analysis to stores located in rural areas of China, where residents rely more on cooperative social interactions.

We illustrate the procedures used to identify stores located in rural areas in Appendix 3. Specifically, the store's precise coordinates are matched with Chinese administrative divisions using geographic information system techniques. Then, the official definition of a rural area in China is applied according to the urban-rural classification code. Panel B of Appendix 3 shows that in the sample of stores with precise location information, 59.5% are located in urban areas and 40.5% are in rural areas. Compared with rural stores, urban stores have higher monthly sales and better ratings. However, as shown in Panel C, the product categories are not significantly different between urban stores and rural stores.

Regarding the second assumption, we use upgrade events as a signal to influence people's decisions regarding entrepreneurship. Unlike sales, store ratings are publicly visible and can be observed directly by other people in the neighborhood. Essentially, the rating reflects the retrospective sales and service quality of a store. However, the rating also provides information to predict the future success of a store. As shown in Figure 5, the rating of a store is monotonically correlated with its monthly sales and website page views. The number of page views represents the level of consumer attention toward the store, which translates to product demand and revenue. Therefore, we argue that an abrupt change in a store's rating provides a valid signal with respect to a peer's performance.

[Figure 5 about here]

Panel A of Table 1 further summarizes store performance by store ratings across the full sample at a monthly frequency. In the first column, we assign a value, *Rating*, to each rating category. To capture the change in performance, we calculate the difference between the values of *Rating* in month *t* and month *t-1* and summarize the results in Panel B. Upgrade events, are rare, accounting for less than 4% of the total store-month observations. This finding suggests that an upgrade event is a milestone in the life cycle of a store. Similarly, we observe very few downgrade events. We identify stores that experienced upgrade events during the sample period and retrieve the monthly observations for this group of stores. The final sample contains 3,048,614 store-month observations corresponding to 133,427 distinct rural stores that experienced upgrade events 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 a neighborhood centered around the store with the upgrade event. First, we define the registration month as the entry month. Second, we restrict the entrants to stores with at least 1 month of non-zero sales in our sample period to mitigate the concern that people may register a store but conduct no further entrepreneurial activities.[7] Third, we construct the number of entrants according to different distance cutoffs. For example, *Entrant [0.5 km]*$_{i,t}$ is the number of registered stores operating within a 0.5-km radius of the upgraded sample store *i* in month *t*.[8] *Upgrade*$_{i,t}$ is an indicator variable set to 1 if the sample store *i* experiences a positive rating change in month *t*, and 0 otherwise. If a store experiences multiple upgrade events, we only consider the first event. We consider an alternative performance measure for an online store. *Store Sales Growth*$_{c,t}$ is defined as the growth rate of the sales from month *t-1* to month *t* for a store *i*. This measure can only be defined when the sales in month *t-1* is non-zero. Table 2 shows the summary statistics of the baseline panel.

[Table 2 about here]

We include the sales level of a store as a control variable. *Store Sales* $_{c,t}$ is the sales level for store *i* in month *t*. We then construct regional time-varying variables to control for 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 of 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 based on satellite data in a given county *c* in month *t*, following Henderson et al. (2012). We capture store performance using the sales level after registration and exit probability. Specifically, *Average Entrant Sales (3m/6m)*$_{i,t,}$ is the average monthly sales within 3 and 6 months after registration among stores that registered within a 0.5-km radius of upgraded store *i* in month *t*. *Entrant Exit Share (3m/6m)*$_{i,t}$ is the percentage of stores that exit the market within the 3- or 6-month period after registration among all registered stores within a 0.5-km radius of the upgraded focal store. These variables can only be defined for store-month observations in which at least one store is registered in the neighborhood; otherwise, they are missing. All variables are defined in Appendix 4.

[7] In a table available upon request, we show that the results are qualitatively similar when this restriction is removed.

[8] In rare cases, newly registered stores and upgraded stores shared the same location. We remove these cases to mitigate concerns of a close connection between the motivated new entrepreneurs and upgraded store owners.

4.3 Model Specifications

Using store-month level sample data, we estimate the following ordinary least square (OLS) regression model:

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

$Entry_{it}^{D}$ is the number of entrants within a *D*-km radius of store *i* in month *t*. $Upgrade_{it}$, the main variable of interest, is an indicator of a positive change in the rating of store *i* in month *t*. A significantly positive value of α_1 indicates the existence of the peer effect, supporting **H1**. X_{it} represents regional characteristics, θ_i denotes store fixed effects, γ_{ct} denotes city-by-month fixed effects, and ε_{it} is the error term. The standard errors are two-way clustered at the store and month levels. In the subsequent analysis regarding **H2**, performance measures are set as the dependent variables.

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

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

where *j* denotes the number of lag periods (n = 5). Other model specifications follow the convention in Equation (1).

5. Results

5.1 Survey of Rural Taobao Merchants

Before investigating the archival online store data from Taobao Marketplace, we begin our analysis with a field survey of Taobao store owners (merchants). We collaborated with a market research company based in mainland China and approached 390 individuals who each owned a Taobao store. For the survey, we restricted our sample of online stores to those in the status of "normally running" and located in rural areas of China. The sample stores are distributed in 12 provinces and 68 cities across China. Using this approach, our sample roughly mirrors the geographical distribution of online stores across Chinese provinces.

Among 390 store owners, 305 individuals provide the valid responses to our questionnaires, and 85 failed to finish the survey in the interviewing process. The response rate of our survey is thus 78.2%. In Appendix 5, we present the statistics of the answers to eight survey questions regarding the motivations for online entrepreneurship. We summarize our findings as follows.

First, the survey results suggest that the decision to become an online entrepreneur is motivated by neighborhood peers. The answers to Question 1 reveal that 46.2% of the interviewees were motivated by observing other locals who run Taobao stores. However, this reason was the third most commonly given, after gaining wealth (86.9%) and becoming one's own boss (55.4%). Other reasons, such as a flexible schedule and government support, appear to be relatively less important. The evidence supports the assumption in our theoretical model that people tend to overweight signals observed in their neighborhood.

Second, we find that neighborhood peers also affect how merchants run their businesses through social interactions. Specifically, the survey results indicate that online merchants often communicate with each other (Question 2) and influence each other in terms of business model and scope (Question 3) and product choice (Question 4). However, the results indicate that many merchants do not analyze the market demand for different products when making operational decisions (Question 5). These findings are consistent with our research design, which is focused on people in rural areas where daily social interactions are perceived to be essential. We also build our tests of heterogeneity on the survey results regarding inattention to product demand.

Third, some findings from the survey support the research design described in Section 4. Most merchants (76.7% in Question 6) reported that they pay attention to the performance of nearby stores (e.g., store ratings), thereby supporting our upgrade event-based identification strategy. Merchants who are motivated by peers' performance are more likely to be disappointed with their own performance (51.3% in Question 7) or even exit the market (39.2% in Question 8), mirroring our tests on aftermarket performance and exit decisions.

5.2 Baseline Estimation

We present our 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 an upgrade event increases the number of newly registered stores within a 0.5-km radius. We then 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 store sales and regional control variables in column 2 and obtain similar results. Interestingly, the sales level of a store does not predict the number of entrants, suggesting entrepreneurs are not particularly motivated by the volume size of a store. In columns 3 and 4, we consider city-by-month fixed effects to capture any city-level trends. The positive peer effect accounts for 3.5% to 3.6% of the sample mean.

[Table 3 about here]

Beside the sales level, a natural extension of our analysis is to explore whether a store's sales growth has explanatory power for the entrepreneurial decision in the neighborhood. One underlying reason is salience theories rely on people's perception on the change instead of the level. In Table 4, we report the results using a store's sales growth. Across all model specifications considered, store's sales growth has significantly positive effect on the number of entrants in the neighborhood. The sample only contains 1,994,776 observations as the measure of sales growth cannot be defined when the store has zero sales in the previous month. In columns (3) and (4), we include both our main variables, *Update*, and *Store Sales Growth*, and find significant coefficients for both. The results suggest the updating event convey additional forward-looking information to the market on top of the sales growth in a given month. We stick to updating events for our identification as it is more visible to people who consider entrepreneurial decisions. The information on sales growth, instead, entails extra efforts in research and comparison.

