# Fintech and Firm Selection: Evidence from E-commerce Platform Lending*

Yi Huang† Ye Li‡ Hongzhe Shan§

July 31, 2018

**Abstract**

This paper provides the first evidence on how fintech powered by big data impacts firm dynamics. With comprehensive data from Alibaba, one of the largest e-commerce platforms and the largest fintech lenders, we estimate the causal impact of platform credit on the size distribution of small businesses. We find that platform credit promotes firm selection – credit leads to stronger growth of market share for online merchants that are larger and better rated by customers. Credit stimulates growth especially when the demand for merchants' products is expanding, and contrary to the conventional wisdom, credit does not affect merchants' product pricing behavior. We also document a dynamic effect: merchants' credit score assigned by Alibaba is highly correlated with their market share and customer ratings. A feedback loop arises between firms' market status and platform credit that accelerates firm selection in this fast growing entrepreneurial space.

**Keywords:** Big Data, Fintech, Firm Dynamics, E-Commerce, Entrepreneurial Finance, Two-Sided Platform, Liquidity, Product Market Competition, Reputation, Firm Quality

*We would like to thank Patrick Bolton, Long Chen, Olivier Darmouni, Bengt Holmström, Tomasz Piskorski, Jun "QJ" Qian, René M. Stulz, Pengfei Wang, and Shangjin Wei for helpful comments. We acknowledge the generous support from Ant Financial and its staff for the data and insightful discussions on the institutional details. We acknowledge the financial support from Bank for International Settlements. **Disclaimer:** The authors highlight that the data and analytics reported in this paper may contain errors and are not suited for the purpose of company valuation or to deduce conclusions about the business success and/or commercial strategy of Ant Financial. All statements made reflect the private opinions of the authors and do not express any official position of Ant Financial and its management. The analysis was undertaken in strict observance of the Chinese law on privacy.

†Graduate Institute of International and Development Studies. E-mail: yi.huang@graduateinstitute.ch

‡Ohio State University. E-mail: li.8935@osu.edu

§Swiss Finance Institute. E-mail: hongzhe.shan@unige.ch

# 1 Introduction

The development of financial technologies ("fintech") has attracted enormous attention, with fintech credit being one of the most prominent areas where big data can add enormous value (Berg, Burg, Gombović, and Puri (2018)). A new breed of lenders rise to compete with traditional players in credit markets, making finance more inclusive and creating business models that demand a new perspective on financial intermediation and regulation (Buchak, Matvos, Piskorski, and Seru (2017); Fuster, Plosser, Schnabl, and Vickery (2018)). On July 31, 2018, the Office of the Comptroller of the Currency, a U.S. banking regulator, started accepting national charter applications from fintech companies, giving them a path to federal oversight.

This paper studies e-commerce and payment platforms as lenders. Platforms, such as Alibaba, Amazon, Paypal, and Square, have extended billions of dollars of credit to their users. As a transaction and data hub, a platform is uniquely positioned to offer credit to online merchants for its advantages in screening, monitoring, and contract enforcement. For example, the Alibaba Group, who owns the largest e-commerce platform in the world, feeds real-time data on the performances of online merchants to its credit-scoring algorithms, screening potential borrowers and monitoring outstanding loans. Those who default can be excluded from the platform, so the value of business continuation enhances borrowers' commitment and enlarges their debt capacity. Such advantages of platform as lender ultimately translate into more accessible credit for small business.[1]

Specifically, we study the allocation of e-commerce platform credit to heterogeneous online merchants and their differential responses. Our findings shed light on the impact of fintech credit on firm dynamics, and in particular, the size distribution of financially constrained firms (Cooley and Quadrini (2001)). More broadly, it offers new insights on the interaction between industrial organization and finance (Phillips (1995); Campello (2006); Fresard (2010)). The last few decades witnessed a startling rise of e-commerce. The U.S. Census Bureau reports that nominal e-commerce spending in the retail sector increased by 216% between 2002 and

[1]In several aspects, platform credit reminds us of trade credit from suppliers (Petersen and Rajan (1997)), but a key difference is that platform credit can be used by online merchants for all purposes, not just expenses on the platform (e.g., advertisement).

2008, while off-line retail spending increased only by 24% (Lieber and Syverson (2012)).[2] In China, online sales increased by more than 25 times between 2008 and 2014. By 2014, 10% of retail sales was online, of which Taobao.com, owned by the Alibaba Group, accounts for more than 82% (Fan, Tang, Zhu, and Zou (2016)).

Our database from Alibaba covers all merchants selling products on Taobao.com, giving a comprehensive picture of online firm dynamics that has not yet been studied by the existing literature. By exploiting a regression discontinuity design, we estimate the causal effect of fintech credit on the size distribution of e-commerce merchants. The size distribution of firms has long been recognized as critical for the organization of production and its efficiency (Simon and Bonini (1958); Lucas (1978)). Enormous efforts have been devoted to document size distribution and its determinants (Evans (1987); Hall (1987); Cabral and Mata (2003); Angelini and Generale (2008)). Our paper is the first to study the size distribution of online small businesses and its response to credit shocks.

E-commerce platforms are viewed by many as business incubators, especially because of the relatively low entry barrier. It is important to understand whether platform credit speeds up the process of firm selection, allowing productive firms to stand out. We find that platform credit leads to greater increases of market share for larger merchants and those with higher customer ratings. In this anonymous online markets, size and customer ratings are two key elements of a merchant's reputation. Platform credit accelerates firm selection, and contributes to the right tail of size distribution, which is in line with the theoretical literature that emphasizes financial friction as a key determinant of size distribution (Albuquerque and Hopenhayn (2004); Clementi and Hopenhayn (2006); Cooley and Quadrini (2001)).

There is a large literature on platform design (Rochet and Tirole (2006)). But credit is ignored as a platform feature. Our findings suggest it is important for the platform to extend credit to online merchants by leveraging on its data advantages. By accelerating firm selection, credit may induce more customers to join the platform, which in turn attracts more merchants through the typical cross-group externality. This effect gives a platform an advantage in the competition with other platforms (Rochet and Tirole (2003)).

[2] Using Visa transaction data, Einav, Klenow, Klopack, Levin, Levin, and Best (2017) document that the share of online spending rose from 12.5% to 22.5% in from 2007 to 2014.

Finally, our study reveals an amplification mechanism that arises from the defining feature of fintech, that is the usage of non-standard data ("big data"). Monitoring the performances of hundreds of thousands online merchants, the platform reacts swiftly to changes of merchants' credit profile. Particularly, the platform benchmarks merchants' performances against their peers when making credit decisions. Outperformers, especially those favored by customers, get picked up by the scoring algorithm to receive more credit from the platform. Therefore, large and reputable merchants not only benefit more from credit, but also have a better chance to receive credit. Next, we will elaborate on our findings.

The starting point of our analysis is a positive correlation between the market-share Herfindahl-Hirschman index (HHI) of a product category and the total platform credit to merchants in the category. Categories range from clothes to digital products, covering every aspect of retail market. An average merchant has a monthly sales of \$6,700 (CNY 45,675), which is close to the average credit line extended by Alibaba to merchants (around \$6,000). The size dispersion is huge. The standard deviation of sales is \$29,000 (CNY 195.970). HHI varies a lot across industries. Clothes and shoes have an HHI equal to 0.01%, while transportation ticket has an HHI of 26.37%. The positive correlation between HHI and credit is robust after we control for category $\times$ time fixed effects, suggesting that given the current conditions of a product category, more platform credit is associated with a higher level of concentration (a longer right tail of size distribution).

To understand how platform credit promotes firm selection, we explore a regression discontinuity design. The credit decision at Ant Financial, the lending affiliate of Alibaba, follows a threshold rule. Firms with a credit score slightly above the threshold obtain credit, while those slightly below do not. Our first step is to estimate the local average treat effect (LATE) of credit access on merchants' market share growth rate. We find a strong positive impact, consistently significant across regression specifications. In the most saturated model, merchants who obtain credit grow $7.14\%$ faster than its peers in the same industry.

We conduct several exercises to validate the identification strategy. First, the observation density is smooth around the threshold, so it is unlikely that merchants manipulate their scores (McCrary (2008)). Moreover, Ant Financial never discloses its credit rule and credit score to merchants.[3] Second, we show that merchant characteristics are smooth around the threshold score (Lee and Lemieux (2010)). Third, we show that loan rate and size are smooth around the threshold score. If the threshold rule were not locally random but rather reflects the lender's recognition of fundamental difference between merchants above and below the threshold, such difference should have manifested into discontinuity in loan rate and size. Last but not least, credit is automatically offered to merchants whenever their score passes the threshold. Firms do not apply for credit. This addresses the concern of selection bias.

To understand the impact of credit on size distribution, we separate merchants into sizes percentiles, and find a stronger positive impact of credit on market share growth rate for larger merchants, suggesting that credit promotes firm selection. This finding echoes the industry-level correlation between credit amount and HHI. It is exactly the opposite to the conventional wisdom that the marginal impact of credit is larger for smaller firms who tend to be more financially constrained. In an anonymous online market, customers infer a merchant's credibility largely by size. A more reliable merchant faces a stronger demand for its product and better expansion opportunities. Indeed, once we sort merchants by customer ratings, we find the credit impact on market share growth rate is stronger for better rated merchants. Platform credit amplifies the voice of customers, helping reputable merchants to stand out.

To understand credit transmission mechanism, we examine the impact of platform credit on merchants' market share under different industry conditions, and to sharpen our inference, we explore a difference-in-difference setting. First, we find that credit impact is stronger on when the product category experienced strong growth in sales. Conventional wisdom predicts the opposite – credit matters more when firms' cash flow is low (e.g., Chevalier and Scharfstein (1996)). However, such models usually leave out time-varying investment opportunities. When industry sales have momentum, which we confirm in data, merchants need more credit to meet the expanding demand for their products. For causal inference, we compare the treatment effect of platform credit on firms' market share in November, the key promotion month of Alibaba (positive demand shock), to the impact in other months (difference-in-difference). The difference is positive and significant, suggesting that when product demand and cus-

[3] It is difficult for the merchants to back out the algorithm, which combines various credit analysis and machine learning techniques, especially given that merchants do not observe their credit scores.

tomers' attention are higher, credit is more effective in helping merchants gain market share.

A particular link between firms' product market performance and access to credit (or liquidity in general) is through product pricing. For example, Bolton and Scharfstein (1990) predict that firms in favorable liquidity conditions lower their price to fend off competitors. Data on product prices is usually difficult to obtain, so evidence on how credit affects product pricing is from particular industries (Chevalier (1995); Phillips (1995)). We observe all product prices, and do not find evidence on such predatory behavior. Rather, the positive impact of credit access on merchants' sales is purely through an increase of quantity sold.

So far, we have characterized how platform credit shapes the size distribution of small businesses online. But what is special about credit powered by fintech or big data? We find that the utilization of non-standard data induces a *dynamic interaction* between platform credit and the size distribution. Merchants' credit score is highly correlated with their market share and customer ratings. In other words, larger and better rated ones are more likely to obtain credit. Given that they already benefit more per unit of credit, a feedback loop arises between market status and platform credit. Moreover, we differentiate two channels of a platform's information acquisition, the customer channel (a merchant's customer ratings) and the platform's proprietary data analysis (approximated by a merchant's cumulative transactions on the platform). We find these two are substitutes in the sense that their interaction term enters the credit-score regression with a negative coefficient. Therefore, even though the dynamic effect of big data accelerates firm selection, data exhibit decreasing return to scale.

**Literature.** Finance is being transformed by the big data revolution. Financial technologies ("fintech") have the potential to reduce the cost and risk of finance (Philippon (2017)). Recent development in credit markets is an leading example. Edelberg (2006) and Einav et al. (2013) study fintech consumer credit. Buchak, Matvos, Piskorski, and Seru (2017) and Fuster, Plosser, Schnabl, and Vickery (2018) focus on online mortgage lending. Berg, Burg, Gombović, and Puri (2018) show that nonstandard data help forecast default. Using data from Ant Financial, previous studies estimate the credit impact on firms' sales growth (Hau, Huang, Shan, and Sheng (2018)) and service quality (Huang, Lin, Sheng, and Wei (2018)). This paper provides the first evidence on how fintech credit, and big data revolution in general, affects industrial organization. Following Petersen and Rajan (2002), we focus on small businesses that are most likely to be financially constrained, and study the evolution of size distribution.

Beyond the literature on firm size distribution that we refer to at the beginning of introduction, our findings on how size distribution responds to credit shocks shed light on the dominantly theoretical literature on credit frictions and wealth/size distribution (Aghion and Bolton (1997); Matsuyama (2000); Piketty (1997); Cooley and Quadrini (2001); Clementi and Hopenhayn (2006); Moll (2014)). Wealth distribution has attracted enormous attention (e.g. Piketty and Saez (2003)). Recent studies unveil rich implications of firm size distribution, especially the rise of superstar firms (Rosen (1981)), on various aspects of economic activities from labor income share (Autor, Dorn, Katz, Patterson, and Reenen (2017)) to the allocation of talent (Choi, Lou, and Mukherjee (2017)).

This paper also contributes in three aspects to the literature on industrial organization and credit (or liquidity in general, e.g., Fresard (2010)). First, using a regression discontinuity design, we estimate the *causal* impact of credit on product market competition. Second, our study utilizes data on firm-level product prices that cover almost every aspect of retail market. The current literature largely focuses on event studies and particular industries (Chevalier (1995); Kovenock and Phillips (1997)), commonly assuming predetermined capital structure of firms and using industry-level product prices (Campello (2003), (2006)). Third, we characterize the feedback effect of data-driven credit decisions by the platform and firms' product market performances.[4] Parsons and Titman (2007) provide a survey on the relation between debt financing and corporate strategy.

Finally, this paper offer new insights on the design of online market. Platform credit, powered by data advantages, accelerate firm selection, and therefore, by affecting the production/seller side, platform credit has impact on the growth of consumer side, as suggested by the effect of cross-group externality (Rochet and Tirole (2006)). Existing studies on online market structure (firm size distribution) ignores the financing aspect. For example, Noe and Parker (2005) find that the winner-takes-all (i.e., right-tail) payoff structure of online markets induces firms to incur large advertisement expenditures. Brynjolfsson et al. (2010) argue that

[4] MacKay and Phillips (2005) highlights a different feedback mechanism: while a firm's financing structure affects product market competition , it in turn depends on industry characteristics.

online market fosters both niche products and superstars (right tails). Related to this paper, Fleder and Hosanagar (2009) examine how big data techniques affect online market structure, but focus on the widely used recommender system and its impact on product diversity.

# 2 Institutional Background and Data Description

## 2.1 Alibaba and Taobao

Founded by Jack Ma in Hangzhou (China), Alibaba Group is a tech-service conglomerate that provides internet infrastructure, digital platforms, and technology solutions to various clients. Its core business focuses on e-commerce platforms, and the associated services, such as cloud computing, logistics, and financial services. Its great commercial success in e-commerce led to a public listing in 2014 on the New York Stock Exchange. The e-commerce platforms encompass both wholesale and retail marketplaces for firms and consumers home and abroad, serving more than 10 million active sellers and 454 million active buyers per year (Ali Research, 2017). Alibaba Group owns three platforms called Alibaba, Tmall, and Taobao.