[Table 4 about here]

It is worthwhile to note that we here investigate the largest online marketplace in mainland China that is the major and obvious option for people who consider online entrepreneurial decision, especially during our sample period. Yet, it is theoretically possible that merchants strategically choose other platforms and avoid direct competition with the local successful incumbent stores. However, this motive would only bias our estimated coefficients downward toward zero. And in summary, the evidence in Tables 3 and 4 is supportive of **H1**.

Although the baseline estimation has considered region-by-time fixed effect to absorb the potential regional trends, unobserved heterogeneity might still exist and confound our results. We continue to 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 alone or in combination with the month level. The results in columns 1 and 2 of Appendix 6 are robust to this alteration. We then replace the dependent variable with the inverse hyperbolic sine transformation of the number of entrants to correct for potential bias due to the distribution of the original count variable. The results in column 3 remain qualitatively similar to the results in the main analysis. We then mitigate the concern that our findings might be driven by demand- or supply-side changes at the product level by constructing a store-product-month level sample and reestimating the effect of neighborhood peers. Our findings hold when we consider trends at the product (column 4) or city-product (column 5) level.

We then estimate the dynamic effect using Equation (2) and present the results in Table 5. As shown in column 2, our baseline model with regional controls indicates that the effect only occurs in the contemporaneous month and that all lagged variables are insignificant. This short-lived effect is inconsistent with explanations based on certain long-term economic confounders; rather, it appears to arise due to behavioral factors related to individual attention (e.g., Busse et al., 2015; Chang et al., 2018).

[Table 5 about here]

5.3 Level of Salience

Our definition of an upgrade event is compatible with salience theories that emphasize the greater attention given to differences rather than absolute values (Gennaioli and Shleifer, 2010). We explore the heterogeneous effect of an upgrade event along two dimensions that capture the level of salience.

Panel A of Table 6 demonstrates the effect of peer performance on the entry decisions of new entrepreneurs according to the magnitude of the store rating increase. As shown in Panel B of Table 1, our sample includes a few cases in which a store's rating was increased by more than one level, thus increasing the salience of the signal to observers in the neighborhood. We split the original treatment variables into two categories: *Upgrade [1 level]* represents one-level upgrade events and *Upgrade [2+ level]* indicates upgrade events of more than one level. As shown in Panel A of Table 5, *Upgrade [2+ level]* events have a large impact, accounting for 9.3% of the sample mean comparing with the effect of *Upgrade [1 level]* (3.1% of the sample mean).

We next evaluate the results according to distance. Specifically, we use alternative distance cutoffs to construct a dependent variable. As shown in Table 5, the effect of an upgrade event within a 0.1-km radius is similar to that in a radius between 0.1 and 0.5 km. However, the effect quickly weakens beyond a radius of 0.5 km and becomes insignificant beyond 1 km. The pattern of effect decay with increasing distance is consistent with both a rational learning story and behavior explanations. A rational learning story implies that the acquisition of better information leads to better decisions, as information quality declines with distance (Giroud, 2013; Huang et al., 2017). Regarding behavioral explanations, the availability heuristic suggests that people tend to rely on immediate examples that come to mind when evaluating a specific topic, concept, method, or decision (Folkes, 1988). In what follows, we will conduct tests to further differentiate the differences implied by the theories.

[Table 6 about here]

Some may raise a more pedestrian explanation for our findings, such as the family expanding their set of outlets selling the same product (or using their experience to help family members get started), particularly when we observe the effect is stronger from those located within 100 meters from the focal business. We argue that this might not be symmetrical pattern that can drive our results. First, we have eliminated any entrants with the same address to mitigate the concern regarding store expansion. Second, the impact is still significant for those entrants between 0.5 km and 1 km that are less likely to be started by family members. Third, we in what follows analyze the performance of those entrants and explore whether the entry decision appear to be rational.

5.4 Types of Entrants

The availability of detailed store product category and owner demographic information enables us to investigate the entrants by type. We first explore the effect of heterogeneity according to product category. The dataset provided by Taobao Marketplace includes the main product category assigned to each store. We then determine the numbers of entrants with the same product category as the upgraded store and those with different product categories. As shown in Panel A of Table 7, upgrade events mainly lead to an increase in the number of stores in the same product category. We argue that this finding is consistent with both behavioral reasons and rational learning. Behavioral bias may lead entrants to mimic the behavior of the upgraded stores. Alternatively, entrants may identify profitable products by observing the upgraded store. Furthermore, in some regions, restrictions may limit the types of products that can be produced.

We next examine whether the cognitive and demographical characteristics of the store owner interact with their choice under risk. The test is related to the analysis in D'Acunto et al.

(2021) in the context of inflation expectation. Figure 6 presents the percentages of entrants in our baseline sample by the owner's age group, education, gender, and marital status. Most of the store owners do not hold a college degree. We explore effect by gender group as previous studies document gender roles significantly affect individual activities (see e.g. D'Acunto et al., 2021). Panels B and C of Table 7 indicate that people who are older than 30 years, less educated, female, and/or unmarried are more subject to the effect of peer performance. Panel D shows the results of an analysis based on the owner's estimated income level, using data provided by Taobao Marketplace. We find that income level is strongly positively correlated with age. Therefore, we define age-adjusted income groups by categorizing the newly registered stores into seven groups according to the estimated monthly income in each age group, as illustrated in Figure 6. Income Groups (1) and (5) indicate the groups with the lowest and highest income levels, respectively. Panel D suggests that the peer effect decreases as the income level increases. This relationship suggests that more sophisticated people (proxied by income level) are less subject to behavioral biases.

[Table 7 about here]

5.5 Performance

We next test **H2** to further differentiate the theories that may better explain our findings.[9] Our purpose is not to measure the economic loss incurred by the "new entrants" is sufficiently large relative to the potential economic gain. For one thing, this practice would be difficult to achieve as the entrepreneurial entry decision depends on exploration-exploitation trade-offs (Vereshchagina and Hopenhayn, 2009; Manso, 2016) that are hard to quantify. For example, the long-run gains of the surviving firms might be quite substantial, while the cost of entry in this industry are small. For another, to differentiate the hypotheses is more tied to the question whether the entry decision of the entrepreneur is perceived to be successful. To this end, we construct a sample, which is also at the store-month level, the

[9] In principle, the performance of the online store depends on the characteristics of the entrepreneurial team. For example, D'Acunto et al. (2020) document the teams with diverse skillsets grow faster than others. But we can only observe the characteristics of the store owner, who is assumed to be the main decision maker in our setting.

same as the baseline sample. But the outcome variables are performance measures of the entrants after registration. Specifically, the sample is again constructed centering around the motivating stores. In our baseline estimation, we look at the number of entrants in the neighborhood of a given store. We in the following tasks examine the performance of those entrants as outcome variable. If there are no entrants registered in the neighborhood of a store, the observation is omitted accordingly. This sample and variable construction approach assures to compare the performance measures among entrants instead of mixing entrants and incumbent. The underlying reason is that entrants, who are likely to be marginal entrepreneurs, theoretically tend to be of poorer quality than the "existing entrants" (Hombert et al., 2020; Hvide and Panos, 2014).

We measure the performance of entrants over the 3- and 6-month periods after registration and report the results in Table 8. The dependent variable is calculated using the sales of newly registered stores rather than those of upgraded stores. *Average Entrant Sales (3m)* and *Average Entrant Sales (6m)* represent the average sales of entrants over the 3- and 6-month periods after registration, respectively. Again, if there are no entrants in a given month, the dependent variable cannot be defined and is dropped. The results shown in Table 7 indicate that entrepreneurs motivated by upgrade events underperform the sample mean by 7.9%–8.7%. The results are supportive of **H2**.

[Table 8 about here]

Store owners who realize the irrationality of their entrepreneurial activities may reverse their decision and exit the market. We calculate the percentage of entrants who exited within 3 and 6 months after registration. The results shown in Table 9 indicate that entrepreneurs motivated by upgraded events are 1.5% to 1.8% more likely to exit the market than the sample mean, suggesting a suboptimal original entry decision. However, as suggested by our survey results (Q8), the majority of the entrepreneurs choose to stay in the business even they believe the performance is not in line with their expectation. The observed exit decisions thus might not be at the optimal time.

[Table 9 about here]

The model proposed in Section 3 predicts that entrepreneurs tend to rely only on signals from their own neighborhood, leading to inadequate economic consequences. We test this prediction by constructing the product-level sales growth in a given city from month *t-1* to month *t*. In Table 9, we interact this variable with the main treatment variable. The results in columns 1 and 2 indicate that the underperformance of entrants in a specific product category is exaggerated when the market is experiencing a downturn. We also define a binary variable, *Low Product Sales Growth*, which is set to 1 if the aggregated sales growth for product *k* in stores located in city *c* from month *t-1* to month *t* for stores is below the median, and 0 otherwise. Columns 3 and 4 of Table 8 consistently confirms the model prediction that peer-motivated store owners who underperform are more likely to ignore the overall market conditions and overweight signals observed in the neighborhood.

Some may worry that it is possible that the increased competition drives underperformance. And we interpret this possibility is consistent with the model prediction. Business owners if expect the expected competition rationally would avoid enter the market in this period. The ignorance of the increased competition thus can be interpreted as a channel to explain the results on underperformance.

[Table 10 about here]

In summary, the results of our analysis of exit decisions support **H2** and thus contradict the explanations for our main findings, such as rational learning and economic confounders. The results in this subsection are mostly consistent with behavioral reasons, particularly salience theories.

5.6 Downgrade Events

We also explore the impact of downgrade events. As shown in Panel B of Table 1, stores rarely experience downgrade events. Our sample contains only 520 stores with downgrade events for which precise location information and other essential variables are available. In

Appendix 7, we present the results of this analysis using an indicator, *Downgrade*, which is set to 1 for the month in which the rating of a given store is downgraded. We do not observe a significant effect of these events on the number of entrants in the neighborhood, suggesting that entrepreneurial decisions tend to be influenced by positive rather than negative surprises. These results are in line with the predictions of Hirshleifer and Plotkin (2021) model of salience in which negative outcomes are sometimes censored.

5.7 Discussions

We provide additional discussions regarding the empirical design and interpretation of our findings as follow. First, our identification strategy is based on the unique features of the rating system of the online marketplace. The rating upgrade events as discussed above are rare, and thus create discrete salience that attract the attention of the people who is considering entrepreneurial event. In addition, comparing with other performance measures, such as sales, the rating of a store is easily accessible publicly and thus more visible to people in the neighborhood. Other performance measures, in contract, usually require extra efforts in research and comparison.

Second, we acknowledge the potential endogeneity concerns that unobserved factors may confound our findings. For instance, local government support at the village level might drive both the upgrade events and increased new entrants. To mitigate the concerns, we build our analysis on the hyper local level and use monthly sample. It is less likely that the monthly intertemporal effect of upgrading events and entrants coincides with local government actions. We also provide evidence, such as the declining effect by distance and underperformance of the entrants, to further support our interpretations.

Last, one key underlying assumption of this study is that the decision of online entrepreneurial decision is not costless. We discuss in Section 2.2 that merchants need to incur some direct financial costs in the process of registration. But it is true that comparing with traditional offline business, online business has lower entry barrier. And the platform continuously aims to lower the costs to attract more merchants. Some may worry the present behavioral biases are less meaningful in an environment where "mistakes" have little costs. To alleviate the concern, in the empirical analysis we only count newly registered stores with non-zero sales in our sample. This practice assures that the real business and transactions indeed happen, imposing operating costs and opportunity costs on entrepreneurs. As indicated by the finding in the survey (Q7 and Q8), merchants do not exit at the optimal time even they realize the performance is not as expected, leading to prolonged underperformance period and additional losses.

6. Conclusion

Online businesses account for a large proportion of the total economy. However, the incentives to create an online business are not well understood. In this study, we leverage a unique dataset from the world's largest C2C platform to explore whether and how social interactions motivate the (excess) entry of new entrepreneurs to the online market.

We find that people who observe an upgrade event in their neighborhood are more likely to become online entrepreneurs. This effect is statistically and economically significant and robust to a range of rigorous model specifications. The impact of peer performance increases with the level of saliency of the upgrade event and decreases with distance. Further testing indicates that such entry decisions might not be rational, as entrants whose owners are motivated by upgrade events underperform their peers in terms of lower sales and have high probability of exit from the market.

Our evidence on the underperformance of entrants, especially in negative overall market conditions, goes against rational explanations, such as regional economic development and rational learning. The results instead lend support to salience theories of choice, which argue that people tend to overweight salient payoffs when making decisions.

Our study contributes to the literature on the drivers of entrepreneurship in the context of online business by exploiting the nuanced measures of an online marketplace platform. We add to the debate on the role of the peer effect in entrepreneurial decisions and emphasize a behavioral channel.

If indeed individuals choose to become entrepreneurs based on salient data, then publicizing stories of successful entrepreneurs may drive entrepreneurship. Our paper suggests one ought to be careful though as at least in our sample entrepreneurs who rely on salient data tend to underperform. However, to the extent that entrepreneurship has positive externalities, such as knowledge spillovers, it might still be in the interest of governments and societies to broadcast such stories. This is an interesting direction for future research.

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.

Bailey, M., Johnston, D.M., Kuchler, T., Stroebel, J. and Wong, A., 2019. Peer effects in product adoption (No. w25843). National Bureau of Economic Research.

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. Working Paper. 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.

Bordalo, P., Gennaioli, N. and Shleifer, A., 2012. Salience theory of choice under risk. *The Quarterly Journal of Economics*, *127*(3), pp.1243-1285.

Bordalo, P., Gennaioli, N. and Shleifer, A., 2013a. Salience and consumer choice. *Journal of Political Economy*, *121*(5), pp.803-843.

Bordalo, P., Gennaioli, N. and Shleifer, A., 2013b. Salience and asset prices. *American Economic Review*, *103*(3), pp.623-28.

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.

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.

Chetty, R., Looney, A. and Kroft, K., 2009. Salience and taxation: Theory and evidence. *American Economic Review*, *99*(4), pp.1145-77.

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.

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*.

D'Acunto, F., D. Hoang, M. P., and Weber, M., 2021. IQ, Expectations, and Choice. Forthcoming in *Review of Economic Studies*.

D'Acunto, F., Malmendier, U., and Weber, M., 2021. Gender Roles and the Gender Expectations Gap, Forthcoming in *Proceedings of the National Academy of Sciences*

D'Acunto, F., Tate, G.A. and Yang, L., 2020. Entrepreneurial Teams: Diversity of Skills and Early-

Stage Growth. Available at SSRN.

DellaVigna, S. and Pollet, J.M., 2009. Investor inattention and Friday earnings announcements. *The Journal of Finance*, 64(2), pp.709-749.

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., Klopack, 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.

Gennaioli, N. and Shleifer, A., 2010. What comes to mind. *The Quarterly Journal of Economics*, *125*(4), pp.1399-1433.

Giroud, X., 2013. Proximity and investment: Evidence from plant-level data. *The Quarterly Journal of Economics*, 128(2), pp.861-915.

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. Internet rising, prices falling: Measuring inflation in a world of e-commerce. *AEA Papers and Proceedings*, *108*, pp. 488-492.

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.

Hirshleifer, D. and Plotkin, J., 2021. Moonshots, Investment Booms, and Selection Bias in the Transmission of Cultural Traits. Working Paper.

Hirshleifer, D., S. S. Lim, and S. H. Teoh. 2011, Limited Investor Attention and Stock Market Misreactions to Accounting Information. *Review of Asset Pricing Studies* 1 : 35-73.

Hirshleifer, D., Teoh, S., 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 35, 337–386

Hombert, J., Schoar, A., Sraer, D. and Thesmar, D., 2020. Can unemployment insurance spur entrepreneurial activity? Evidence from France. *The Journal of Finance*, 75(3), pp.1247-1285.

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.

Hvide, H.K. and Panos, G.A., 2014. Risk tolerance and entrepreneurship. *Journal of Financial Economics*, 111(1), pp.200-223.

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.

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.

Kuchler, T. and Stroebel, J., 2020. Social Finance, forthcoming in *Annual Review of Financial Economics.*

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. (March 15, 2016). Harvard Business School NOM Unit Working Paper, (12-016).

Manso, G., 2016. Experimentation and the Returns to Entrepreneurship. *The Review of Financial Studies,* 29, pp. 2319-2340.

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.

Robinson, D.T. and Stuart, T.E., 2007. Network effects in the governance of strategic alliances. *The Journal of Law, Economics, & Organization*, 23(1), pp.242-273.