The merchants selling products on Alibaba's platforms are largely small businesses. According to Ali Research (2017), about 97% of the merchants on Alibaba platforms employ less than 5 workers, and 95% of them invest less than \$4,500. Among the three platforms, Alibaba is a business-to-business platform, while Tmall and Taobao are consumer-driven. By 2016, the gross merchandise trading volume (total trading value) in the two trading platforms Tmall and Taobao exceeded 3 trillion RMB a year, which amounts to 4% of Chinese GDP. On Tmall, brand name merchants sell products directly to customers. Taobao is distinguished by its focus on small merchants, who do not have an established brand name, and usually cannot access traditional credit from the banking system. Our study uses data on Taobao merchants.

The e-commerce platform generates a huge volume of data, thanks to its product diversity that ranges from computer games to basic manufacturing products for household uses. Database is a core asset of Alibaba Group, based upon which its financial services affiliate, Ant Financial, make credit decisions.

## 2.2 Ant Financial

Merchants conduct transactions using a payment system, called "Alipay", owned by Alibaba Group. Alipay was launched in 2004. The success of Alibaba's platforms is often attributed to the design of escrow accounts in the Alipay system, because online trading counterparties are anonymous and usually do not long-term relations off-line. Given the relatively low confidence on the part of consumers in merchant credibility, securing online payments through escrow accounts turned out to be a very important business design. By 2014, Alipay had become the worlds largest mobile and online payment platform and accounted for approximately half of all online transactions in China.[5]

In 2015, Alipay was merged into a new company, Ant Financial Service Group, that not only provides payment services but also extend credit to merchants on Alibaba platforms. Ant Financial conduct its lending business through a subsidiary, MYbank, that obtained the banking license from CBRC (China Banking Regulatory Commission) in 2014. MYbank largely finances its lending by tapping into the money market and asset-backed security market, and thereby, is considered as shadow banking operation by the regulatory authorities in China. Its lending decision is largely based on the huge volume of data from Alipay that include transactions between merchants and customers and transaction between merchants and the Alibaba Group (for example, merchants' investing in advertisement on the platform). Our study focuses on the loans extended by Ant Financial to merchants on Taobao platform, and our sample starts from August 2014 and ends in June 2015.

**Lending at Ant Financial.** Ant Financial extended unsecured credit line to merchants operating on Taobao.com. A credit score is assigned to each merchant. The scoring model incorporates a large scale of data from the various platforms of Alibaba, including the merchants trading behaviors, operational conditions, and financial performances from Alibaba's e-commerce platforms, especially their payment records and past credit records (e.g., historical default information) from Ant Financial. While the specific algorithms are unknown to us, the most important consideration hinges on the recent sales performance of the merchants, and

[5]See Bobsguide, February 12, 2014: http://www.bobsguide.com/guide/news/2014/Feb/12/alipay-surpassespaypal-as-leading-mobile-payments-platform/

the algorithms include not only traditional credit analysis models but also machine learning algorithms, from the basic support vector machine to random forest. A credit score is assigned, typically in a range from 380 to 680, and it changes instantaneously when new data feed in. There exists a fuzzy discontinuity in the decision rule: firms with score higher than 480 are significantly more likely to obtain credit. We discard the data in December 2014 because the 480 cutoff was briefly suspended at Ant Financial. We explore such discontinuity for causal inference. Note that merchants do not observe their credit scores, and are not aware of such score-based rule of credit approval at Ant Financial.

*[ Figure 1 here ]*

Figure 1 plots the evolution of total loans outstanding. Ant Financial extends loans in the form of credit line with a maturity of one year. Credit is offered to online merchants once their credit profiles meet the standards. Merchants may pay back loans before maturity whenever they prefer. In case of delinquency or default, Ant Financial may coordinate with Taobao.com to terminate the online operation of the delinquent merchant, and impose an exclusion penalty that may last forever. The potential exclusion from the largest e-commerce platform serves as a threat that incentivizes merchants to repay these unsecured loans.

# 3 Methodology

## 3.1 Fuzzy discontinuity design and its validity

To identify the causal effect of credit access on various outcome variables, we explore a fuzzy discontinuity in the score-based lending rule of Ant Financial. Figure 2 shows the distribution of credit scores for all Taobao merchants that had passed the first-step screening. For any given credit score, the blue bars show the average number of firms approved for credit in our sample, and the red bars show the average number of firms without credit approval. A discontinuity appears at the score of 480.

*[ Figure 2 here ]*

While the probability of obtaining credit is discontinuous at the score of $480$, the discontinuity is not sharp. Some firms with a credit score above 480 still do not obtain credit, and others with a score below 480 are nevertheless eligible for credit, There are two reasons. First, the credit decision of Ant Financial depends on information beyond the credit score. Unfortunately, we cannot access the complete set of information at Ant Financial. The second issue is data frequency. Ant Financial makes credit approval decisions on daily basis, but we only obtain data at the end of each month. Firms have a credit score higher than $480$ and obtain credit in the middle of the month, but their score drops below 480 at the end of the month. In the next subsection, we validate the fuzzy discontinuity design from various angles.

When properly implemented, the regression discontinuity design yields an unbiased estimate of the local treatment effect (Shadish, Clark, and Steiner (2008)). Let $Y_{i,j,t}$ denote the outcome variable of firm $i$ in month $t$ and industry $j$, for example, the growth rate of firm $i$'s market share. Using credit score ("$CS$") as the running variable and its value of $480$ as threshold, we define an indicator variable $I_{\{CS_{i,t} \geq 480\}}$ for firm $i$ in month $t$ that equals one if $CS_{i,t} \geq 480$ and zero otherwise. To implement the fuzzy discontinuity design, we use two-stage least-squares to estimate the treatment effect (Hahn, Todd, and der Klaauw (2001); Lee and Lemieux (2010)). Porter (2003) discusses the attractive properties of regression discontinuity estimation by local linear regression. Our local bandwidth of credit score is $[460, 500]$.[6]

In the first stage, we estimate the probability of treatment (i.e., obtaining credit access) by regressing an indicator variable of credit access, denoted by $X_{i,j,t}$, on $I_{\{CS_{i,t} \geq 480\}}$, and a third-order polynomial of credit score that is recentered at $480$. We include the polynomial because the estimated effects are only unbiased if the functional form of the relationship between the treatment and outcome is correctly modeled. The most popular caveats are non-linear relationships that are mistaken as a discontinuity. Gelman and Imbens (2017) discuss how to choose the order of polynominal.

We examine specifications that control for industry fixed effects, month fixed effects (or industry $\times$ month fixed effects) and firm characteristics. As summarized in Imbens and Lemieux (2008), covariates can be used to eliminate small sample biases present in the basic

[6]Imbens and Kalyanaraman (2012) discuss the optimal bandwidth for regression discontinuity estimation by local linear regression,

specification and improve the precision (see Frölich and Huber (2018) for a recent discussion). Moreover, if parameter estimates are sensitive to removing or adding covariates, then this may cast doubt on the validity of the regression discontinuity design (Lee and Lemieux (2010)).

Let $C_{i,t}$ denote firm characteristics as control variables, and $FE_{j,t}$ denote the fixed effects for industry $j$ and month $t$, which can be either the sum of industry and month fixed effects or the industry $\times$ month fixed effects. The first stage regression is formalized below:

$$X_{i,j,t} = \alpha + \tau I_{\{CS_{i,t} \geq 480\}} + \sum_{k=1}^{3} \delta_k CS_{i,t}^k + FE_{j,t} + C_{i,t} + \epsilon_{i,j,t}^X$$

In the second stage, the indicator variable, $\tau I_{\{CS_{i,t} \geq 480\}}$ is the excluded variable:

$$Y_{i,j,t} = \mu + \beta \hat{X}_{i,j,t} + \sum_{k=1}^{3} \lambda_k CS_{i,t}^k + FE_{j,t} + C_{i,t} + \epsilon_{i,j,t}^Y.$$

Following Imbens and Lemieux (2008), we use the same bandwidth, i.e., $[460, 500]$, and order of polynomial for both stages of regressions.

## 3.2 Validating local randomization

We conduct several exercises to validate the discontinuity design. First, in our setting, credit is *offered* by Ant Financial to merchants whenever they qualify. Merchants do not voluntarily apply credit from Ant Financial. This addresses the concern of selection bias. We observe all four types of firms (interacting whether credit is needed and whether credit is obtained).

*[ Figure 3 here ]*

Regression discontinuity design is invalid if agents can precisely manipulate the running variable, that is, more generally, whether $I_{i,t}$ reflects firms' choice. Manipulation is very unlikely in our setting. Ant Financial never discloses to its borrowers how credit score is constructed, and Taobao merchants do not know their scores or the threshold rule used by Ant Financial. The algorithm is extremely difficult for potential borrowers to guess the algorithm, given that it combines various machine learning and credit analysis models. Therefore, credit access is unlikely to be subject to borrowers' manipulation. Moreover, we examine the density of observations following McCrary (2008). Figure 3 shows that the distribution of firms across credit scores. We distinguish firms below and above the cutoff using different colors to shade respective density areas. As shown by the plot, firms are smoothly distributed around the threshold score of 480. Specifically, we do not observe a greater density of firms right above the threshold. Therefore, there is no evidence that firms can precisely manipulate the credit score around the cutoff.

*[ Figure 4 here ]*

The second concern is that the threshold score of 480 is *designed to be special* by Ant Financial so that firms on different sides of the threshold are recognized by the lender as fundamentally different. If such a difference exists, it may show up in firms' product market performances, and market share in particular. We address this concern by plotting the average loan size and rate for each 5-point bin of credit score in Figure 4. The range of $[460, 500]$ is highlighted, and in the middle, the cutoff value of 480 is marked. If firms on two sides of the cutoff value are fundamentally different and recognized by Ant Financial in its credit score calculation, we would expect discontinuity in loan size and rate because such difference is likely to matter not only for granting access or not but also other aspects of the loan. Both loan size and rate show continuity in the interval, suggesting this is unlikely to be the case.

*[ Figure 5 here ]*

Our hypothesis focuses on how the *access* to credit affects firms' outcome variables. A potential concern over our identification strategy is that the threshold policy of Ant Financial may not only apply to credit access, but also to the size and cost of credit (loan rate). If so, the treatment effect we find may be a mixed impact of access, loan size, and loan rate. Figure 4 also addresses this concern by showing the continuity of loan size and rate. Moreover, Figure 5 shows that for around $80\%$ of firm-month observations, firms who can access credit do not exhaust the full credit line. This suggests that in most cases, once firms obtain credit access, their financial constraint is relaxed, so the size of credit line is no longer a binding concern for them. Note that for more than 60% firm-month observations, firms do not draw on the credit line at all. This does not mean that credit access is not important for those firms, because in a dynamic setting, firms' decisions are affect by the availability of credit line as a source of liquidity. Overall, we conclude that our 0-1 variable, credit access, is likely to be the most important aspect of platform lending for from firm size distribution.

*[ Figure 6 here ]*

Following Lee and Lemieux (2010), we plot several merchant characteristics in Figure 6 to show that around the threshold score of 480, firm characteristics move smoothly with credit score, suggesting that the discontinuity cannot be attributed to firm-level differences other than the threshold-rule of credit allocation.

Lastly, we discuss external validity. One important aspect of the regression discontinuity design is that at best, it provides estimates of the average treatment effect for a subpopulation of the compliers. Lee (2008) points out that the more the running variable is measured with error about the individuals "true type", the more generalizable the regression discontinuity estimate is. The intuition behind is very simple, if the running variable contains more randomness that is exogenous, the threshold-based assignment is closer to a randomized experiment. In our setting, Ant Financial revised their credit score algorithm in July 2015. This suggests that the previous credit score system, based on which we formulate the discontinuity design, had not adequately reflected the credit profile of borrowers, and thus, is likely to contain errors or exogenous randomness that favors the generalizability of our estimate.

## 3.3 Heterogeneous treatment effect

The key outcome variable of our study is market share. After establishing the causal effect of credit access on a firm's market share, we are interested in whether the effect differs by firm size. If the credit impact on market share is weaker for larger firms, credit is an equalizing force; otherwise, it accelerates firm selection by increasing concentration. Therefore, we estimate the heterogeneous treatment effect.

We adopt a parametric approach. Specifically, we augment the second-stage regression with an interaction term. The interaction variable, $Z_{i,j,t}$, itself which may vary by time ("$t$"), firm ("$i$"), or industry ("$j$"):

$$Y_{i,j,t} = \mu + \beta_1 \hat{X}_{i,j,t} + \beta_2 \hat{X}_{i,j,t} \times Z_{i,j,t} + \beta_3 Z_{i,j,t} + \sum_{k=1}^{3} \lambda_k CS_{i,t}^k + FE_{j,t} + C_{i,t} + \epsilon_{i,j,t}^Y.$$

Note that since our interaction variable is either the percentile ranking of merchants' market share in the previous month or the merchants' customer ratings (ranging from one to five), adding the interaction terms is equivalent to separately estimating the treatment effects in subsamples divided by market share percentiles and customer ratings respectively.

# 4 Results

**Sample description and summary statistics.** Our sample of firm-month observations spans August 2014 to June 2015. Observations in December 2014 are discarded because the 480 threshold rule was temporarily suspended in that month. In our analysis, we often use the past three-month growth rate of sales at firm or industry (product category level) as right-hand side variables, so the sample for regression discontinuity analysis has a total of seven months of data, which starts in November 2014 with the aforementioned right-hand side variable constructed from observations in August to October 2014 for the starting month.

Another irregularity of our sample is that credit score is only available starting September 2014. This does not affect the regression discontinuity analysis since it starts in November 2014. However, it affects our last exercise, the full-sample projection of credit score on firm characteristics. Therefore, the sample for credit score project (9) is September - November 2014, and January - June 2015.

Table 1 provides the summary statistics. Column (1), (2), and (3) provide statistics for the full sample, i.e., all firm-month observations in August - November 2014, and January - June 2015. Column (4), (5), and (6) provide statistics for the regression discontinuity sample, i.e., all firm-month observations with credit score in $[460, 500]$, the local band around the credit cutoff value of 480, in November 2014 and January - June 2015. Panel A shows that in the full sample, 77.5% firm-month observations have credit access, and 63.8% in our sample for regression discontinuity analysis. The average size of credit line is CNY 40,879 (approximately $6,000) in the full sample, and CNY 25,767 (approximately $ 3,800) in the sample for discontinuity analysis. The credit usage is 14.4% of credit line on average in the full sample, and 20.9% in the sample for discontinuity analysis.

In Panel B Column (2), the average monthly sales is CNY 45,675 (approximately $6,700), which is close to the average size of credit line. The standard deviation of sales is large, CNY 195,970 (approximately $29,000) in Column (3). Moreover, the size distribution is highly skewed. Figure 7 plots the firm size distribution by counting the number of firms in each CNY 1,000 bin of monthly sales. Firms with sales larger than CNY 150,000 (approximately $22,000) are summed up in the last bin. The distribution is calculated every month, and then, averaged over time. Taobao platform is dominated by small enterprises, and as shown in the appendix, most product categories belong to retail markets. In Panel B, we also provide statistics for other variables used in our analysis, such as firms' market share and customer ratings (on a scale from 1 to 5).