Pistaferri, L., Guiso, L. and Schivardi, F., 2021. Learning entrepreneurship from other entrepreneurs? *Journal of Labor Economics*, 39(1), pp.135-191.

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.

Vereshchagina, G. and Hopenhayn, H.A., 2009. Risk taking by entrepreneurs. *American Economic Review*, 99(5), pp.1808-30.

Table 1. Store Rating System Used by Taobao Marketplace

Notes: This table presents a summary of the store rating system. Panel A shows the numbers of stores in different rating categories. We also present the average monthly sales and average monthly page views of stores with different ratings. Panel A is based on the full sample of stores on Taobao Marketplace. Panel B summarizes rating change (upgrade/downgrade) events. The first column of Panel B shows the change in the rating level from month *t-1* to month *t*. Our baseline sample comprises stores with upgrade events.

Panel A. Summary of Performance by Store Rating

Rating	Rating Name	Number of Stores	Average Monthly Sales	Average Monthly Page Views
0	no rating	5,545,977	436	116
1	1 heart	1,136,325	660	278
2	2 hearts	1,880,862	1,317	419
3	3 hearts	1,278,553	2,135	684
4	4 hearts	784,573	3,182	942
5	5 hearts	740,805	4,289	1,243
6	1 crystal	1,031,379	6,037	1,627
7	2 crystals	857,647	10,115	2,501
8	3 crystals	689,709	16,260	3,831
9	4 crystals	641,062	28,032	6,222
10	5 crystals	292,391	44,315	10,466
11	1 crown	185,064	66,502	16,696
12	2 crowns	117,906	112,366	31,684
13	3 crowns	37,038	193,000	61,692
14	4 crowns	16,031	307,926	117,786
15	5 crowns	7,211	586,484	276,163
16	1 gold crown	1,660	1,156,974	639,958
17	2 gold crowns	548	2,309,458	1,529,410
18	3 gold crowns	214	4,402,325	3,292,865
19	4 gold crowns	48	6,465,623	4,211,521
20	5 gold crowns	24	8,180,417	10,129,738

Table 1. Store Rating System Used by Taobao Marketplace (Continued)

Panel B. Summary of Rating Changes

Rating (t) - Rating (t-1)	Frequency	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 at the store-month level. The sample comprises 133,427 unique online stores located in rural areas of China that experienced increased store ratings during the August 2014 to August 2016 period. The panel sample contains 3,048,614 store-month observations. The sample construction process is detailed in Appendix 2. The variables are defined in Appendix 4.

	N	Mean	SD	Median
Entry Variables				
Entry [<0.5km]	3,048,614	1.298	4.541	0.000
Entry [0.1km]	3,048,614	0.622	2.744	0.000
Entry [0.1km - 0.5km]	3,048,614	0.676	2.656	0.000
Entry [0.5km - 1km]	3,048,614	1.070	2.924	0.000
Entry [1km - 2km]	3,048,614	2.371	5.213	0.000
Treatment Variables				
Upgrade	3,048,614	0.044	0.204	0.000
Upgrade [1 level]	3,048,614	0.040	0.196	0.000
Upgrade [2+ level]	3,048,614	0.004	0.062	0.000
Control Variables				
Store Sales Growth	1,994,776	0.751	3.858	-0.103
Store Sales	3,048,614	27.472	212.484	1.100
County Store Number	3,048,614	8086.081	9582.933	5085.000
County Sales	3,048,614	10.319	7.822	8.580
County Light	3,048,614	10.698	1.080	10.717
Performance				
Average Entrant Sales (3m)	1,014,705	4.894	28.932	0.500
Average Entrant Sales (6m)	677,820	9.677	68.181	0.200
Entrant Exit Share (3m)	1,014,705	0.267	0.381	0.000
Entrant Exit Share (6m)	677,820	0.513	0.442	0.500

Table 3. Baseline Estimation: Rating Upgrades and Online Entrepreneurship

Notes: This table presents a baseline estimation of the spillover effect of store rating increases (upgrade events) on the entry of other neighborhood stores into the online market. The sample is at the store-month level and consists of stores that experienced upgrade events during the sample period. The dependent variable, *Entry [< 0.5 km]*, is the number of entrants in a 0.5-km radius of a given store *i* in month *t*. The independent variable, *Upgrade*, is an indicator set to 1 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 level of a store's monthly sales, the number of active stores in a county, the average monthly sales (in thousands) of active stores in a county, and the log value of night light in a county. The *t*-statistics are reported in brackets and are based on standard errors clustered at both the store and month levels. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
	Entry [<0.5km], t			
Upgrade, t	0.099**	0.101***	0.046***	0.047***
	[2.67]	[2.80]	[3.45]	[3.50]
Store Sales		-0.000		-0.000
		[-0.34]		[-1.47]
County Store Number		-0.001***		-0.001***
		[-3.50]		[-3.30]
County Sales		0.017***		0.005
		[3.19]		[1.20]
County Light		0.252**		0.291*
		[2.33]		[2.01]
N	3,048,614	3,048,614	3,048,614	3,048,614
R^2	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	7.78	3.54	3.62

Table 4. Robustness Checks: Alternative Performance Measure

Notes: This table presents the robustness checks for the main estimation using a store's sales growth as performance measure. The sample is at the store-month level and consists of stores that experienced upgrade events during the sample period. The dependent variable, *Entry [< 0.5 km]*, is the number of entrants in a 0.5-km radius of a given store *i* in month *t*. The independent variable, *Store Sales Growth*, is the monthly sales growth for a store. The observations are dropped when the sales level in the last month is zero. *Upgrade* is an indicator set to 1 if the given store *i* is upgraded in month *t*. Other model specifications follow the conventions in Table 3.

	(1)	(2)	(3)	(4)
	Entry [<0.5km], t			
Store Sales Growth, t	0.004**	0.003***	0.004**	0.003***
	[2.21]	[2.91]	[2.26]	[2.95]
Upgrade, t			0.082**	0.033***
			[2.40]	[3.21]
Store Sales	0.000	-0.000	-0.000	-0.000
	[0.04]	[-1.31]	[-0.08]	[-1.35]
County Store Number	-0.001***	-0.001***	-0.001***	-0.001***
	[-3.35]	[-3.11]	[-3.35]	[-3.11]
County Sales	0.000***	0.000	0.000***	0.000
	[2.95]	[0.71]	[2.95]	[0.71]
County Light	0.287**	0.293*	0.286**	0.293*
	[2.48]	[1.84]	[2.48]	[1.84]
N	1,994,776	1,994,776	1,994,776	1,994,776
R^2	0.757	0.816	0.757	0.816
Store FE	X	X	X	X
Month FE	X		X	
City-Month FE		X		X

Table 5. The Dynamic Effect

Notes: This table presents the dynamic effects of store rating increases (upgrade events) by including related lagged variables. 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 thousands) of active stores in a county, and the log value of night light in a county. The *t*-statistics are reported in brackets and are based on standard errors clustered at both the store and month levels. *** indicates significance at the 1% level.

	(1)	(2)
	Entry [<0.5km], t	
Upgrade, t	0.043***	0.044***
	[3.81]	[3.85]
Upgrade, t-1	-0.008	-0.006
	[-0.51]	[-0.43]
Upgrade, t-2	-0.026	-0.024
	[-1.41]	[-1.36]
Upgrade, t-3	-0.014	-0.013
	[-1.16]	[-1.08]
Upgrade, t-4	-0.006	-0.005
	[-0.59]	[-0.48]
Upgrade, t-5	0.004	0.005
	[0.41]	[0.53]
N	3,048,614	3,048,614
R^2	0.791	0.792
Controls		X
Store FE	X	X
City-Month FE	X	X

Table 6. Level of Salience

Notes: This table presents the effect of the level of salience of an upgrading event on the number of entrants to the online market. Panel A shows the results according to the magnitude of the upgrade. Panel B presents the results according to the distance from 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 thousands) of active stores in a county, and the log value of night light in a county. The *t*-statistics are reported in brackets and are based on standard errors clustered at both the store and month levels. *** indicates significance at the 1% level.

Panel A. Upgrade Level

	(1)	(2)
	Entry [<0.5km]	Beta/Mean (%)
Upgrade [1 level]	0.040***	3.08
	[3.31]	
Upgrade [2+ level]	0.121***	9.32
	[3.49]	
N	3,048,614	
R^2	0.792	
Controls	X	
Store FE	X	
City-Month FE	X	

Panel B. Distance

	(1)	(2)	(3)	(4)
	Entry [<0.1km]	Entry [0.1km-0.5km]	Entry [0.5km-1km]	Entry [1km-2km]
Upgrade	0.022***	0.025***	0.020***	0.015
	[4.11]	[2.82]	[3.46]	[1.68]
N	3,048,614	3,048,614	3,048,614	3,048,614
R^2	0.768	0.750	0.699	0.774
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	1.07	2.371
Beta/Mean (%)	3.54	3.7	1.87	0.63

Table 7. Types of Entrants

Notes: This table presents the number of entrants by store owners' demographic characteristics. Panel A shows the number of entrants according to whether the new and incumbent stores have the same product category. Panel B presents the number of entrants by the age and education level of the owner. "Non-college" and "college" indicate the lack and possession of a college degree, respectively. Panel C shows the number of entrants according to the gender and marital status of the owner. Panel D presents the peer effect using sub sample based 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 Groups (1) and (5) indicate the groups with the lowest and highest income levels, respectively. 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 thousands) of active stores in a county, and the log value of night light in a county. The *t*-statistics are reported in brackets and are based on standard errors clustered at both the store and month levels. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Product Category

	(1)	(2)
	Entry [<0.5km], t	
	Same Product	Different Product
Upgrade, t	0.041**	0.005
	[2.38]	[0.82]
N	3,048,614	3,048,614
R^2	0.805	0.675
Controls	X	X
Store FE	X	X
City-Month FE	X	X
H_0: beta(1)=beta(2)	p<0.01	
Mean of Dep. Var.	0.541	0.757
Beta/Mean (%)	7.58	0.66

Panel B. Age (years) and Education

	(1)	(2)	(3)	(4)
	Entry [<0.5km], t			
	Age [<=30]	Age [>30]	Non-College	College
Upgrade, t	0.020***	0.026***	0.046***	0.000
	[3.63]	[3.04]	[3.37]	[0.20]
N	3,048,614	3,048,614	3,048,614	3,048,614
R^2	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
H_0: beta(1)=beta(2)	p=0.01			
H_0: beta(3)=beta(4)			p<0.01	
Mean of Dep. Var.	0.760	0.488	1.128	0.170
Beta/Mean (%)	2.63	5.33	4.08	0.17

Table 7. Types of Entrants (Continued)

Panel C. Gender and Marital Status

	(1)	(2)	(3)	(4)
	Entry [<0.5km], t			
	Male	Female	Married	Not Married
Upgrade, t	0.021***	0.025***	0.016***	0.021***
	[3.54]	[3.19]	[3.17]	[3.59]
N	3,048,614	3,048,614	3,048,614	3,048,614
R^2	0.744	0.735	0.685	0.748
Controls	X	X	X	X
Store FE	X	X	X	X
City-Month FE	X	X	X	X
H_0: beta(1)=beta(2)	p=0.08			
H_0: beta(3)=beta(4)			p=0.02	
Mean of Dep. Var.	0.698	0.551	0.531	0.467
Beta/Mean (%)				

Panel D. Income Groups

	(1)	(2)	(3)	(4)	(5)
	Entry [<0.5km], t				
	Income Group (1)	Income Group (2)	Income Group (3)	Income Group (4)	Income Group (5)
Upgrade, t	0.018***	0.009***	0.008***	0.005*	0.001
	[2.89]	[4.19]	[3.27]	[1.85]	[0.93]
N	3,048,614	3,048,614	3,048,614	3,048,614	3,048,614
R^2	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.	0.298	0.246	0.229	0.208	0.151
Beta/Mean (%)	6.04	3.66	3.49	2.4	0.66

Table 8. Performance of Entrants

Notes: This table presents the results of a performance analysis of the entrants around the upgraded stores within 3 and 6 months after registration. *Average Entrant Sales (3m)* and *Average Entrant Sales (6m)* represent the average monthly sales (in thousands) in the 3- and 6-month periods after registration, respectively, of stores within a 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 thousands) of active stores in a county, and the log value of night light in a county. The *t*-statistics are reported in brackets and are based on standard errors clustered at both the store and month levels. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(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	1,014,705	1,014,705	677,820	677,820
R^2	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 9. Exit Decisions of Entrants

Notes: This table presents an analysis of the exit decisions of entrants around the upgraded stores. *Entrant Exit Share (3m)* and *Entrant Exit Share (6m)* represent the percentages of entrants that exited the market within 3 and 6 months after registration, respectively. 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 thousands) of active stores in a county, and the log value of night light in a county. The *t*-statistics are reported in brackets and are based on standard errors clustered at both the store and month levels. *** indicates significance at the 1% 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	1,014,705	1,014,705	677,820	677,820
R^2	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 10. Entrant Performance and Product Market Performance

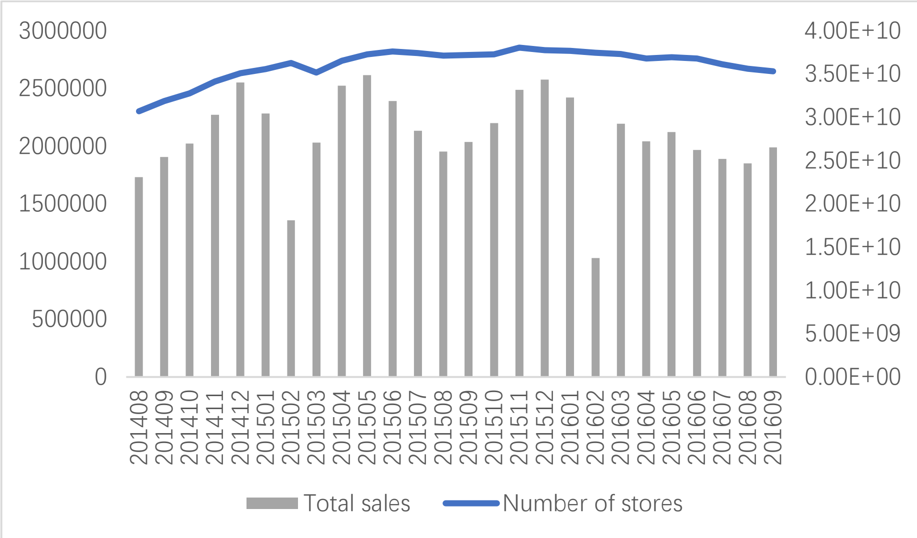
Notes: This table presents the results of an analysis of entrants' aftermarket and product market performance. We use the aggregated sales growth in a given product category and city to measure the product market performance. The sample is at the store-product-month level. The dependent variable, *Average Entrant Sales (3m)*, is the average monthly sales in product category *k* in the 3-month period after registration and applies to stores that registered within a 0.5-km radius of the upgraded sample store i in month *t*. *Product Sales Growth* is the aggregated sales growth of 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 1 if the aggregated sales growth of product *k* from month *t-1* to month *t* for stores located in city *c* is below the median, and 0 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 thousands) of active stores in a county, and the log value of night light in a county. The *t*-statistics are reported in brackets and are based on standard errors clustered at both the store and month levels. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(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	2,058,898	2,058,898	2,058,898	2,058,898
R^2	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 numbers and the performance of the full sample of online stores with a "normal status" on Taobao Marketplace. Panel A shows the total sales (right scale) and the number of stores (left scale) over time. Panel B presents the number of entrants in the sample period.