**Outline and result overview.** We examine the impact of credit on firm size distribution by taking the following steps. First, as a simple motivation, we show that as the market-share HHI is positively correlated with the total amount of credit to a product cateogry (Table 2). Next, our analysis moves forward to firm-level observations. Using the fuzzy discontinuity design, we establish a positive causal effect of credit on firms' market share growth rate (Table 3), and find that the treatment effect is heterogeneous – credit helps disproportionately the larger firms and those with higher customer ratings (Table 4 and 5). This echoes the positive correlation between industry-level HHI and the amount of platform credit.

To understand the credit transmission mechanism, we examine how credit impact varies with industry conditions (Table 6). We find that credit matters more for firms' market share growth when an industry is growing or experiences positive demand shocks. Moreover, we find that firms do not change product prices after obtaining credit (Table 7), so the positive impact of credit on market share works through quantity, i.e., more products sold (Table 8).

Next, we look into how the dynamic effect of platform credit. We find that a firm's credit score is positively correlated with relative performances, such as market share and customer ratings (Table 9). This finding suggests a positive feedback loop between market status and credit that accelerates firm selection.

## 4.1 Credit and market share

Table 2 reports the results of a simple regression of a product category's HHI on the ratio of its total credit line (normalized by sales). The positive correlation is consistently significant across regression specifications, suggesting that more credit is associated with a longer right tail. To understand this industry-level correlation, we try to identify the the causal effect of platform credit on size distribution. In the following, we explore a regression discontinuity design for our analysis of firm-month observations.

In the first-stage regression of the fuzzy discontinuity design, we instrument credit access with an indicator function that equals one when the firm's credit score is above 480. Note that credit access, a 0-1 variable, is at the end of the current month, while a firm's credit score is obtained at the end of the previous month. This is to leave a period of time for the lender, Ant Financial, to respond to changes in merchants' credit profile. Also, using lagged variable as instrument alleviates potential concerns of endogeneity.

In the second stage, the dependent variable is the firm's sales growth in the current month and next month relative to the growth rate of its industry during the two-month period,

$$Y_{firm,t+1} = \Delta \ln \left(Sale_{firm,t+1}\right) - \Delta \ln \left(Sale_{industry,t+1}\right).$$

We consider both the current month and next month because our observations of credit access are at the end of every month. If a firm obtains credit at the month end, the credit impact happens in the next month.

A firm's sales growth relative to the industry average is the growth rate of market share

(Campello (2003); Fresard (2010)):

$$\Delta \ln \left(Sale_{firm,t+1}\right) - \Delta \ln \left(Sale_{industry,t+1}\right) = \ln \left(\frac{Sale_{firm,t+1}}{Sale_{industry,t+1}}\right) - \ln \left(\frac{Sale_{firm,t}}{Sale_{industry,t}}\right).$$

A positive impact of instrumented credit access in month $t$ on $Y_{firm,t+1}$ means that on average, firms with credit access outgrow their peers in the same product category.

We focus on the relative size of firms within a product category instead of the whole platform, because across categories, the type of products is very different. It is unlikely that firms in different categories (or "industry") compete directly with each other, or that a firm operates in two or more categories at the same time. As previously discussed, we adopt the product categorization by Alibaba that use it to facilitate the management of its platform. In the appendix, we list all the categories.

*[ Table 3 here ]*

Table 3 reports the results. Column (1), (2), and (3) show the results of second-stage regression with the right hand side including respectively the industry and month fixed effects, industry $\times$ month fixed effects, and these fixed effects together with firm characteristics. The corresponding first-stage results are reported in column (4), (5), and (6).

The industry $\times$ month fixed effects in Column (2) and (3) control for industry-level unobservables that potentially vary over time, such as shocks to consumer demand in a product category, the conditions of upstream (supplier) sectors, and variation in government regulation and intervention targeted at particular products. For the relevant firm characteristics, we include as many from the database as possible: the logarithm of sales in month $t$, sales growth in the past quarter, and customer ratings (more details on these later). Firms' off-line financial resources and sales may also influence firms' online performances, but obtaining such information is challenging, especially given that Taobao vendors are dominantly small, private firms. As a future direction of research, we will investigate how online resources and business transactions interact with online performances and credit.

*[ Figure 8 here ]*

The first-stage results show a very strong relation between our threshold indicator and credit access, and the relation is numerically stable across different specifications of the right-hand side. We visualize the first-stage results by plotting in Figure 8 the percentage of firms that become eligible for credit lines for different levels of credit score ("*CS*") in the range of [460, 500] (1-point bins). We fit a third order polynomial to both sides of the 480 threshold value. There is a jump of probability of credit access by more than 20% at the threshold. Note that even though in our regressions, the first stage has fixed effects and other firm characteristics as control variables, the estimated coefficient on the instrumented indicator in Table 3 is similar to the jump in Figure 8. The fact that control variables do not have significant impact lends further support to the threshold indicator as an instrument variable.

We are interested in the impact of credit access on a firm's market share growth to credit access. Specifically, in the second stage, we test the following hypothesis:

**H1**: Firms outgrow their peers in the same product category when they are granted credit.

In the basic specification (Column (1)) of Table 3, on average, a firm grow $6.12\%$ faster than its peers if it can access the credit provided by the platform. In Column (3). the most saturated specification, all else equal, the relative sales growth caused by credit access is larger, $7.41\%$, and significant. This treatment effect is large in magnitude given an average sales growth rate of 20.6% (Table 1). Figure 9 plots the average market share growth rate for each 1-point bin of credit score, showing a jump at the threshold value that corresponds to an estimated treatment effect from a sharp discontinuity approach. The size of jump, c.6%, is close to our estimate of treatment effect from the fuzzy discontinuity approach in Table 3.

*[ Figure 9 here ]*

As previously discussed, there can be three potential concerns over our identification strategy. The first is whether a firm can manipulate its credit score to be just above 480. Such manipulation is very unlikely because Ant Financial never discloses to its borrowers how credit score is constructed and the threshold rule. The scoring-algorithm extremely difficult to be backed out since it combines various machine learning models, from the basic support vector machine to more advanced random forest, and traditional credit analysis models. Therefore, manipulation by Taobao firms is unlikely.

The second concern is that the threshold score of 480 is *designed to be special* by Ant Financial so that firms on different sides of the threshold are recognized by the lender as fundamentally different. If such a difference exists, it may show up in firms' product market performances, and market share in particular. We address this concern by plotting the average loan size and rate for each 5-point bin of credit score in Figure 4. The range of $[460, 500]$ is highlighted, and in the middle, the cutoff value of 480 is marked. If firms on two sides of the cutoff value are fundamentally different and recognized by Ant Financial in its credit score calculation, we would expect discontinuity in loan size and rate because such difference is likely to matter not only for granting access or not but also other aspects of the loan. Both loan size and rate show continuity in the interval, suggesting this is unlikely to be the case.

Our hypothesis focuses on whether or not a firm can access platform credit. A third concern over our identification strategy is that the threshold policy of Ant Financial may not only apply to credit access, but also to the size and cost of credit (loan rate). If so, the treatment effect we find may a mixed impact of access, loan size, and loan rate. Figure 4 also addresses the third concern by showing the continuity of loan size and rate. Moreover, Figure 5 shows that for around 80% of firm-month observations, firms who can access credit do not exhaust the full credit line. This suggests that in most cases, once firms obtain credit access, their financial constraint is relaxed, so the size of credit line is no longer a binding concern for them. Note that for more than 60% firm-month observations, firms do not draw on the credit line at all. This does not mean that credit access is not important for those firms, because in a dynamic setting, firms' decisions are affect by the availability of credit line as a source of liquidity. Overall, we conclude that our 0-1 variable, credit access, is likely to be the most important aspect of platform lending for from firm size distribution.

## 4.2 Credit and firm selection

As demonstrated in Table 3, credit gives firms a competitive advantage. How credit affects firm size distribution crucially depends on the heterogeneity of its treatment effect. Conventional wisdom suggests that credit has stronger impact on smaller firms that tend to face more difficulties in financing. This implicitly assumes that small firms faces similar (if not better) investment opportunities as big firms. However, in our setting of anonymous online market, size is a key indicator of credibility or product quality.(e.g., Cabral and Hortaçsu (2010)).[7] So, big firms may have better opportunities to expand business because they face stronger product demand, and the credit impact on their market share growth can be stronger.

If firms are larger and receive better ratings because of higher productivity, concentration simply reflects the process of resources being allocated towards more efficient producers, admittedly with potential inefficiencies from monopolistic behavior. However, equally productive but smaller firms can simply be unlucky in their past sales or ratings.[8] In such cases, history dependence is an inefficient force.

Direct evidence on the causal impact of credit on firm size distribution is rare. A credit *shock* is difficult to obtain. Here, we provide evidence on the distributional effect of credit in an increasingly important entrepreneurial space – the online retail market. As previously discussed, our discontinuity design resembles a semi-experimental setting. Moreover, from the perspective of e-commerce platform design, it is important to understand how credit as a platform feature affects online market structure. Specifically, we test the following hypothesis:

**H2.1**: Platform credit has stronger positive impact on larger firms' market share growth.

**Credit effect and size.** To capture the cross-section heterogeneity, we rank firms by sales (or equivalently, market share) within a product category in each month. Firms with higher market share percentile rank are larger in sales relative to their peers. We also measure the size of firms simply by the logarithm of sales. Following the parametric approach laid out in Section 3, we are interested in the coefficient of the interaction between instrumented credit and firm size. A positive coefficient suggests the impact of credit access on market share growth rate is higher for larger firms, that is credit promotes firm selection. We also record firms' sales growth rate in the past quarter, and interact it with the instrumented credit access to examine whether growth momentum is amplified by credit access.

*[ Table 4 here ]*

[7]Cabral (2012) reviews the literature on reputation in online environments.

[8]A particular example of history dependence is information cascade (Welch (1992)).

Table 4 reports the results. In Column (1), (4), and (7), we use the percentile rank of firms' market share as a measure of size to interact with the instrumented credit access, and control respectively, industry and month fixed effects, industry $\times$ month fixed effects, and in the most saturated specification, both industry $\times$ month fixed effects and the firm characteristics from Table 3. The coefficient on the interaction term is statistically significant across specifications, with the largest magnitude from the most saturated specification (Column (7)).

In Column (1), the basic specification, a firm ranked one decile higher by market share gains a growth rate that is 4.107% higher than its peers. This suggests that credit has a stronger positive impact among larger firms. This coefficient of interaction term and the coefficient of instrument credit alone (2.13%) almost perfectly decompose the coefficient of credit access in Table 3, the average treatment effect. Note that our dependent variable is a growth *rate*, so the fact that larger firms tend to have larger credit lines, borrow more, and grow more in absolute terms, should not be a concern.

In Column (2), (5), and (8), we replace the market share percentile rank with $\ln(sales)$ as a measure of firm size for a robustness check, and interact it with the instrumented credit. The coefficient of interaction term is positive, and statistically significant across specifications. In terms of economic magnitude, a firm's credit access adds 1% to growth relative to peers when the firm's sales increase by 20%.

In Column (3), (6), and (9) of Table 4, we examine the impact of credit on the momentum of firms' growth. Specifically, we interact the instrumented credit access with the growth of sales in the past quarter. Same as before, Column (3) reports the basic specification with industry and month fixed effects, while Column (6) controls for industry $\times$ month fixed effects and Column (9) control additionally the firm characteristics as in Table 3. All three specifications exhibit growth momentum as the coefficients on the standalone term of firms' past growth rate are all positive and statistically significant. In Column (3), a 20% increase in a firm's past-quarter growth rate translate into more than 1% faster growth relative to other firms in the same product category.

We are interested in the coefficient of the interaction term: a positive estimate suggests that credit amplifies the momentum. In Column (3), credit access amplifies the momentum by around 50%. The effect varies across specifications, with the largest magnitude from the most saturated specification with both industry $\times$ month fixed effects and control variables (Column (9)) – credit access amplifies growth momentum relative to peers by 9 times.

Overall, our findings suggest that platform credit promotes firm selection. Standard models of competition suggest firms with cost advantage have large market shares. For example, in Bertrand competition, the firm with lowest marginal cost of production sets price and dominates the market. Platform credit can be a force that removes financial frictions and facilitates this natural process of firm selection.[9]

**Credit effect and reputation.** Next, we examine whether credit benefits reputable firms more. Reputation is a key determinant of a firm's online performance. Using eBay data, Cabral and Hortaçsu (2010) estimate the response of sales to negative online ratings, the serial correlation of negative rating, and how negative ratings lead to exit. Using data from the B2B platform of Alibaba Group, Chen and Wu (2017) find that sales of Chinese T-shirt exporters are concentrated among superstar firms. Using data from Taobao, Fan et al. (2016) find substantial return to reputation. The reputation system on Taobao.com is largely based on customer ratings. Specifically, based on their shopping experience, customers rate the quality of merchandise, service, and delivery on a scale from 1 to 5. We test the following hypothesis:

**H2.2**: Platform credit has stronger positive impact on better rated firms' market share growth.

*[ Table 5 here ]*

We examine the heterogeneous effect of credit by interacting the ratings from last month with instrumented credit access. Table 5 reports the results. In Column (1), (4), and (7), we focus on merchandise rating, and control respectively the industry and month fixed effects, industry $\times$ month fixed effects, and firm characteristics together with the interacting fixed effects. The estimates of coefficient of the interaction between merchandise rating and instrumented credit access are positive and significantly across specifications. Firms with more reputable products gain more market share when they obtain their credit. In Column (1), the

[9]Even if we entertain the potential heterogeneity of products within a category, for example, in the form of Dixit and Stiglitz (1977), firms with lower production cost still have larger sales in monopolistic competition.

basic specification, credit access adds 16.70% more to a firm's market share growth rate when the firm's merchandise rating increases by one. Once we add the interaction between merchandise rating and credit access, the coefficient of credit access alone become insignificant.

Similar results are obtained with service rating (Column (2), (5), and (8)) and delivery rating (Column (3), (6), and (9)). The right-hand side specifications follow that of merchandise ratings. The coefficients of the interaction between rating and credit access are positive and statistically significant across specifications, and large in magnitude. For example, in Column (2), credit leads to 17.32% more growth in market share for firms whose service rating is one level higher, and in Column (3), an one-level increase in delivery rating augments the credit effect on market share growth by 18.61%. What we find are consistent with the hypothesis that more reputable firms gain larger competitive advantage from access to credit as they tend to have better expansion opportunities.

Overall, our results suggest that platform credit amplifies the voice of customers. Since there is a strong correlation between firm size and customer rating, this finding refines our understanding of how platform credit shapes the firm size distribution. From the perspective of platform design, if credit allows good firms to stand out, it is a desirable feature of e-commerce platform, especially due to the positive cross-side externality – good sellers tend to attract more buyers to join the platform (Rochet and Tirole (2006)).