Panel A. Sales and the Number of Stores



Panel B. Number of Entrants

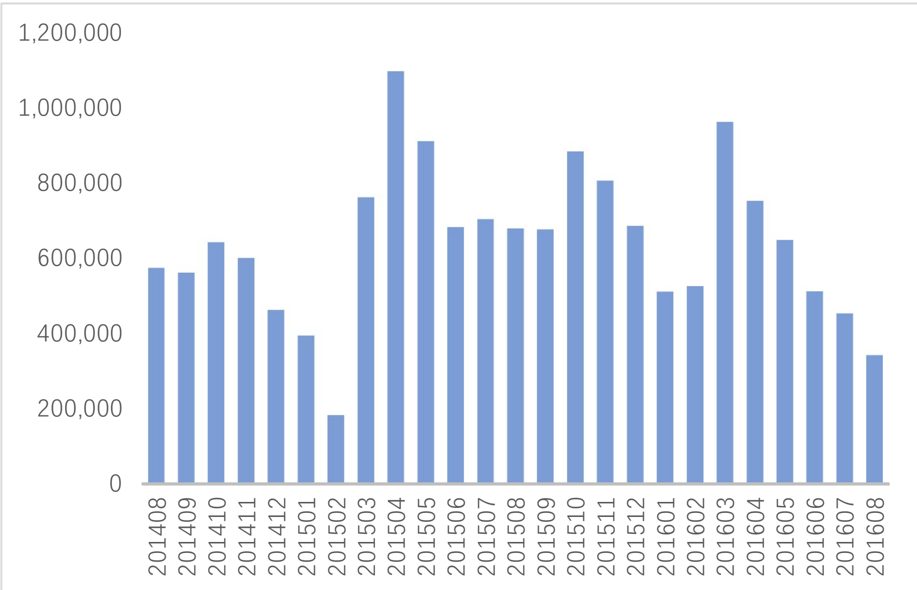
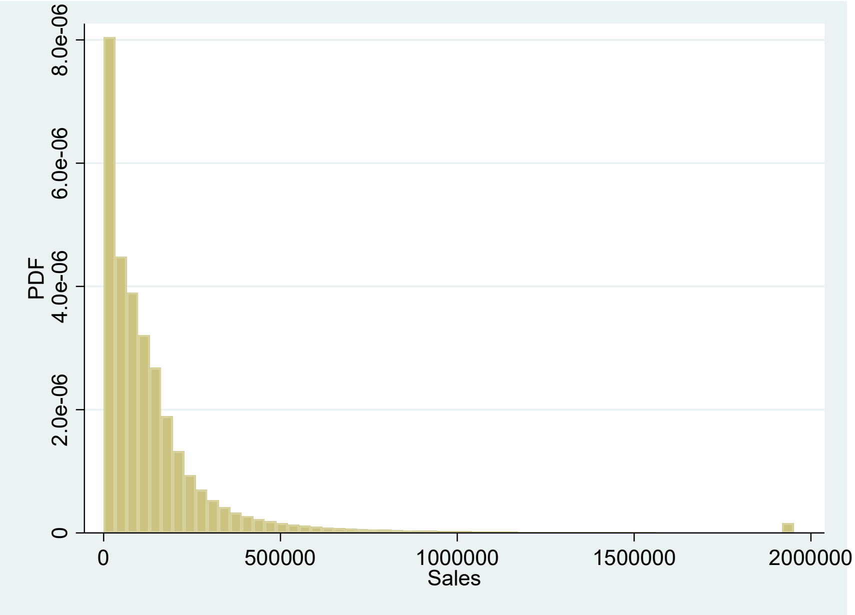


Figure 2. Sales Distribution

Notes: This figure depicts the distribution of sales across online stores on Taobao Marketplace. Panel A presents the probability density function of the monthly sales across stores. This density function is plotted on a log-log scale in Panel B.

Panel A. Probability Density Function



Panel B. Log-log Scale Plot

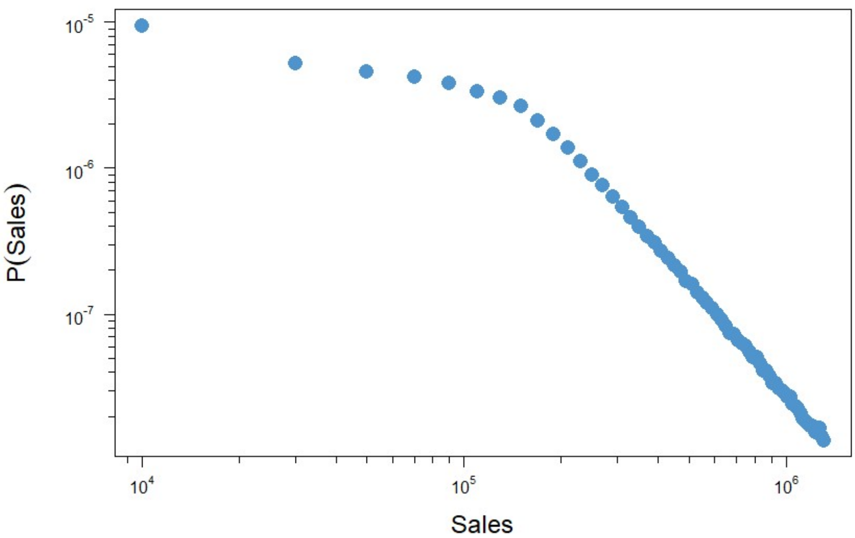
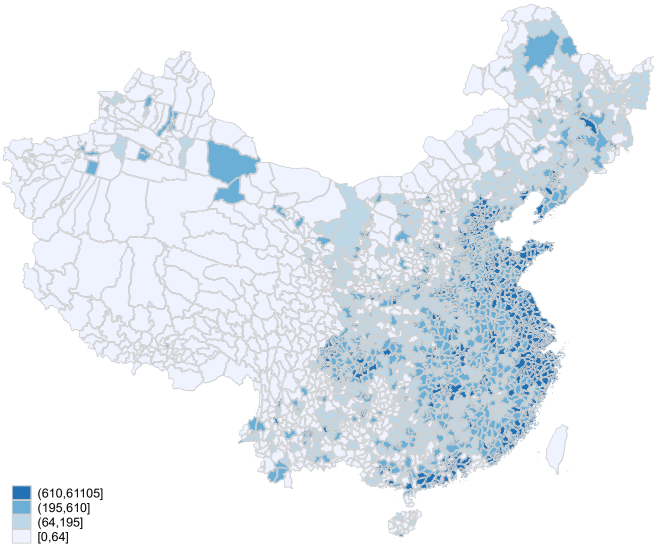


Figure 3. Spatial Distribution of Stores on Taobao Marketplace

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

Panel A. Number of Stores per County



Panel B. Number of Entrants per County

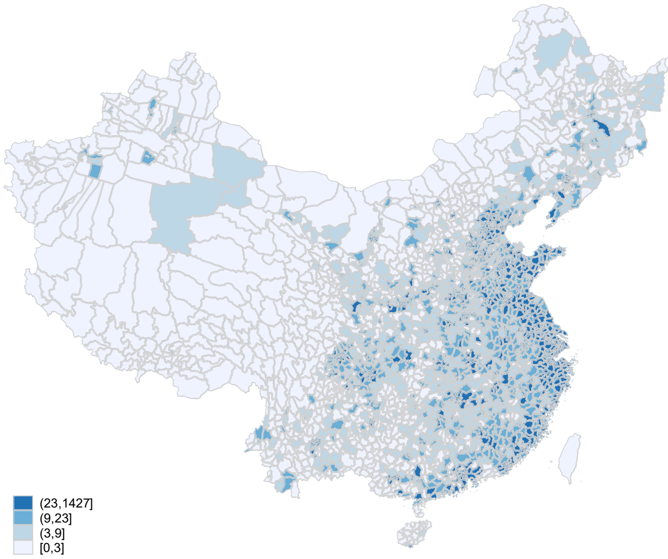


Figure 4. Store Rating System

Notes: Panel A illustrates the store rating system used by Taobao Marketplace. Column 1 lists the internal score ranges used to assign ratings. Columns 2 and 3 present the symbols and names of the store ratings, respectively. Column 4 shows the values we assigned to the rating categories. Panel B depicts a representative store rating, which is publicly available to all customers.