## 4.3 Credit transmission mechanisms

So far, we have characterized how platform credit powered by big data affects on the firm size distribution. In particular, we show that credit benefits disproportionately larger and more reputable firms. It is important to understand the mechanisms through which credit enhances firms' competitive advantages. We approach this question from two different angles.

First, we distinguish credit impact on firms' market share in good and bad industry conditions. We find that in contrast to the traditional wisdom, credit matters more in when the industry is growing, especially due to demand shocks (Table 6), suggesting that credit access may help firms capture new customers during a period of demand growth and heightened consumer attention. Next, we find that credit access does not affect firms' product pricing

(Table 7), which indicates that other channels of credit usage, such as advertisement expense and inventory build-up, play more prominent roles in market share growth.

**Credit impact and demand shocks.** There is a large literature that studies the cyclicality of firms' market power (Rotemberg and Woodford (1999)). In particular, the fluctuation of product demand is supposed to influence firms' investments in market share. The traditional wisdom is that the expansion of product demand increases firms' cash flow, and thereby, weakens the connection between firms' financing conditions and investments in market share (Chevalier and Scharfstein (1996); Campello (2006)). This implies that the impact of credit on market share is stronger during a contraction of product demand.

However, this cash-flow-centric argument ignores the time-varying investment opportunity set of firms. When product demand grows, a firm has opportunities to capture new customers, which has persistent effect on its market share when customers face switching costs à la Klemperer (1987) and Farrell and Shapiro (1988) (and more recently, Dou, Ji, Reibstein, and Wu (2018)). When product demand declines, many customers retreat from the market, so any effort to capture customers, for example through advertisement, tends to be less effective. Next, we test our hypothesis that credit matters more in booms. Because of sample spans a relative short period of time, we explore the cross-sectional difference across industries in the sales growth of the past quarter.

**H3.1**: The positive impact of credit is stronger in industries with higher recent growth.

Our hypothesis can be casted in the theoretical framework of Chevalier and Scharfstein (1996). In that paper, it is assumed that demand shocks are not correlated cross time, and the size of firms' initial investment does not affect their market share. Once we introduce a positive serial correlation of the demand shock and link firms' investment (e.g., advertisement) to market share, external credit will affect firms' market share growth especially when firms face positive demand shocks.

*[ Table 6 here ]*

Table 6 reports the heterogeneous effect of credit across different industry conditions. To explore the cross-sectional difference, we calculate the percentile rank of industries by their sales growth in the previous quarter, and interact it with the instrumented credit. Effectively, we calculate the local treatment effect of credit in different subsamples dividend by previous industry-level sales growth. In Column (1) of Table 6, credit access contributes $1.419\%$ *more* to the growth rate of a firm's market share when the past-quarter growth of its industry is ranked 10 percentile higher, in line with our hypothesis. The estimate is statistically significant, and robust after we replace the industry and month fixed effects with industry $\times$ month fixed effects (Column (3)) and control for the firm characteristics (Column (5)).

There are two potential issues of using lagged industry growth to proxy for product demand shocks. First, an industry can expand due to reduced cost of production instead of demand shock. Second, serial correlation in the sales growth of industries can bias the estimate in small sample by the logic of Stambaugh (1999).

To address these issues, we explore a difference-in-difference setting. Specifically, the percentile ranking of industry sales growth is replaced by a November indicator variable that captures the heightened consumer attention around November 11th, a shopping holiday called "Singles Day" (name derived from four "1"s in 11/11). While the event is not an officially recognized holiday in China, it has become the largest shopping festival in the world. On November 11th 2017, the total sales on Alibaba platforms amounted to $25.3 billion. In the same year, the two largest shopping holidays in the United States, Black Friday and Cyber Monday, registered total online sales of $5.0 billion and $6.6 billion respectively. "Double eleven" was trademarked in China by Alibaba Group in 2012. Around that day, Alibaba rolls out various promotion programs across its platforms including Taobao.com. It enhances its logistic services and attracts customers with various events, such as televised gala that mark the hours leading up to November 11th. This period of time is characterized by heightened customer attention and spending, so we use the November indicator for product demand shock.

**H3.2**: The positive impact of credit is stronger in the presence of a positive demand shock.

Column (2) of Table 6 shows that credit impact is stronger in November. Relative to other months, credit access contributes $5.6\%$ more to the growth rate of a firm's market share in November. The estimate is highly statistically significant, and robust once we replace the industry and month fixed effects with industry $\times$ month fixed effects (Column (3)) and add the firm characteristics as control variables (Column (5)). Note that in Column (3) and (5), the product category $\times$ month fixed effect absorb any industry-level variation so we do not include industry sales growth as a standalone control variable.

Overall, our results stand in contrast with the tradition wisdom. Credit, or liquidity in general, matters more for a firm's market share when an industry is growing, especially due to positive demand shocks. Customer attention is a scarce resource particularly in online markets (Dinerstein, Einav, Levin, and Sundaresan (2017)). During an expansion of product demand and customer attention, firms with credit access may invest more in advertisements to gain market share.[10] When customers face costs of switching between sellers, investing in market shares can be highly profitable during demand expansion.[11] Therefore, the shadow value of financial constraints, and in particular, the value of credit, can be higher in booms because there are stronger investment needs. Accordingly, we find credit helps firms more in booms.

**Credit and product pricing.** Models of product market competition under financial constraints mostly predict a negative relationship between firms' access to liquidity and product price. The tension is always between lowering price to gain market share and raising price to preserve liquidity. In Chevalier and Scharfstein (1996), firms lower prices now to gain market share for the future when they experience positive demand shocks that reduce the probability of bankruptcy. In Bolton and Scharfstein (1990), firms that are less liquidity constrained may engage in predatory pricing to starve liquidity-constrained competitors. Next, we test whether credit leads to product price change.

**H3.3**: Credit access does not affect firms' product pricing.

For each firm-month observation, we obtain a product price level that is transaction volume-weighted average across a firm's product lines. This allows us to comprehensively study the product pricing behavior of all firms on the platform. It is usually difficult to obtain

[10]See Bagwell (2007) for a review on the economics of advertisement, especially the informative role of advertisements and how advertisements expand customer base.

[11]Many emphasize customers' switching costs in the modeling market share dynamics. See Klemperer (1987) and (1995), Farrell and Shapiro (1988), Chen and Rosenthal (1996), Dasgupta and Titman (1998), and Campello and Fluck (2003).

data on product prices. Chevalier (1995) and Phillips (1995) examine how firms change product prices after leveraged buyouts in particular industries. Many studies are done at industry level, focusing on markups (Rotemberg and Woodford (1999); Campello (2003)).

*[ Table 7 here ]*

We use the same fuzzy discontinuity methodology with the dependent variable equal to the difference between a firm's product price change and the product price change of the whole industry (sales-weighted average across firms). It measures how firms change product prices relative to their peers. This benchmarking implicitly controls industry-level common shocks, but we still include industry $\times$ time fixed effects in some specifications to control for shocks that survive a simple industry-level demean. Column (1), (2), and (3) of Table 7 report the results of second-stage regression with the right-hand side controlling respectively for industry and month fixed effects, industry $\times$ month fixed effects, and firm characteristics together with industry $\times$ time fixed effects. Column (4), (5), and (6) of Table 7 report the corresponding first-stage results that are the same as those in Table 3.

Credit access does not affect firms' product pricing. The estimate is consistently indistinguishable from zero across specifications in Table 7. Accordingly, credit impact on firms' market share growth works through quantity, i.e., a higher number of products sold (Table 8).

*[ Table 8 here ]*

Overall, we find that access to credit provides firms with competitive advantages especially in booms when product demand expands, but firms do not gain market share by cutting prices as typically modeled in the literature. These findings suggest that firms use credit for other forms of investments in market share, for example advertisement, can be more important in this online market where customer attention tends to the scarce resource (Bagwell (2007); Dinerstein et al. (2017)).

## 4.4 Dynamic firm selection and the return-to-scale of big data

A special advantage of platform as a lender is that it can easily benchmark the performances of potential borrowers who operate on the platform, for example, by comparing a firm's sales growth with that of the whole industry to tease out any shocks at the industry level, and thereby, have clearer picture on a firm earning power. Relative performances can provide valuable information to lenders, and it may accelerates firm selection – firms with better relative performances not only benefit more from credit access, but also obtain credit more easily. Here, we study how the platform's lending decision is spanned by information on relative performances, and whether this information advantage amplifies the selection effect of credit. Specifically, we regress firms' credit score on relative performances, such as market share and ratings.

**H5.1**: A firm's credit score is positively correlated with its relative performance measures.

Rating is one particular type of relative performance information. On a scale from one to five, a firm's product, service, and delivery are rated by customers. We already know from Table 5 that the positive impact of credit on market share is stronger for firms with higher ratings. Do ratings affect whether a firm obtains credit access in the first place?

The platform has two channels of information aggregation. The first is the *big data channel*, i.e., the platform's proprietary algorithm that amasses all the variables available to form a signal of creditworthiness. The second is the *customer channel*. Ratings aggregate customers' evaluation of the firm in different dimensions, albeit that they can be noisy and biased for various reasons. These two channels of information acquisition can be substitutes. The more data the platform has on a firm, the less the firm's credit score depends on customer ratings. This hypothesis can be tested by adding an interaction term between a measure of data amount and ratings in the credit score regression.

We use a firm's cumulative transaction volumes (number of products sold) as a proxy for the volume of data the platform has on the firm. If the big data channel and the customer channel of information aggregation are substitutes, we should find a negative estimate on the interaction term between cumulative transaction and ratings.

These two channels can also be complementary. The more transactions a firm has done (i.e., a larger sample size), the more *precise* customer ratings are in evaluating a firm's competitiveness. A five-star rating based on two transactions contains certainly more noise than a five-star based on a million of transactions. If the complementarity hypothesis holds, we should find a positive coefficient on the interaction term between cumulative transaction and customer rating. Therefore, we also test the following hypothesis:

**H5.2**: Big data and customer ratings are substitutes as information sources for the platform: the correlation between credit score and ratings is higher for firms with less transactions.

Note that in these exercises, we think of credit score as a statistic that is measurable with respect to the platform's information set, and the aim is to characterize the conditioning function instead of identifying causality.

*[ Table 9 here ]*

Table 9 reports the results from the full-sample regression of a firm's credit score on various relative performances measures, and the interaction between ratings and cumulative transactions. The sample covers all firm-month observations, including those with credit score outside $[460, 500]$, in the period of September 2014 to November 2014, and January 2015 to June 2015. As previously discussed, our data start from August 2014, but credit score is only available starting September 2014. Moreover, observations in December 2014 are discarded because the credit decision rule was suspended at Ant Financial in that month. The total number of firm-month observations is 3,990,000, which is smaller than the total number of observations in Table 1, because due to data confidentiality issues, we have to conduct regression analysis on the computation platform at Alibaba and we are assigned with limited space that can only accommodates linear regression with a maximum of 3,990,000 observations. These observations are randomly selected.

We focus on the statistical significance of coefficient estimates and use $R^2$ as a measure of credit score variation explained by the variables of interest. In Column (1), the logarithm of sales explains almost 17% of credit score variation. We use statistical significance and $R^2$ of this *absolute* performance measure as a benchmark, so that we understand how strongly the *relative* performances factor into the platform's lending decision. In the unreported results, we find the level of sales explains a merely 2% of credit score variation, suggesting that the way credit score depends on sales is highly non-linear. In other words, the larger a firm is, the less the platform's lending decision depends on the size of the firm.

As the most common measure of relative performance, market share is positively correlated with credit score as shown in Column (2), suggesting that relative performance enters the platform's information set, consistent with hypothesis 5.1. Firms with higher market share tend to have higher credit score, and thereby, are more likely to obtain credit. Given our findings of a stronger positive impact of credit on market share growth for larger firms (Table 4), the following amplification mechanism emerges:

$$\textit{Larger firms grow faster} \begin{cases} \leftarrow \textit{ Easier credit access } \leftarrow \textit{ higher market share} \\ \leftarrow \textit{ Bigger impact of credit on larger firms' market share} \end{cases}$$

In Column (3), (4), and (5) of Table 9, the regression coefficient and t-statistics of merchandise rating, service rating, and delivery rating, and their interaction with cumulative transaction are reported respectively in each column. The message across three types of ratings is consistent. Consider a benchmark credit score of 470, approximately the constant term of our credit score regression. An increase of ratings by one level is strongly associated the firm obtaining credit access, given the 480 cutoff score for credit access. Similarly, if we consider a borderline firm with credit score slightly above 480, a decrease of ratings by one level is strongly associated with the firm losing credit access. Together with the results in Table 5, we observe the following amplification mechanism:

$$\textit{Better rated firms grow faster} \begin{cases} \leftarrow \textit{Easier credit access } \leftarrow \textit{ higher ratings} \\ \leftarrow \textit{Bigger impact of credit on better rated firms' market share} \end{cases}$$

Amplification mechanism is weakened when the platform obtains more data to gauge the quality of firms. The coefficients on the interaction between ratings and cumulative transaction ranking are significantly negative across all three types of ratings. These findings suggest that the two channels of information aggregation are substitutes. When more data is available, the platform relies less on customer ratings as a source of information on firms' quality. Firms that have done more businesses in a product category than others (i.e., with higher rank of cumulative transactions) see their credit access less dependent on customer ratings.

## 5 Conclusion

This paper provides the first evidence on how fintech powered by big data impacts firm dynamics. We obtain comprehensive data from Alibaba, one of the largest e-commerce platforms and the largest fintech lenders, and estimate the causal impact of fintech credit on the size distribution of online merchants. We find that credit extended by the platform promotes firm selection – credit leads to stronger growth of market share for e-commerce merchants that are larger and have higher customer ratings. To study the credit transmission mechanism, we examine the impact of credit under different industry conditions and its impact on product pricing. Moreover, a feedback effect highlights the dynamic aspect of big data: firms' credit score, assigned by platform's proprietary algorithm, is highly correlated with their market share and customer ratings. A feedback loop arises between firms' market status and platform credit that accelerates firm selection in this fast growing entrepreneurial space.

Our findings unveil an interesting set of interactions among big data, credit, and market structure, and speak to the broad issues of industrial organization and finance in the era of fintech. Our paper studies a new business paradigm: e-commerce platform extends credit to its online merchants. As a data hub, a platform has advantages in screening and monitoring borrowers, and as a trade hub, it has advantages in contract enforcement (i.e., imposing threat to exclude default merchants). Does platform credit improve the efficiency of online markets, and more specifically, stimulate entrepreneurial activities and enhance consumer surplus? Our angle is the size distribution of online merchants. Answers from alternative perspectives shall also be informative for both policy making and platform design.

# References


Aghion, P. and P. Bolton (1997). A theory of trickle-down growth and development. *The Review of Economic Studies 64*(2), 151–172.

Albuquerque, R. and H. A. Hopenhayn (2004). Optimal lending contracts and firm dynamics. *The Review of Economic Studies 71*(2), 285–315.

Angelini, P. and A. Generale (2008, March). On the evolution of firm size distributions. *American Economic Review 98*(1), 426–38.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. V. Reenen (2017). The fall of the labor share and the rise of superstar firms. Working paper, MIT.