Panel A. The Rating System

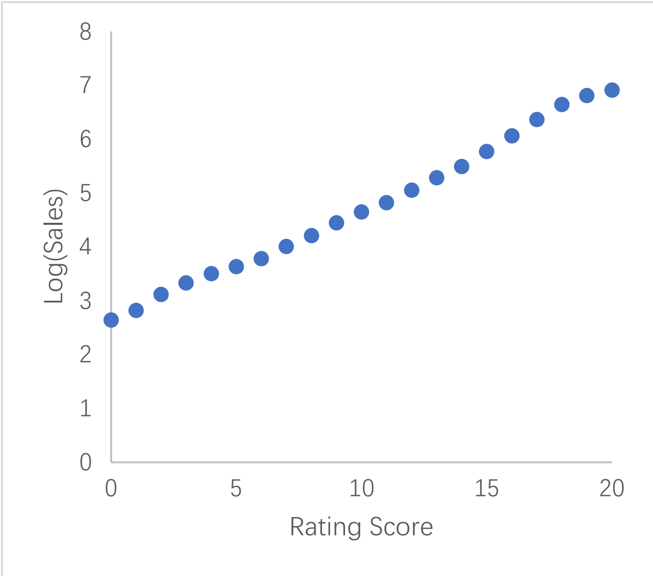
所积分数	等级图标	信誉等级	Rating
4分-10分 Score	Symbol	一星 Name	1
11分-40分		二星	2
41分-90分		三星	3
91分-150分		四星	4
151分-250分		五星	5
251分-500分		一钻	6
500分-1000分		二钻	7
1001分-2000分		三钻	8
2001分-5000分		四钻	9
5001分-10000分		五钻	10
10001分-20000分		一皇冠	11
20001分-50000分		二皇冠	12
50001分-100000分		三皇冠	13
100001分-200000分		四皇冠	14
200001分-500000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. Example Store Rating

Figure 5. Store Rating and Performance

Notes: This figure demonstrates the relationship between store rating and performance. Panel A plots the average log value of monthly sales by store rating score. Panel B shows the average monthly page views of stores in different rating categories.

Panel A. Store Ratings vs. Sales



Panel B. Store Ratings vs. Page Views

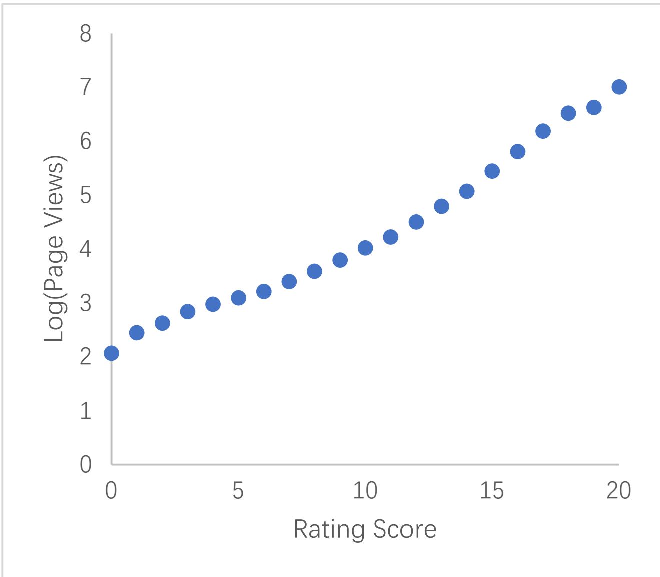
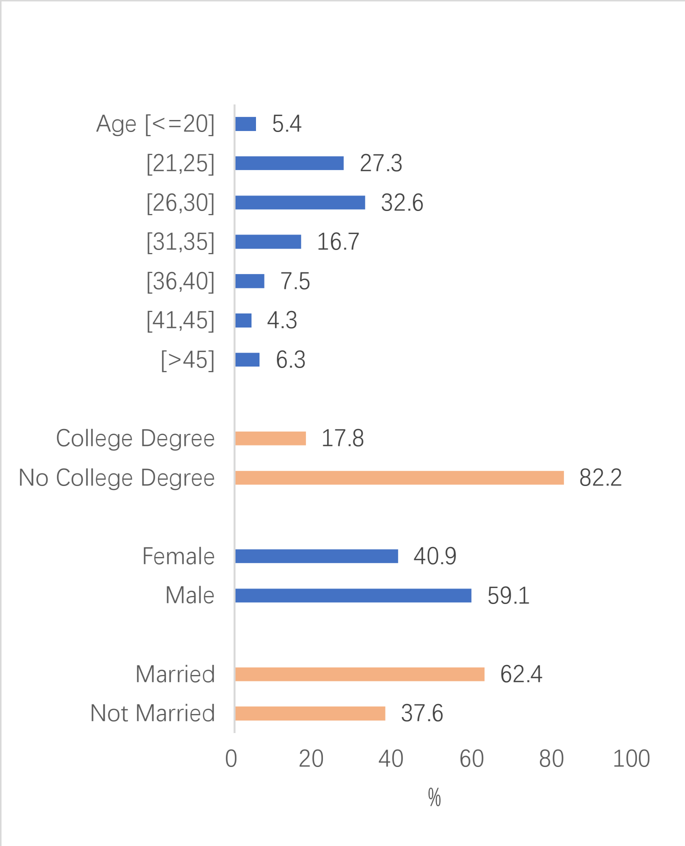


Figure 6. Breakdown of Entrants by Owners' Characteristics

Notes: This figure presents the distribution of entrants by the owners' characteristics.



Appendix 1. Registration of an Online Store on Taobao Marketplace

Notes: This appendix depicts the process of registering an online store. The initial cost of store registration is RMB 1000 (approximately USD 144). This registration fee is refundable when the store is dissolved. The owner is also required to pay a monthly technical support fee of RMB 50 (approximately USD 7.2) and to provide the following documentation upon registering: personal ID, certified phone number, and bank account information. The process is illustrated below.

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

Appendix 2. Sample Construction

Notes: The baseline sample is constructed as follows.

1. Choose stores with a "normal" status.

2. Restrict sample to stores in rural areas with available location information.

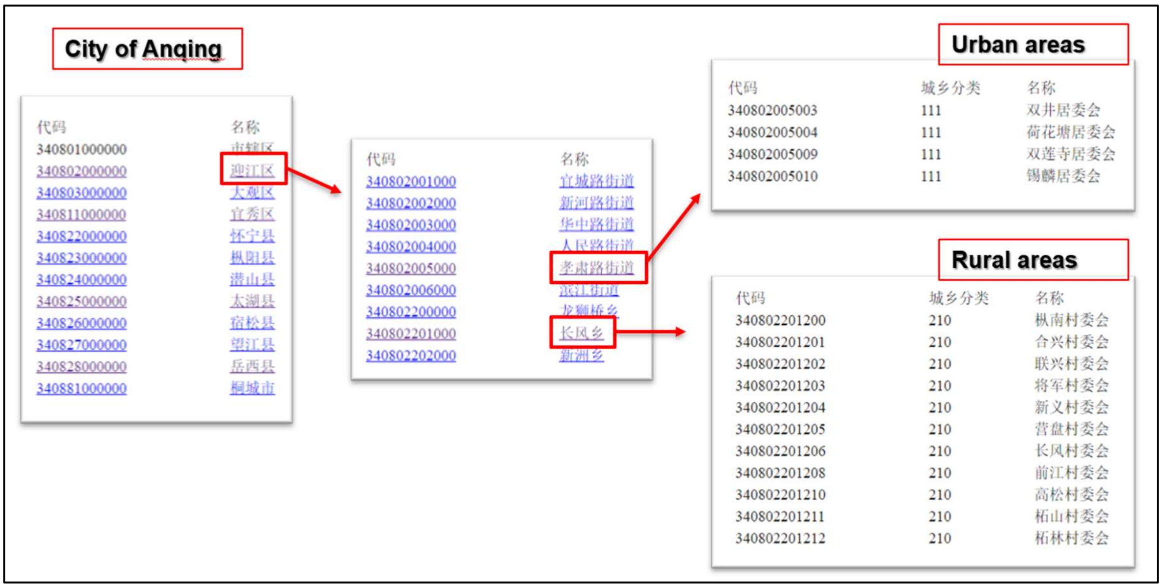
3. Restrict sample to stores that experience an increase in rating during the sample period.