Bagwell, K. (2007). Chapter 28 the economic analysis of advertising. Volume 3 of *Handbook of Industrial Organization*, pp. 1701 – 1844. Elsevier.

Berg, T., V. Burg, A. Gombović, and M. Puri (2018). On the rise of fintechs – credit scoring using digital footprints. Working paper, Duke University.

Bolton, P. and D. S. Scharfstein (1990). A theory of predation based on agency problems in financial contracting. *The American Economic Review 80*(1), 93–106.

Brynjolfsson, E., Y. J. Hu, and M. D. Smith (2010). Research commentary – long tails vs. superstars: The effect of information technology on product variety and sales concentration patterns. *Information Systems Research 21*(4), 736–747.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2017, March). Fintech, regulatory arbitrage, and the rise of shadow banks. Working Paper 23288, National Bureau of Economic Research.

Cabral, L. (2012). Chapter 13 reputation on the internet. The Oxford Handbook of the Digital Economy, pp. 344 – 354. Oxford University Press.

Cabral, L. and A. Hortaçsu (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics 58*(1), 54–78.

Cabral, L. M. B. and J. Mata (2003, September). On the evolution of the firm size distribution: Facts and theory. *American Economic Review 93*(4), 1075–1090.

Campello, M. (2003). Capital structure and product markets interactions: evidence from business cycles. *Journal of Financial Economics 68*(3), 353 – 378.

Campello, M. (2006). Debt financing: Does it boost or hurt firm performance in product markets? *Journal of Financial Economics 82*(1), 135 – 172.

Campello, M. and Z. Fluck (2003). Market share, financial leverage and the macroeconomy: Theory and empirical evidence. Working paper, University of Illinois.

Chen, M. X. and M. Wu (2017). The value of reputation in trade: Evidence from alibaba. Working paper, George Washington University.

Chen, Y. and R. W. Rosenthal (1996). Dynamic duopoly with slowly changing customer loyalties. *International Journal of Industrial Organization 14*(3), 269 – 296.

Chevalier, J. A. (1995). Capital structure and product-market competition: Empirical evidence from the supermarket industry. *The American Economic Review 85*(3), 415–435.

Chevalier, J. A. and D. S. Scharfstein (1996). Capital-market imperfections and countercyclical markups: Theory and evidence. *The American Economic Review 86*(4), 703–725.

Choi, D., D. Lou, and A. Mukherjee (2017). The effect of superstar firms on college major choice. Working Paper DP12296, CEPR.

Clementi, G. L. and H. A. Hopenhayn (2006). A theory of financing constraints and firm dynamics*. *The Quarterly Journal of Economics 121*(1), 229–265.

Cooley, T. F. and V. Quadrini (2001, December). Financial markets and firm dynamics. *American Economic Review 91*(5), 1286–1310.

Dasgupta, S. and S. Titman (1998). Pricing strategy and financial policy. *The Review of Financial Studies 11*(4), 705–737.

Dinerstein, M., L. Einav, J. Levin, and N. Sundaresan (2017). Consumer price search and platform design in internet commerce. *American Economic Review*, forthcoming.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *The American Economic Review 67*(3), 297–308.

Dou, W. W., Y. Ji, D. Reibstein, and W. Wu (2018). Customer capital, financial constraints and stock returns. Working paper, Wharton School of the University of Pennsylvania.

Edelberg, W. (2006). Risk-based pricing of interest rates for consumer loans. *Journal of Monetary Economics 53*(8), 2283–2298.

Einav, L., M. Jenkins, and J. Levin (2013). The impact of credit scoring on consumer lending. *The RAND Journal of Economics 44*(2), 249–274.

Einav, L., P. J. Klenow, B. Klopack, J. D. Levin, L. Levin, and W. Best (2017). Assessing the gains from e-commerce. Working paper, Stanford University.

Evans, D. S. (1987). Tests of alternative theories of firm growth. *Journal of Political Economy 95*(4), 657–674.

Fan, J., L. Tang, W. Zhu, and B. Zou (2016). The alibaba effect: Spatial consumption inequality and the welfare gains from e-commerce. Working paper, University of Maryland.

Fan, Y., J. Ju, and M. Xiao (2016). Reputation premium and reputation management: Evidence from the largest e-commerce platform in china. *International Journal of Industrial Organization 46*(Supplement C), 63 – 76.

Farrell, J. and C. Shapiro (1988). Dynamic competition with switching costs. *The RAND Journal of Economics 19*(1), 123–137.

Fleder, D. and K. Hosanagar (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science 55*(5), 697–712.

Fresard, L. (2010). Financial strength and product market behavior: The real effects of corporate cash holdings. *The Journal of Finance 65*(3), 1097–1122.

Frölich, M. and M. Huber (2018). Including covariates in the regression discontinuity design. *Journal of Business & Economic Statistics 0*, 0–0.

Fuster, A., M. C. Plosser, P. Schnabl, and J. Vickery (2018). The role of technology in mortgage lending. Staff Reports 836, ederal Reserve Bank of New York.

Gelman, A. and G. Imbens (2017). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics 0*(ja), 0–0.

Hahn, J., P. Todd, and W. V. der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica 69*(1), 201–209.

Hall, B. H. (1987). The relationship between firm size and firm growth in the us manufacturing sector. *The Journal of Industrial Economics 35*(4), 583–606.

Hau, H., Y. Huang, H. Shan, and Z. Sheng (2018). Techfin in China: Credit market completion and its growth effect. Working paper, Graduate Institute of International and Development Studies.

Huang, Y., C. Lin, Z. Sheng, and L. Wei (2018). Fintech credit and service quality: Evidence from Alibaba online merchants. Working paper, Graduate Institute of International and Development Studies.

Imbens, G. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies 79*(3), 933–959.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615 – 635. The regression discontinuity design: Theory and applications.

Klemperer, P. (1987). Markets with consumer switching costs*. *The Quarterly Journal of Economics 102*(2), 375–394.

Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *The Review of Economic Studies 62*(4), 515–539.

Kovenock, D. and G. M. Phillips (1997). Capital structure and product market behavior: An examination of plant exit and investment decisions. *The Review of Financial Studies 10*(3), 767–803.

Lee, D. S. (2008, February). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics 142*(2), 675–697.

Lee, D. S. and T. Lemieux (2010, June). Regression discontinuity designs in economics. *Journal of Economic Literature 48*(2), 281–355.

Lieber, E. and C. Syverson (2012). Online versus offline competition. *Oxford Handbook of the Digital Economy*, 189–223.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics 9*(2), 508–523.

MacKay, P. and G. M. Phillips (2005). How does industry affect firm financial structure? *The Review of Financial Studies 18*(4), 1433–1466.

Matsuyama, K. (2000). Endogenous inequality. *The Review of Economic Studies 67*(4), 743–759.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698 – 714. The regression discontinuity design: Theory and applications.

Moll, B. (2014, October). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review 104*(10), 3186–3221.

Noe, T. and G. Parker (2005). Winner take all: Competition, strategy, and the structure of returns in the internet economy. *Journal of Economics and Management Strategy 14*(1), 141–164.

Parsons, C. and S. Titman (2007). Capital structure and corporate strategy. Working paper, University of Texas.

Petersen, M. A. and R. G. Rajan (1997). Trade credit: Theories and evidence. *The Review of Financial Studies 10*(3), 661–691.

Petersen, M. A. and R. G. Rajan (2002). Does distance still matter? the information revolution in small business lending. *The Journal of Finance 57*(6), 2533–2570.

Philippon, T. (2017). The FinTech Opportunity. Working paper 655, BIS.

Phillips, G. M. (1995). Increased debt and industry product markets an empirical analysis. *Journal of Financial Economics 37*(2), 189 – 238.

Piketty, T. (1997). The dynamics of the wealth distribution and the interest rate with credit rationing. *The Review of Economic Studies 64*(2), 173–189.

Piketty, T. and E. Saez (2003). Income inequality in the united states, 19131998*. *The Quarterly Journal of Economics 118*(1), 1–41.

Porter, J. (2003). Estimation in the regression discontinuity model. Working paper, Harvard University Department of Economics.

Rochet, J. and J. Tirole (2003). Platform competition in twosided markets. *Journal of the European Economic Association 1*(4), 990–1029.

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

Rosen, S. (1981). The economics of superstars. *The American Economic Review 71*(5), 845–858.

Rotemberg, J. J. and M. Woodford (1999). The cyclical behavior of prices and costs. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1 of *Handbook of Macroeconomics*, Chapter 16, pp. 1051–1135. Elsevier.

Shadish, W. R., M. H. Clark, and P. M. Steiner (2008). Can nonrandomized experiments yield accurate answers? a randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association 103*(484), 1334–1344.

Simon, H. A. and C. P. Bonini (1958). The size distribution of business firms. *The American Economic Review 48*(4), 607–617.

Stambaugh, R. F. (1999). Predictive Regressions. *Journal of Financial Economics 54*, 375–421.

Welch, I. (1992). Sequential sales, learning, and cascades. *The Journal of Finance 47*(2), 695–732.

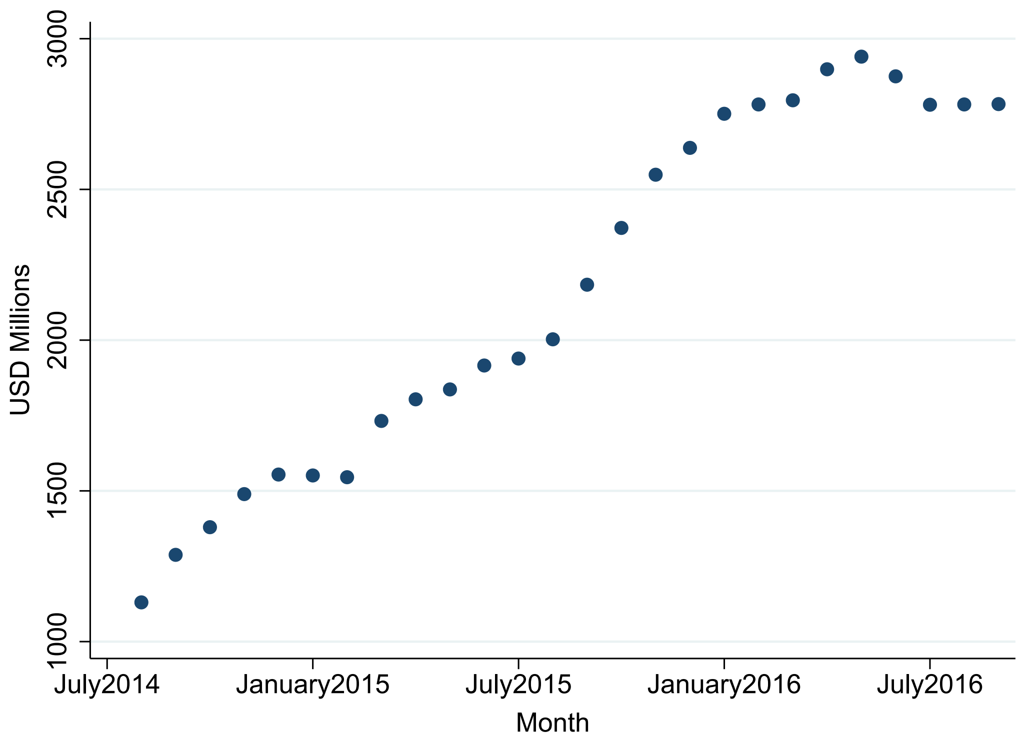


Figure 1: Total Volume of Credit Lines Outstanding by Month.

This figure plots the evolution of total amount of loans outstanding extended by Ant Financial to Taobao merchants for each month starting in August 2014 and ending in September 2016. The data is from Ant Financial.

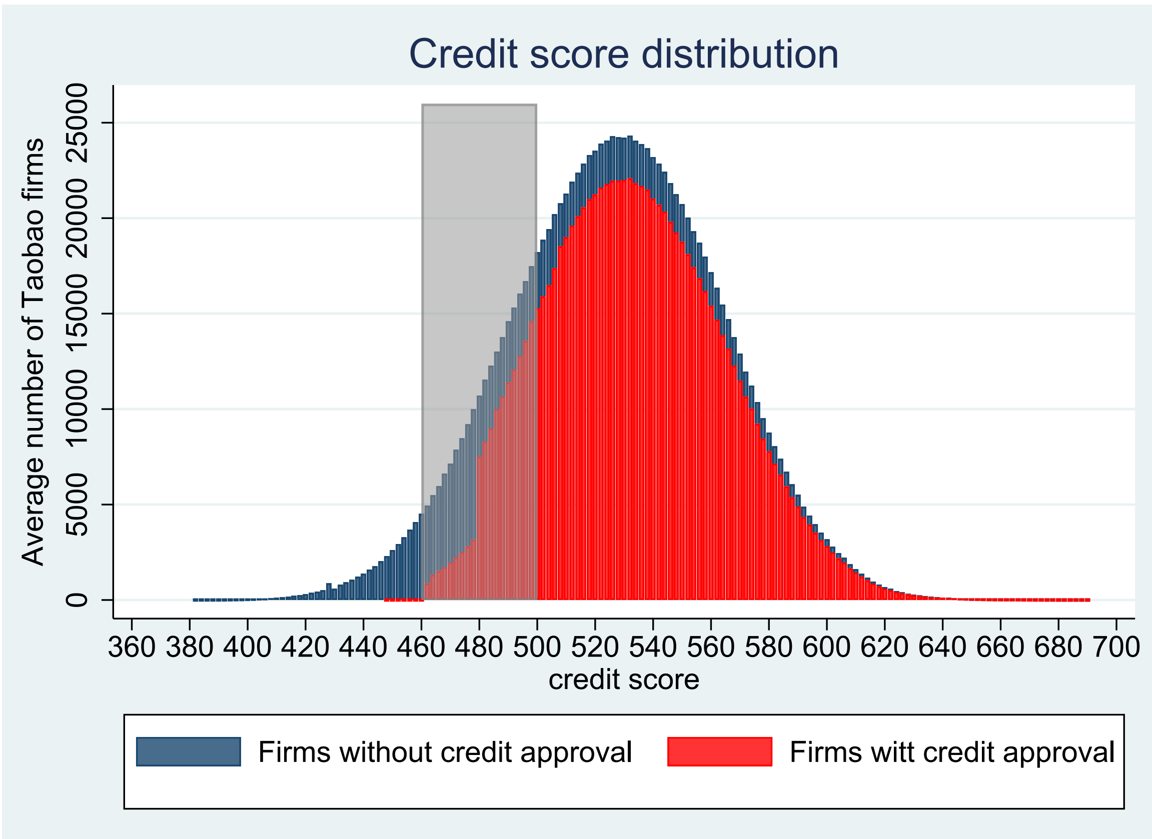


Figure 2: Firm Distribution.

This figures shows the distribution of credit scores for all Taobao merchants in our sample from September 2014 to November 2014, and January 2015 to June 2015. For any given credit score, the blue bars show the average number of firms approved for credit in our sample, and the red bars show the average number of firms without credit approval. Note that our data of credit score starts in September 2014 (while other variables start in August 2014), and data in December 2014 are discarded because of the suspension of credit scoring rule at Ant Financial. The credit score of a firm in month $t$ is obtained at the very end of month $t-1$ to leave a period of time for Ant Financial to respond in its credit approval decision. A discontinuity appears at the score of $480$.

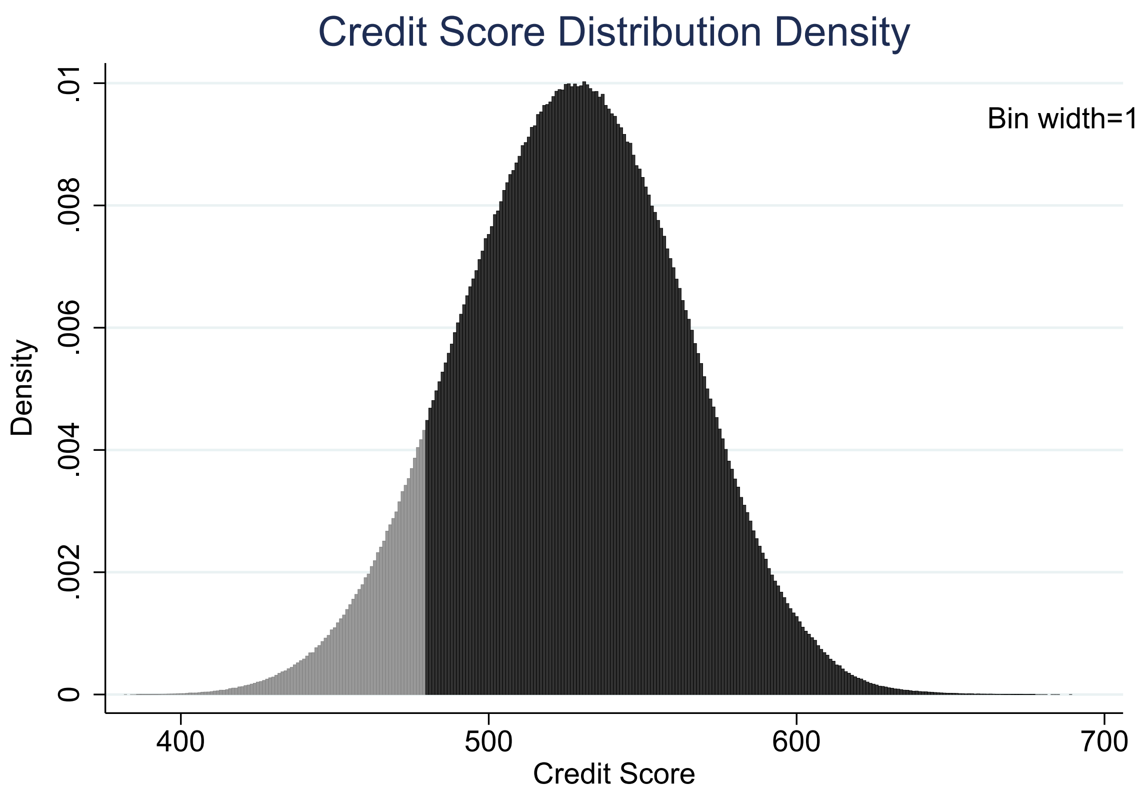


Figure 3: Credit Score Distribution Density.

This figure shows that the distribution of firms across credit scores. The density function is calculated by normalizing the number of firms at each value of credit score by the total number of firms. For any given credit score, we calculate the average (over time) number of firms. Note that our data of credit score starts in September 2014 (while other variables start in August 2014), and data in December 2014 are discarded because of the suspension of credit scoring rule at Ant Financial. The credit score of a firm in month $t$ is obtained at the very end of month $t-1$ to leave a period of time for Ant Financial to respond in its credit approval decision. We mark the area above the credit score of $480$ in black, and the remaining area in gray.

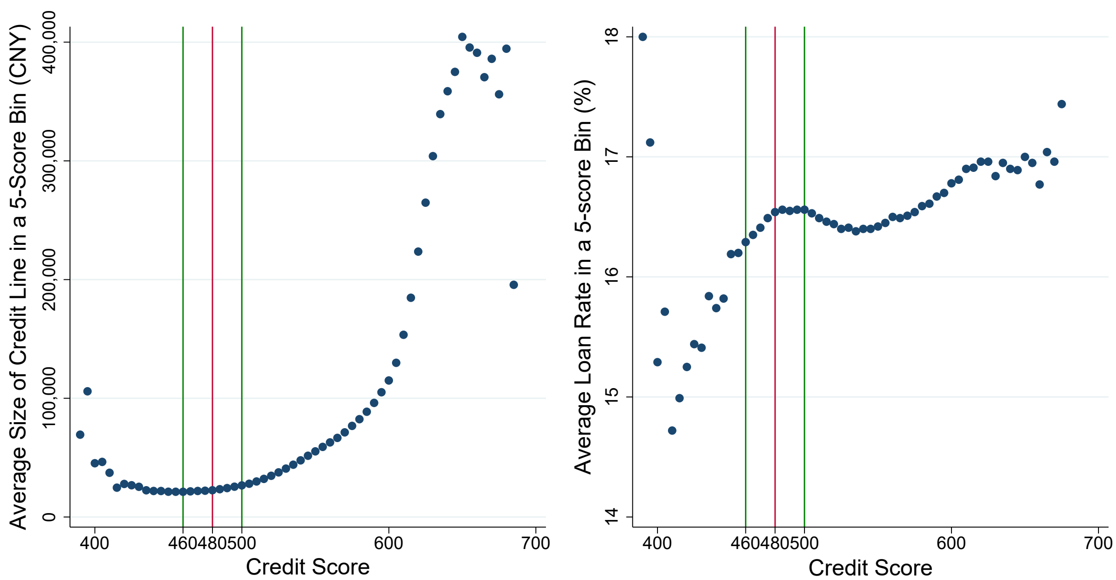


Figure 4: The Average of Loan Size and Rate for Each Credit Score Bin (5 Scores).

This figure plots the average (over time and firm) credit line in the left panel and the average (over time and firm) loan rate in the right panel for each 5-score bin on the horizontal axis of credit score. The sample for regression discontinuity analysis, i.e., observations with credit score in $[460, 500]$, is highlighted by the green lines, and the threshold value of 480 is highlighted by the red line in the middle. Note that our data of credit score starts in September 2014 (while other variables start in August 2014), and data in December 2014 are discarded because of the suspension of credit scoring rule at Ant Financial. The credit score of a firm in month $t$ is obtained at the very end of month $t-1$ to leave a period of time for Ant Financial to respond in its credit approval decision.

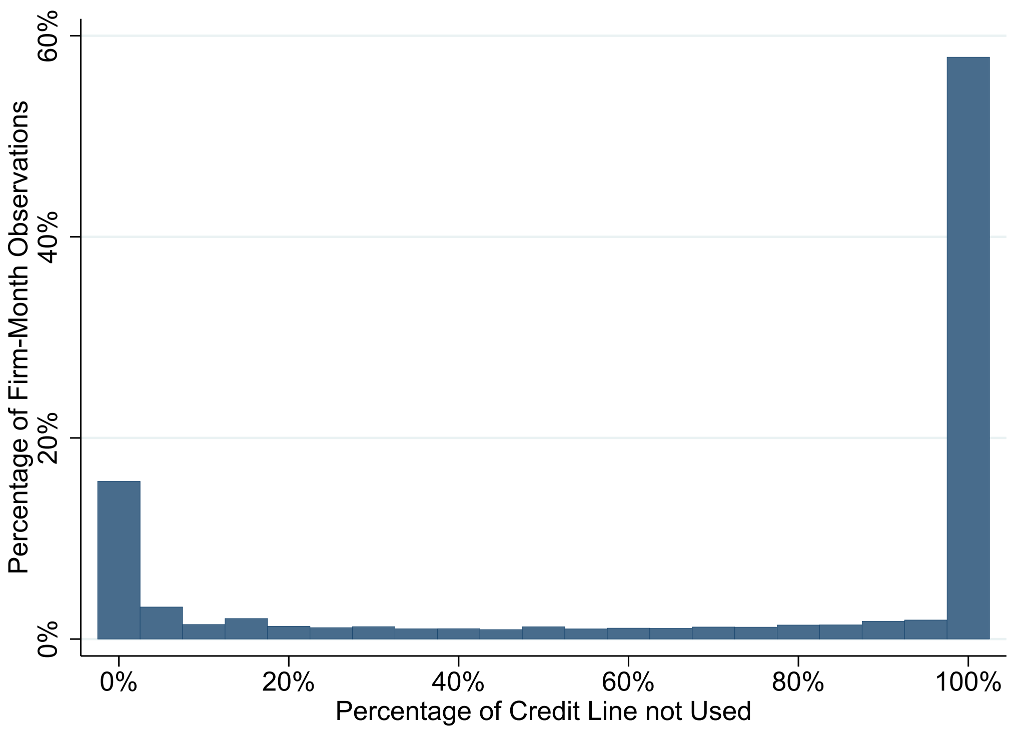


Figure 5: The Distribution of Financial Slackness (Percentage of Credit Line Not Used).

This figure shows the percentage of firm-month observations for twenty one equally distributed binds of credit usage. Credit usage is defined as the percentage of credit line drawn by the firm in a given month. The data is from our full sample, i.e., from August 2014 to November 2014, and from January 2015 to June 2015. Note that data in December 2014 are discarded because of the suspension of credit scoring rule at Ant Financial.

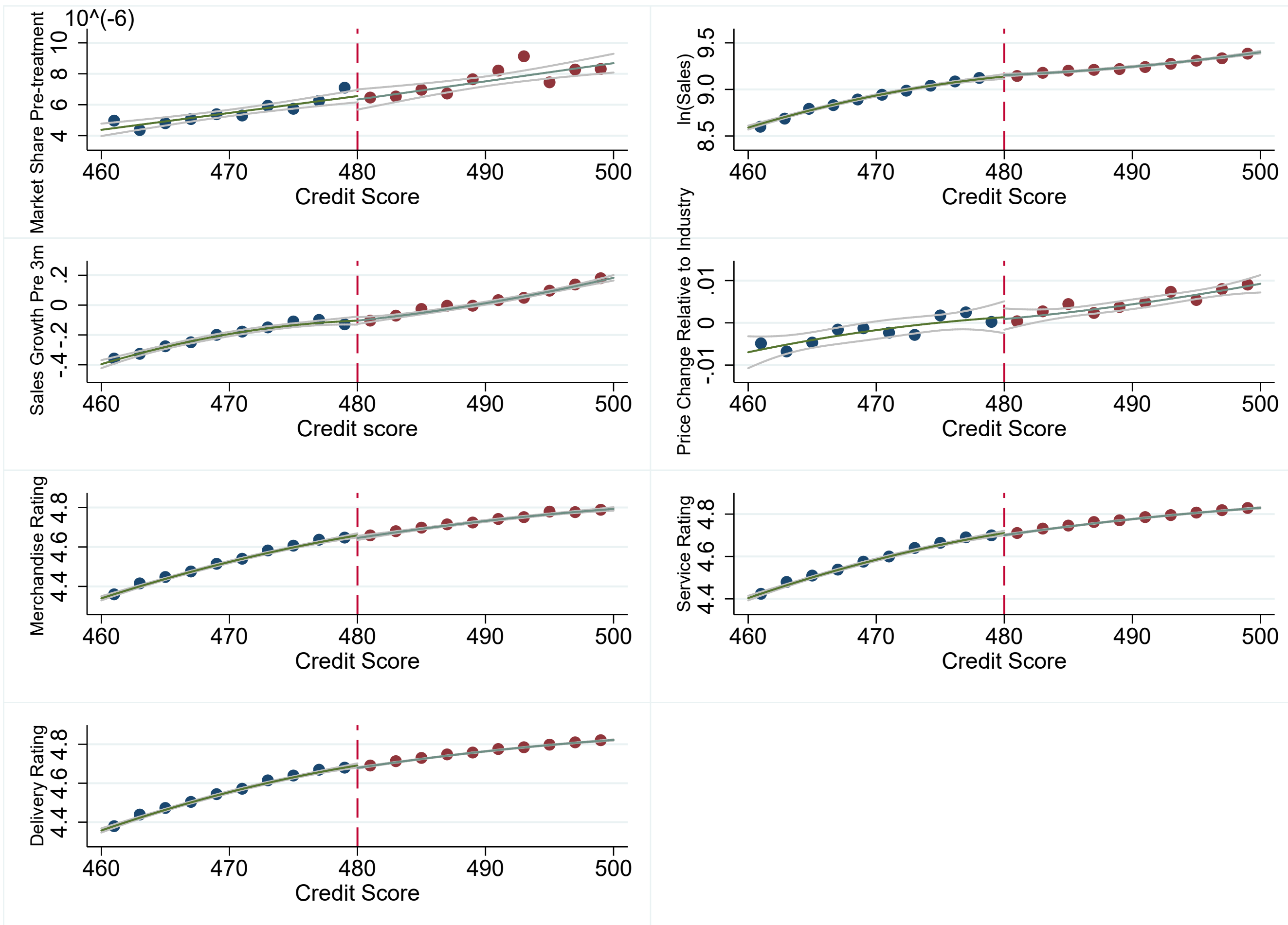


Figure 6: Merchant Characteristic Continuity.

This figure plots the average (over time and firm) value of firm characteristics for ten evenly distributed binds of credit score on each side of the threshold value of 480. The data is from our sample for regression discontinuity analysis, described in Section 4.

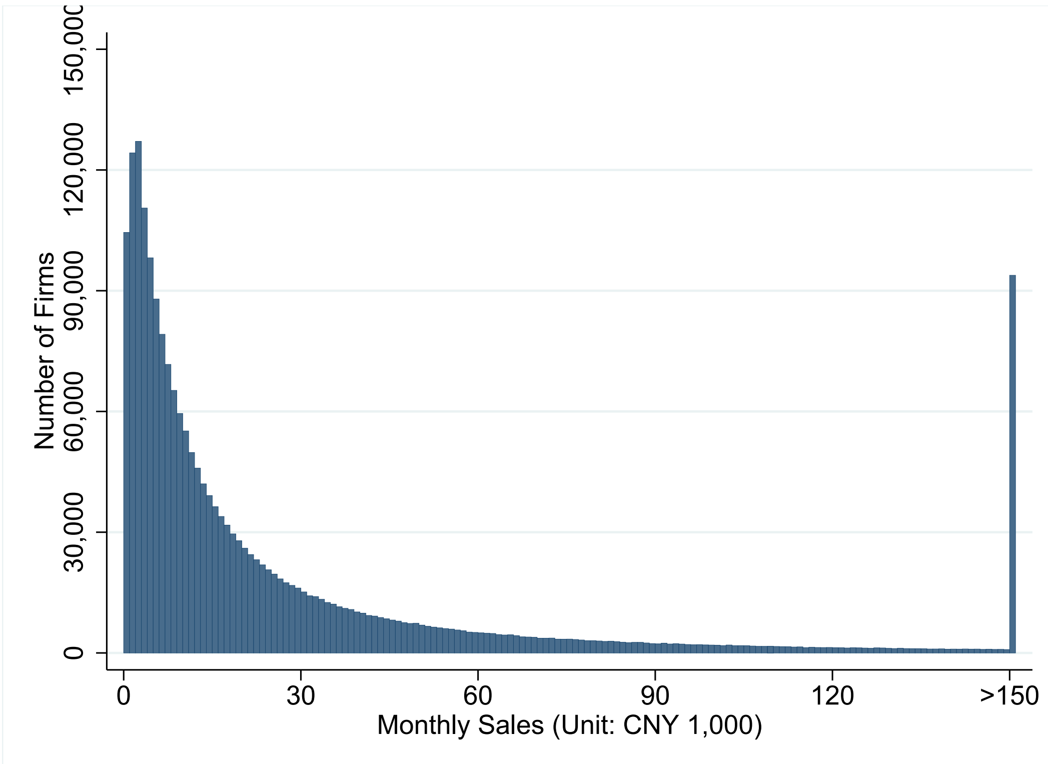


Figure 7: Firm Size Distribution (Averaged over Months).