4. The final sample contains 3,048,614 store-month observations and 133,427 distinct stores.

Appendix 3. Rural vs. Urban Areas

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

Panel A. Example: Identifying Rural Areas



Panel B. Rural vs. Urban: Performance

	Number of Stores	Percentage of Stores	Average Monthly Sales (RMB)	Average Rating
Urban	2,513,396	59.5%	11,337.0	3.9
Rural	1,709,807	40.5%	9,229.6	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 vegetables
17	Antiques	Furniture
18	Toiletries	Digital Accessories
19	Diapers	Toys
20	Household necessities	Men's shoes

Appendix 4. Variable Definitions

Variable	Definition
Entrant [X km] $_{i,t}$	The number of newly registered stores operating in an X km radius of upgraded store *i* in month *t*. The variable is defined at the store (*i*), month (*t*), and product (*q*) levels.
Upgrade $_{i,t}$	Indicator variable set to 1 if store *i* experiences positive changes in its rating in month *t* and 0 otherwise.
Upgrade [Z level] $_{i,t}$	Indicator variable set to 1 if the rating of store *i* is upgraded by Z levels in month *t* and 0 otherwise. For example, *Upgrade [2+ level]* indicates that the rating of the sample store is upgraded by two or more levels in month *t*.
Store Sales Growth $_{c,t}$	The growth rate of the monthly sales from month *t-1* to month *t* for a given firm *i*.
Store Sales $_{c,t}$	The level of sales for store *i* in month *t*.
County Store Number $_{c,t}$	The number of active stores in county *c* in month *t*.
County Sales $_{c,t}$	The average monthly sales of active stores in county *c* in month *t*.
County Light $_{c,t}$	The log value of the night light based on satellite data in county *c* in month *t*.
Rating	The rating of store *i* in month *t*. There are 20 rating levels, as defined in Table 1.
Average Entrant Sales (3m/6m) $_{i,t}$	The average monthly sales within 3 and 6 months after registration among stores that registered within a 0.5-km radius of upgraded store *i* in month *t*.
Entrant Exit Share (3m/6m) $_{i,t}$	The percentage of stores that exit the market within 3 and 6 months after registration among stores that registered within a 0.5-km radius of upgraded store *i* in month *t*.
Product Sales Growth $_{c,k,t}$	The aggregated sales growth of 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 1 if the aggregated sales growth for product *k* from month *t-1* to month *t* is below the median for stores located

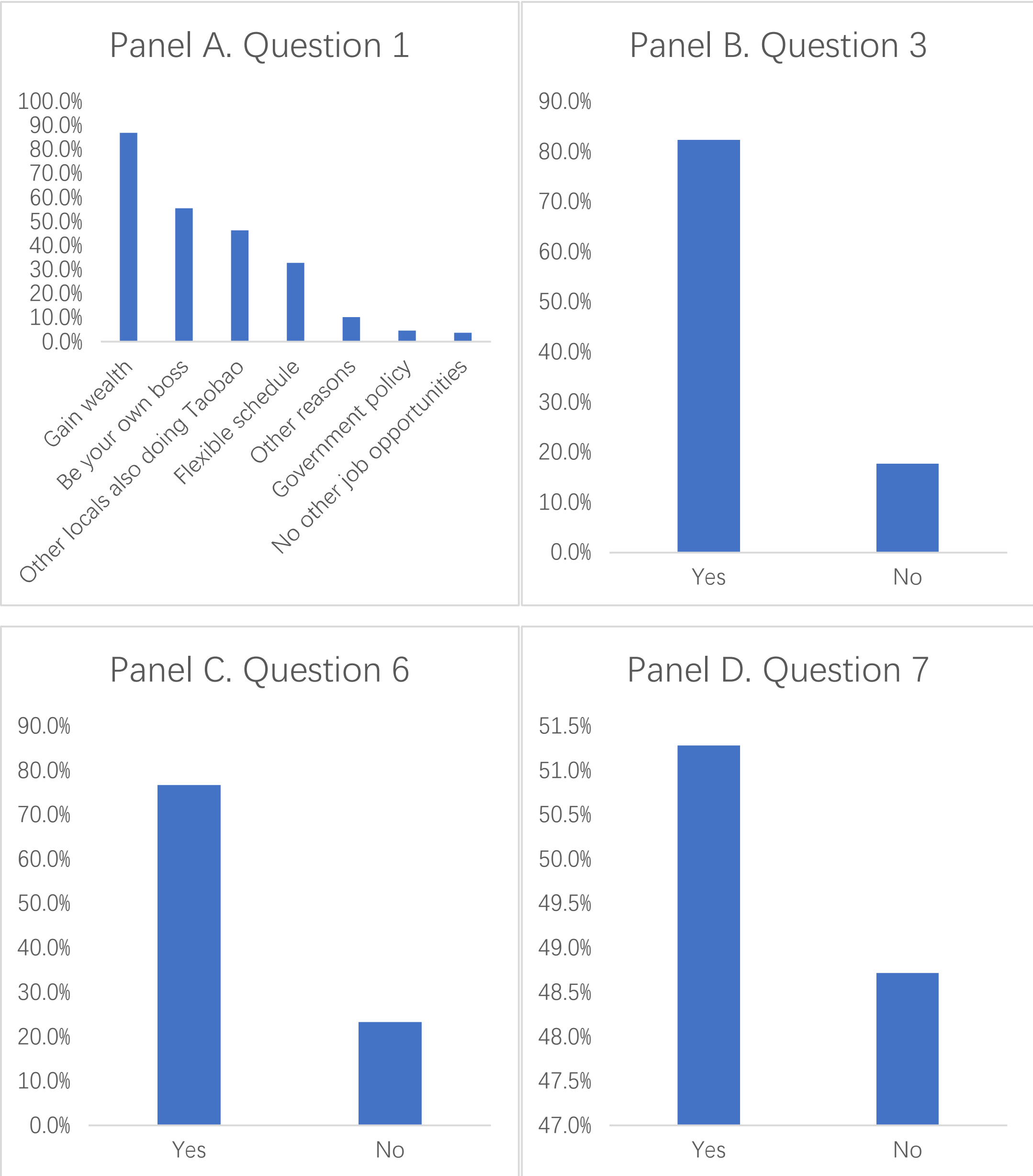
in city *c* and 0 otherwise.

Appendix 5. Results of a Survey of Rural Taobao Merchants

Notes: We conducted a field survey of Taobao merchants across 12 provinces and 68 cities in China and construct a sample of 305 reliable observations out of 390 surveyed individuals, each corresponding to the owner of a Taobao store. The results of survey responses to eight questions are reported as follows.

Q1. What are the main factors that motivated you to open a Taobao store?	
A. Gain wealth	86.9%
B. No other job opportunities	3.6%
C. I saw other locals using 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%
Q2. Do you often ask your friends about e-commerce issues?	
A. Never	0.0%
B. Occasionally	26.9%
C. Normally	28.2%
D. Often	34.4%
E. Frequently	10.5%
Q3. Do you learn from others in terms of the online store business model and business scope?	
A. Yes	82.3%
B. No	17.7%
Q4. Did you consider the products of nearby residents' Taobao stores when choosing the product to sell in your Taobao store?	
A. Yes	74.4%
B. No	25.6%
Q5. Do you analyze the market demand for different products when you choose the types of products to sell?	
A. Yes	71.5%
B. No	28.5%
Q6. Did you consider the performance of nearby residents' Taobao stores (e.g., store ratings) when deciding to open a Taobao store?	
A. Yes (Continue with Q7)	76.7%
B. No	23.3%
Q7. If you considered the performance of nearby residents' Taobao stores (e.g., store ratings) when deciding to open a Taobao store, did you feel that it was not in line with your expectations of your future operations?	
A. Yes (Continue with Q8)	51.3%
B. No	48.7%
Q8. If you considered the performance of nearby residents' Taobao stores (e.g., store ratings) when deciding to open a Taobao store and you feel that your business performance is not in line with your expectations, will you terminate your store operation?	
A. Yes	39.2%
B. No	60.8%

Appendix 5. Results of a Survey of Rural Taobao Merchants (Continued)



Appendix 6. Robustness Checks

Notes: This table presents the results of robustness checks of the baseline estimation in Table 3. Columns (1) and (2) report the results based on standard errors clustered at the township level alone and in combination with the month level, respectively. Column (3) uses the inverse hyperbolic sine (HIS) transformation of the number of entrants as the dependent 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 brackets and are based on standard errors clustered at both the store and month levels. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
	Entry [< 0.5 km]				
	SE cluster: town	SE cluster: town-month	IHS Transformation	Store-Product-Month Sample	
Upgrade	0.047***	0.047***	0.004**	0.003**	0.004***
	[3.54]	[2.97]	[2.34]	[2.65]	[2.86]
N	3,048,614	3,048,614	3,048,614	24,270,264	24,270,264
R^2	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 downgraded store ratings (downgrade events) on entry decisions. In our sample period, downgrade events are identified for 520 stores for which precise location information and other essential variables are available. Downgrade is an indicator set to 1 in the month of the downgrade event. Other model specifications follow the convention in Table 3.

	(1)	(2)
	Entry [< 0.5 km]	
Downgrade	0.074	0.056
	[0.58]	[0.39]
N	8,997	8,074
R^2	0.843	0.851
Controls		X
Store FE	X	X
City-Month FE	X	X