This figure plots the number of firms for each level of monthly sales. Firms with sales larger than CNY 150,000 (approximately $22,000) are summed up in the last bin. The distribution is calculated every month, and then, averaged over time. . The data is from our full sample, August 2014 to November 2014, and January 2015 to June 2015. Note that data in December 2014 are discarded because of the suspension of credit scoring rule at Ant Financial.

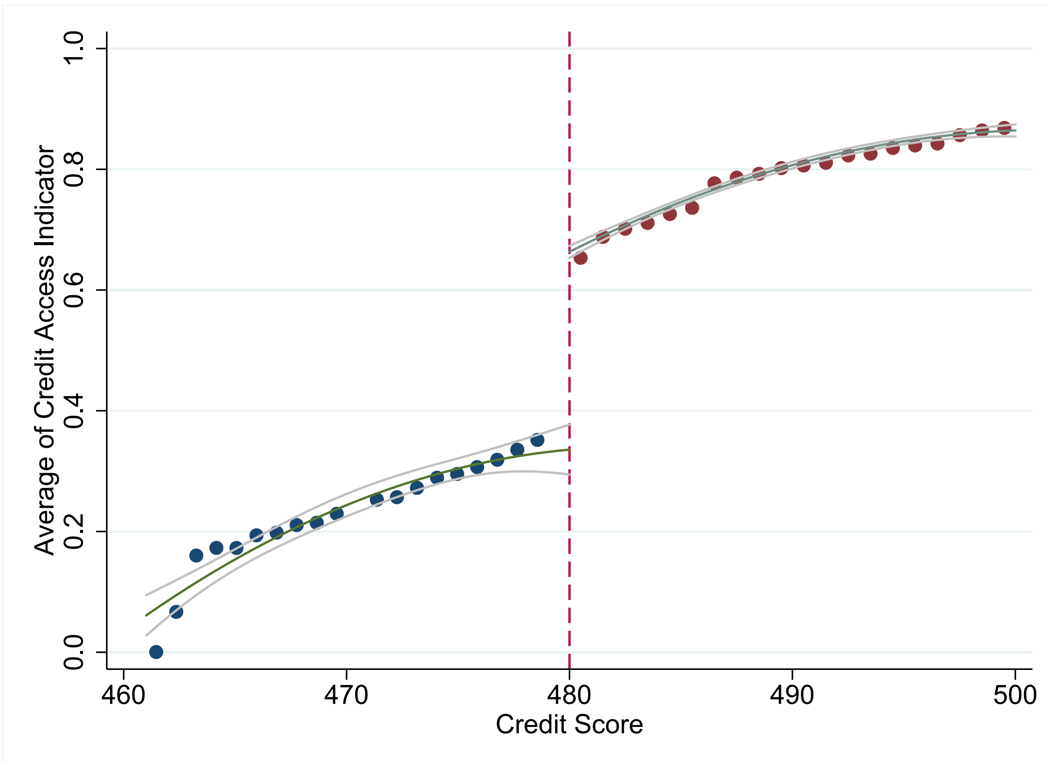


Figure 8: The Average of Credit Access Indicator in Each 1-point Credit Score Bin.

This figure plots the average (over time and firm) of credit access indicator at each level of credit score in the interval from 460 to 500. A polynominal curve is fit the scattered dots on each side the cutoff score of 480. The data is from our sample for regression discontinuity analysis, described in Section 4.

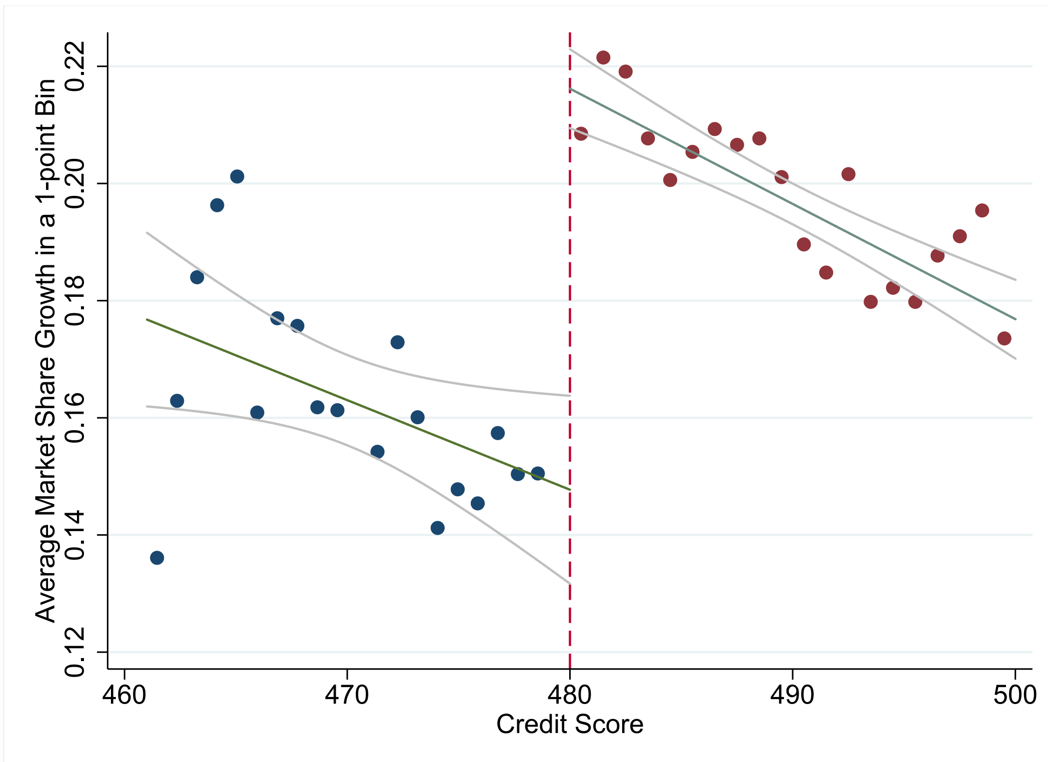


Figure 9: The Average of Market Share Growth in Each 1-point Credit Score Bin.

This figure presents the discontinuity plots of the average (over time and firm) market share in the month subsequent to the credit allocation decision of Ant Financial (i.e. post-treatment) events against different levels of credit scores. A polynominal curve is fit the scattered dots on each side the cutoff score of 480. The data is from our sample for regression discontinuity analysis, described in Section 4.

## Table 1: Summary Statistics

| | All Taobao Firms | | | Credit Score in $[460, 500]$ | | |
|---|---|---|---|---|---|---|
| | Observations | Mean | STD | Observations | Mean | STD |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A: Online credit information | | | | | | |
| *Credit Approval (0/1)* | 12,014,748 | 0.775 | 0.417 | 1,146,740 | 0.638 | 0.481 |
| *Credit Line, Approval =1 (CNY)* | 9,315,393 | 40,879 | 112,057 | 731,730 | 25,767 | 71,942 |
| *Credit Use/ Credit Line, Approval =1 (CNY)* | 9,315,393 | 0.144 | 0.483 | 731,730 | 0.209 | 0.618 |
| Panel B: Firm characteristics | | | | | | |
| *Sales (CNY)* | 12,014,696 | 45,675 | 195,970 | 1,146,740 | 31,944 | 121,210 |
| *Market share* | 9,534,712 | 5.654E-5 | 8.239E-4 | 1,146,740 | 9.65E-6 | 1.78E-6 |
| *Ln(sales+1)* | 12,014,696 | 8.894 | 2.719 | 1,146,740 | 9.134 | 1.593 |
| *Credit score* | 11,970,625 | 523 | 37.35 | 1,146,740 | 487 | 16.48 |
| *Deliver rating* | 12,014,748 | 4.495 | 1.245 | 1,146,740 | 4.699 | 0.571 |
| *Service rating* | 12,014,748 | 4.504 | 1.256 | 1,146,740 | 4.715 | 0.586 |
| *Merchandise rating* | 12,014,748 | 4.478 | 1.241 | 1,146,740 | 4.669 | 0.574 |

Table 2: Herfindahl-Hirschman Index (HHI) and Credit Availability

| Dependent variable: | HHI | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Constant | $0.0401^{***}$ | $0.0393^{***}$ | $0.0459^{***}$ |
| | (5.805) | (2.849) | (30.069) |
| Credit Line / Total Sales | $0.0519^{***}$ | $0.0516^{***}$ | $0.0329^{***}$ |
| | (9.020) | (8.902) | (8.377) |
| Product Category (Industry) FE | No | No | Yes |
| Month FE | No | Yes | Yes |
| Observations | 1,185 | 1,185 | 1,185 |
| $R^2$ | 0.0684 | 0.0725 | 0.3701 |

Table 3: Credit Impact on Market Share Growth

| Dependent variable: | $\Delta \ln(sales_{firm,t}) - \Delta \ln(sales_{industry,t})$ | | | | | |
|---|---|---|---|---|---|---|
| | second stage | | | first stage | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Intercept | $0.1048$ | $0.1784^{***}$ | $0.0994^{***}$ | $0.3916^{***}$ | $0.4346^{***}$ | $-0.0333^{***}$ |
| | $(1.20)$ | $(23.24)$ | $(4.10)$ | $(4.41)$ | $(365.6)$ | $(-4.717)$ |
| Instrumented credit access | $0.0612^{***}$ | $0.0512^{***}$ | $0.0741^{***}$ | | | |
| | $(2.68)$ | $(3.83)$ | $(5.59)$ | | | |
| If Creditscore above 480 | | | | $0.2551^{***}$ | $0.2335^{**}$ | $0.2126^{***}$ |
| | | | | $(122.2)$ | $(113.4)$ | $(105.6)$ |
| Control variables | No | No | Yes | No | No | Yes |
| Product Category (Industry) FE | Yes | No | No | Yes | No | No |
| Month FE | Yes | No | No | Yes | No | No |
| Product Category $\times$Month FE | No | Yes | Yes | No | Yes | Yes |
| Observations | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ |
| $R^2$ * | 0.012 | 0.00009586 | 0.0165 | 0.3174 | 0.3045 | 0.3355 |

* $R^2$ is overall $R^2$ for specifications with separate fixed effects of category and month. $R^2$ is within $R^2$ for specifications with interacting fixed effects of category and month.

## Table 4: Credit Impact and Firm Size

| Dependent variable: | | $\Delta \ln(sales_{firm,t}) - \Delta \ln(sales_{industry,t})$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Intercept | | 0.0275 | −0.1218 | 0.118 | 0.0575*** | −0.0209 | 0.0609*** | 0.0748*** | 0.0176 | 0.006 |
| | | (0.32) | (−1.39) | (1.36) | (5.288) | (−1.219) | (3.822) | (3.186) | (0.7035) | (0.2558) |
| Instrumented credit access | | 0.0213 | −0.0896*** | 0.0738** | 0.0265 | −0.051 | 0.2761*** | 0.0318 | 0.0012 | 0.003 |
| | | (0.73) | (−2.72) | (2.53) | (1.102) | (−0.1946) | (9.64) | (0.918) | (0.018) | (0.0618) |
| 100× Mkt share percentile | $(\times 10^{-3})$ | 0.515** | | | 1.905*** | | | −0.7456** | | |
| | | (2.5) | | | (9.58) | | | (−2.3) | | |
| 100× Mkt share percentile× IV credit access | $(\times 10^{-3})$ | 4.107*** | | | 2.299*** | | | 4.792*** | | |
| | | (19.14) | | | (10.5) | | | (19.97) | | |
| ln (sales+1) | $(\times 10^{-3})$ | | 8.931*** | | | −18.39*** | | | −33.75*** | |
| | | | (31.24) | | | (−5.997) | | | (−9.81) | |
| ln (sales+1)*IV credit access | $(\times 10^{-3})$ | | 57.46*** | | | 77.57*** | | | 49.69*** | |
| | | | (20.85) | | | (24.01) | | | (16.83) | |
| Firm sales growth | $(\times 10^{-3})$ | | | 53.01*** | | | 59.55*** | | | 10.86*** |
| (Past three-month) | | | | (41.58) | | | (45.87) | | | (6.504) |
| Firm sales growth*IV credit access | $(\times 10^{-3})$ | | | 26.70*** | | | 19.88*** | | | 92.28*** |
| | | | | (14.14) | | | (10.36) | | | (44.57) |
| Control variables | | No | No | No | No | No | No | Yes | Yes | Yes |
| Product Category | | Yes | Yes | Yes | No | No | No | No | No | No |
| Month FE | | Yes | Yes | Yes | No | No | No | No | No | No |
| Product Category × Month FE | | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | | 1,146,740 | 1,146,740 | 1,146,740 | 1,146,740 | 1,146,740 | 1,146,740 | 1,146,740 | 1,146,740 | 1,146,740 |
| $R^2$ * | | 0.0186 | 0.0175 | 0.0247 | 0.0069 | 0.0063 | 0.0144 | 0.0169 | 0.0162 | 0.0179 |

* $R^2$ is overall $R^2$ for specifications with separate fixed effects of category and month. $R^2$ is within $R^2$ for specifications with interacting fixed effects of category and month.

Table 5: Credit Impact and Reputation

| Dependent variable: | | $\Delta \ln(sales_{firm,t}) - \Delta \ln(sales_{industry,t})$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Intercept | | $-0.0852$ | $-0.0778$ | $-0.0767$ | $0.3588^{***}$ | $-0.0494^{***}$ | $0.1703^{***}$ | $0.0263$ | $0.0346$ | $0.0304$ |
| | | $(-0.9731)$ | $(-0.8889)$ | $(-0.9119)$ | $(25.85)$ | $(-3.599)$ | $(15.45)$ | $(1.057)$ | $(1.42)$ | $(1.228)$ |
| Instrumented credit access | | $-0.0607$ | $-0.0686$ | $-0.0712$ | $-2.706^{***}$ | $-0.0088$ | $-3.306^{***}$ | $-0.0011$ | $-0.0040$ | $-0.0016$ |
| | | $(-1.133)$ | $(-1.256)$ | $(-1.294)$ | $(-52.0)$ | $(-0.1654)$ | $(-61.14)$ | $(-0.0189)$ | $(-0.0680)$ | $(-0.0266)$ |
| Merchandise rating | $(\times 10^{-3})$ | $-26.83^{***}$ | | | $-19.86^{***}$ | | | $46.53^{***}$ | | |
| | | $(-7.685)$ | | | $(-5.465)$ | | | $(5.652)$ | | |
| Merchandise rating $\times$ IV credit access | | $0.167^{***}$ | | | $0.5516^{***}$ | | | $0.1723^{***}$ | | |
| | | $(20.32)$ | | | $(69.83)$ | | | $(18.72)$ | | |
| Service rating | $(\times 10^{-3})$ | | $-34.01^{***}$ | | | $-55.91^{***}$ | | | $-58.41^{***}$ | |
| | | | $(-10.03)$ | | | $(-15.88)$ | | | $(-7.63)$ | |
| Service rating $\times$ IV credit access | | | $0.1732^{***}$ | | | $0.1995^{***}$ | | | $0.1759^{***}$ | |
| | | | $(20.53)$ | | | $(24.46)$ | | | $(19.11)$ | |
| Delivery rating | $(\times 10^{-3})$ | | | $-40.16^{***}$ | | | $41.16^{***}$ | | | $-28.13^{***}$ |
| | | | | $(-11.54)$ | | | $(10.16)$ | | | $(-3.405)$ |
| Delivery rating $\times$ IV credit access | | | | $0.1861^{***}$ | | | $0.635^{***}$ | | | $0.1641^{***}$ |
| | | | | $(22.03)$ | | | $(79.45)$ | | | $(17.44)$ |
| Control variables | | No | No | No | No | No | No | Yes | Yes | Yes |
| Product Category | | Yes | Yes | Yes | No | No | No | No | No | No |
| Month FE | | Yes | Yes | Yes | No | No | No | No | No | No |
| Product Category $\times$Month FE | | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ |
| $R^2$ * | | $0.0144$ | $0.014$ | $0.0143$ | $0.0058$ | $0.0028$ | $0.0067$ | $0.0172$ | $0.0171$ | $0.017$ |

* $R^2$ is overall $R^2$ for specifications with separate fixed effects of category and month. $R^2$ is within $R^2$ for specifications with interacting fixed effects of category and month.

# Table 6: Credit Transmission Mechanism

| Dependent variable: | | $\Delta \ln(sales_{firm,t}) - \Delta \ln(sales_{industry,t})$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| Intercept | | $0.1466^{*}$ | $0.0794$ | $0.1461^{***}$ | $0.1462$ | $0.000769$ | $0.0935^{***}$ |
| | | $(1.67)$ | $(0.91)$ | $(8.936)$ | $(0.7602)$ | $(0.033)$ | $(4.017)$ |
| Instrumented credit access | | $0.0724^{**}$ | $0.0762^{***}$ | $0.076^{**}$ | $0.0922$ | $0.000381$ | $0.0205$ |
| | | $(2.46)$ | $(2.73)$ | $(2.412)$ | $(0.2616)$ | $(0.00527)$ | $(0.4124)$ |
| $100 \times$ Industry sales growth percentile | $(\times 10^{-3})$ | $-2.327^{***}$ | | | | | |
| (Past three-month) | | $(-20.21)$ | | | | | |
| $100 \times$ Industry sales growth percentile $\times$ IV credit access | $(\times 10^{-3})$ | $1.419^{***}$ | | $0.7465^{***}$ | | $1.969^{***}$ | |
| | | $(9.00)$ | | $(4.639)$ | | $(9.245)$ | |
| November $\times$ IV credit access | | | $0.056^{***}$ | | $0.1307^{***}$ | | $0.1449^{***}$ |
| | | | $(4.426)$ | | $(4.211)$ | | $(11.58)$ |
| Control variables | | No | No | No | No | Yes | Yes |
| Product Category | | Yes | Yes | No | No | No | No |
| Month FE | | Yes | Yes | No | No | No | No |
| Product Category $\times$ Month FE | | No | No | Yes | Yes | Yes | Yes |
| Observations | | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ |
| $R^2$ * | | $0.0112$ | $0.0122$ | $0.000137$ | $0.000229$ | $0.016$ | $0.0167$ |

* $R^2$ is overall $R^2$ for specifications with separate fixed effects of category and month. $R^2$ is within $R^2$ for specifications with interacting fixed effects of category and month.

Table 7: Credit Impact on Product Pricing

| Dependent variable: | | $\Delta \ln(productprice_{firm,t}) - \Delta \ln(productprice_{industry,t})$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | Second Stage | | | First Stage | | |
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| Intercept | | 0.0349 | 0.0010 | 0.00000992 | $0.3916^{***}$ | $0.4346^{***}$ | $-0.0333^{***}$ |
| | | (0.4329) | (0.4845) | (−0.0014) | (4.41) | (365.6) | (−4.717) |
| Instrumented credit access | $(\times 10^3)$ | 12.4 | 0.616 | 0.00677 | | | |
| | | (1.673) | (0.164) | (0.00179) | | | |
| If Creditscore above 480 | | | | | $0.2551^{***}$ | $0.2335^{***}$ | $0.2126^{***}$ |
| | | | | | (122.2) | (113.4) | (105.6) |
| Control variables | | No | No | Yes | No | No | Yes |
| Product Category | | Yes | No | No | Yes | No | No |
| Month FE | | Yes | No | No | Yes | No | No |
| Product Category $\times$Month FE | | No | Yes | Yes | No | Yes | Yes |
| Observations | | 1,146,740 | 1,146,740 | 1,146,740 | 1,146,740 | 1,146,740 | 1,146,740 |
| $R^2$ * | | 0.0027 | 0.000105 | 0.000120 | 0.3174 | 0.3045 | 0.3355 |

* $R^2$ is overall $R^2$ for specifications with separate fixed effects of category and month. $R^2$ is within $R^2$ for specifications with interacting fixed effects of category and month.

Table 8: Credit Impact on Product Quantity

| Dependent variable: | $\Delta \ln(transaction_{firm,t}) - \Delta \ln(transaction_{industry,t})$ | | | | | |
|---|---|---|---|---|---|---|
| | Second Stage | | | First Stage | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Intercept | $0.1988$ | $0.3461^{***}$ | $0.1032^{***}$ | $0.3916^{***}$ | $0.4346^{***}$ | $-0.0333^{***}$ |
| | $(0.4822)$ | $(31.32)$ | $(2.941)$ | $(4.41)$ | $(365.6)$ | $(-4.717)$ |
| Instrumented credit access | $0.1168^{***}$ | $0.0659^{***}$ | $0.082^{***}$ | | | |
| | $(3.078)$ | $(3.426)$ | $(4.279)$ | | | |
| If Creditscore above 480 | | | | $0.2551^{***}$ | $0.2335^{***}$ | $0.2126^{***}$ |
| | | | | $(122.2)$ | $(113.4)$ | $(105.6)$ |
| Control variables | No | No | Yes | No | No | Yes |
| Product Category | Yes | No | No | Yes | No | No |
| Month FE | Yes | No | No | Yes | No | No |
| Product Category $\times$Month FE | No | Yes | Yes | No | Yes | Yes |
| Observations | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ | $1,146,740$ |
| $R^2$ * | $0.0128$ | $0.00004008$ | $0.0087$ | $0.3174$ | $0.3045$ | $0.3355$ |

* $R^2$ is overall $R^2$ for specifications with separate fixed effects of category and month. $R^2$ is within $R^2$ for specifications with interacting fixed effects of category and month.

Table 9: Information Aggregation and the Dynamic Effect of Platform Credit

| Dependent variable: | | Credit score | | | | | |
|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| *Intercept* | | 470.5*** | 470.4*** | 469.6*** | 471.4*** | 469.8*** | 469.0*** |
| | | (7731) | (6617) | (6868) | (6914) | (6860) | (6537) |
| *Ln(Sales+1)* | | 5.856*** | | | | | 10.17*** |
| | | (896.4) | | | | | (568.9) |
| *Market share* | | | 61.79*** | | | | −81.03*** |
| | | | (758.2) | | | | (−339.8) |
| *Merchandise rating* | | | | 11.92*** | | | −2.292*** |
| | | | | (742.3) | | | (−13.55) |
| *Merchandise rating* Cum transaction ranking* | $(\times 10^{-3})$ | | | −2.078*** | | | 327.2*** |
| | | | | (−15.67) | | | (82.6) |
| *Service rating* | | | | | 11.5*** | | 2.386*** |
| | | | | | (720.5) | | (19.67) |
| *Service rating* Cum transaction ranking* | $(\times 10^{-3})$ | | | | −2.773*** | | −325.3*** |
| | | | | | (−20.95) | | (−100.9) |
| *Delivery rating* | | | | | | 11.85*** | 8.623*** |
| | | | | | | (738.8) | (49.58) |
| *Delivery rating* Cum transaction ranking* | $(\times 10^{-3})$ | | | | | −2.59*** | −34.23*** |
| | | | | | | (−19.59) | (−8.364) |
| $R^2$ | | 0.1685 | 0.1258 | 0.1407 | 0.1335 | 0.1393 | 0.219 |
| Observations | | 3,990,000 | 3,990,000 | 3,990,000 | 3,990,000 | 3,990,000 | 3,990,000 |

# Appendix I: Product Categories on Taobao.com

| Category name | HHI | Category name | HHI | Category name | HHI |
|---|---|---|---|---|---|
| Men Cloth | 0.0001 | Milk Powder/Food Supplements/Nutritions/Snacks | 0.001 | Takeaway / Delivery / Ordering | 0.0127 |
| Women Shoes | 0.0001 | Large household electronic appliances | 0.0011 | Gaming: equipments, currency, account, delegate player | 0.0134 |
| Women Cloth | 0.0001 | Books/Magazines/Newspapers | 0.0011 | Discount hotels and hostels | 0.0138 |
| Home Decoration Materials | 0.0002 | Children's shoes/parent-child shoes | 0.0011 | Education and training | 0.0146 |
| Auto Accessories and Supplies | 0.0002 | Dietary products | 0.0012 | Electronic game accessories | 0.0148 |
| Hardware Tools | 0.0002 | Hair Care/Wigs | 0.0012 | Attractions Tickets / Live Performances / Theme Parks | 0.015 |
| Bags leather goods / wemon handbags / men bag | 0.0002 | Network equipment/network related | 0.0012 | Online shop, web service, and software | 0.0172 |
| Home decoration products | 0.0003 | Audio and video appliances | 0.0012 | Digital products (domestic brands) | 0.0181 |
| Home textile products | 0.0003 | Jewelry / Diamond / Jade / Gold | 0.0012 | Used goods | 0.0182 |
| Men Shoes | 0.0003 | Sportswear / Casual Wear | 0.0013 | Mobile number, package, related services | 0.0187 |
| Cosmetics | 0.0003 | Computer hardware, monitors, other accessories | 0.0014 | Global delegate shopping | 0.0205 |
| Women's underwear / Men's underwear / Indoor clothes | 0.0003 | Motorcycle/Electric Vehicle/Equipment/Accessories | 0.0014 | Other food and beverage | 0.0261 |
| Toys/Carton | 0.0003 | Home devices | 0.0014 | Others | 0.0272 |
| Household furniture | 0.0003 | Maternity and nutrition | 0.0014 | Supermarket and shopping mall cards | 0.0316 |
| Tableware | 0.0004 | Pet/Pet food and supplies | 0.0015 | Cake bread and other shopping gift cards | 0.0341 |
| Outdoor and travel products | 0.0004 | Home customization | 0.0016 | Leisure and entertainment | 0.0373 |
| Jewelry / Fashion Jewelry / Fashion accessories | 0.0004 | Kitchen appliances | 0.0017 | Family services and insurance | 0.042 |
| Office equipments, consumables, and related | 0.0005 | Bicycle and related equipments | 0.0017 | Online shop payment/coupon | 0.0533 |
| Bed Linings | 0.0005 | Sports shoes | 0.0019 | Decoration design / Construction Supervision | 0.0762 |
| Electronic dictionary / electronic books / stationery | 0.0005 | Fish and meat / fresh fruits and vegetables / cooked food | 0.0022 | Mobile / Unicom / Telecom recharge center | 0.1266 |
| Clothing Accessories, belts, hats, scarves | 0.0005 | Musical instruments | 0.0026 | Online game card | 0.1663 |
| Daily household products | 0.0005 | Storage consolidation | 0.0027 | Public service and charity | 0.1777 |
| Commercial/office furniture | 0.0005 | Sports Bags/Outdoor Bags/Accessories | 0.0027 | Game Item Trading Platform | 0.209 |
| Sports/Yoga/Fitness/Sports fan products | 0.0005 | Cell phone | 0.0028 | Transportation ticket | 0.2637 |
| Tea / coffee / Drink Mixes | 0.0006 | Household cleaning products | 0.0033 | Insurance (remittance charges) | 0.2746 |
| Nursing Cleanser/Sanitary Napkin/Aromatherapy | 0.0006 | Wine and spirits | 0.0033 | Digital reading | 0.3427 |
| Flower and gardening | 0.0006 | Special crafts | 0.0033 | QQ (instant chat) service related | 0.3529 |
| Digital Accessories | 0.0007 | Laptop | 0.0034 | Property / Rent / Commission Service | 0.3763 |
| Kitchen Appliances | 0.0007 | Watch | 0.004 | New / used car | 0.4408 |
| Electronic and Electrical | 0.0007 | Flash card / U disk / storage / mobile hard disk | 0.0043 | Service market | 0.4534 |
| Personal Care / Health / Massage Equipment | 0.0007 | Food delivery services | 0.0043 | Crowdfunding | 0.5568 |
| Antique/Bills/Paintings/Collections | 0.0007 | Movies / Shows / Sports Events | 0.0046 | Taobao Business Number | 0.5644 |
| Festive supplies/gifts | 0.0007 | Digital Camera/SLR Camera/Camera | 0.0056 | Other service goods | 0.5853 |
| Snacks/Nuts/Local food | 0.0007 | MP3/MP4/iPod/recording pen | 0.0066 | Asset sale | 0.7231 |
| Children's shoes & clothes | 0.0007 | Local living services | 0.0067 | Taobao fashion model | 0.7252 |
| ZIPPO, Swiss Army Knife / Glasses | 0.0008 | Brand name machines / Web server | 0.0077 | Taobao food service coupon | 0.8237 |
| Diapers / Nursing / Feeding / Beds | 0.0008 | Holiday, visa and other travel services | 0.0078 | Taobao partner business | 0.8504 |
| Traditional nourishing products | 0.0009 | Music / Movies / Audiovisual | 0.008 | | |
| Grain, oil, rice, noodles, dry goods, spices | 0.0009 | DIY computer | 0.0096 | | |
| Perfume/Beauty products | 0.001 | Adult products / contraception product | 0.0096 | | |
| Customization/Design Services/DIY | 0.001 | Photography/camera services | 0.0103 | | |
| Basic building materials | 0.001 | Tablet/MID | 0.0122 | | |