DISCUSSION PAPER SERIES

DP15262

DATA VS COLLATERAL

Leonardo Gambacorta, Yiping Huang, Zhenhua Li,
Han Qiu and Shu Chen

FINANCIAL ECONOMICS

ISSN 0265-8003

DATA VS COLLATERAL

Leonardo Gambacorta, Yiping Huang, Zhenhua Li, Han Qiu and Shu Chen

Discussion Paper DP15262
Published 09 September 2020
Submitted 07 September 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

DATA VS COLLATERAL

Abstract

The use of massive amounts of data by large technology firms (big techs) to assess firms' creditworthiness could reduce the need for collateral in solving asymmetric information problems in credit markets. Using a unique dataset of more than 2 million Chinese firms that received credit from both an important big tech firm (Ant Group) and traditional commercial banks, this paper investigates how different forms of credit correlate with local economic activity, house prices and firm characteristics. We find that big tech credit does not correlate with local business conditions and house prices when controlling for demand factors, but reacts strongly to changes in firm characteristics, such as transaction volumes and network scores used to calculate firm credit ratings. By contrast, both secured and unsecured bank credit react significantly to local house prices, which incorporate useful information on the environment in which clients operate and on their creditworthiness. This evidence implies that a greater use of big tech credit – granted on the basis of machine learning and big data – could reduce the importance of collateral in credit markets and potentially weaken the financial accelerator mechanism.

JEL Classification: D22, G31, R30

Keywords: big tech, Big Data, Collateral, banks, asymmetric information, credit markets

Leonardo Gambacorta - leonardo.gambacorta@bis.org
Bank for International Settlements and CEPR

Yiping Huang - yhuang@nsd.pku.edu.cn
Institute of Digital Finance and National School of Development, Peking University

Zhenhua Li - sunny.lzh@antgroup.com
Ant Group

Han Qiu - hqiu93@163.com
Institute of Digital Finance and National School of Development, Peking University

Shu Chen - emily.cs@antgroup.com
Ant Group

Acknowledgements
We thank Claudio Borio, Gong Chen, Sebastian Doerr, Torsten Ehlers, Jon Frost, Catherine Koch, Amit Seru, Hyun Song Shin and participants at a BIS research seminar for useful comments and insights. We thank Giulio Cornelli for excellent research assistance. We thank Guangyao Zhu, Fang Wang, Yongguo Li, Zhiyun Chen, Jinyan Huang, Yiteng Zhai, Yanming Fang, Xin Li, Zhengjun Nie, Liang Guo, Ting Xu, Peng Liu, Ma Li, Ruiping Wu for help with the data, comments and suggestions. The views in this paper are those of the authors only and do not necessarily reflect those of the Bank for International Settlements or Ant Group. The authors

highlight that the data and analysis 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 Group other firms. All statements made reflect the private opinions of the authors and do not express any official position of Ant Group and its management. The analysis was undertaken in strict observance of the Chinese law on privacy. Yiping Huang and Han Qiu gratefully acknowledge financial support by the National Social Science Foundation of China (project number 18ZDA091). The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. Zhenhua Li and Shu Chen disclose an employment relationship in Ant Group. Ant Group did not exercise any influence on the content of this paper, but has ensured confidentiality of the (raw) data.

1. Introduction

Collateral is used in debt contracts to mitigate agency problems arising from asymmetric information. Banks usually require their borrowers to pledge tangible assets, such as real estate, to lessen ex-ante adverse selection problems (Bester 1985, Chan and Kanatas, 1985; Besanko and Thakor, 1987) or as a way to reduce ex-post frictions, such as moral hazard (Aghion and Bolton, 1997; Holmström and Tirole, 1997), costly state verification (Gale and Hellwig, 1985; Boyd and Smith 1994; Cooley et al., 2004) and imperfect contract enforcement (Albuquerque and Hopenhayn, 2004).[1]

The use of collateral is more widespread for opaque firms, such as small and medium-sized enterprises (SMEs). It is common for SME owners to pledge their homes to finance their firms (Bahaj et al, 2020). According to a recent survey conducted by the Financial Stability Board (FSB, 2019), the percentage of bank loans to SMEs that are collateralised amounts to 90% in the US, 82% in Switzerland and 65% in Canada. The percentage drops to 53% in China (OECD, 2019), where many SMEs lack basic documentation and are geographically remote from bank branches.

With the development of fintech, especially the entry of large technology firms (big techs) into the provision of financial services, non-traditional data play an increasingly important role in credit assessment for SMEs (BIS, 2019). The business model of big techs rests on enabling direct interactions among a large number of users. An essential by-product of their business is the large stock of user data, which are used as an input to offer a range of services that exploit natural network effects, generating further user activity. Increased user activity then completes the circle, as it generates yet more data. The mutually reinforcing data-network-activity (DNA) feedback loop helps big tech firms to identify the characteristics of their clients and offer them financial services that best suit their needs. As a result, big techs can have a competitive advantage over banks and serve firms that otherwise would remain unbanked. Recent work suggests that big techs' credit scoring applied to small vendors outperforms models based on credit bureau ratings and

[1] For a review of the literature on the effects of collateral in credit markets see, amongst others, Ioannidou et al (2019).

traditional borrower characteristics (Frost et al, 2019). All this could help to significantly advance financial inclusion and improve firms' performance (see Luohan Academy, 2019; Hau et al, 2018).

By leveraging artificial intelligence, big techs could address asymmetric information problems differently from banks. They can use machine learning and big data to infer the credit quality of a borrower more precisely in real time (Bazarbash, 2019). For example, Ant Group in China and Mercado Libre in Argentina claim that their credit quality assessment and granting of loans typically involve more than 1,000 data series per loan applicant. Fintech credit platforms (which include peer-to-peer and marketplace lending, and share characteristics with big tech credit) may use alternative data sources, including insights gained from social media activity (U.S. Department of the Treasury, 2016; Jagtiani and Lemieux, 2018a) and users' digital footprints (Berg et al, 2018). Moreover, monitoring can be conducted almost in real time and credit scoring quickly adjusted (Gambacorta et al, 2019). In this new way to conduct financial intermediation, the use of data could substitute that of collateral. This is the reflection of the general principle that financial intermediaries substitute information for collateral when collateral is relatively more expensive (Holmström and Tirole, 1997).

However, their access to big data is not the only potential advantage for big techs over banks. Big techs have the further advantage of being able to monitor borrowers once they are within a big tech's ecosystem. For example, when a borrower is closely integrated into an e-commerce platform, it may be relatively easy for a big tech to deduct the (monthly) payments on a credit line from the borrower's revenues that pass through its payment account. This is useful in enforcing repayment and reducing the moral hazard problem. By contrast, banks may not be in a position to do likewise as the borrower could have accounts with other banks. Given network effects and high switching costs, big techs could also enforce loan repayments by the simple threat of a downgrade or maybe an exclusion from their ecosystem if in default. Anecdotal evidence from Argentina and China suggests that the combination of massive amounts of data and network effects may allow big techs to mitigate the information and incentive problems that are traditionally addressed through the posting of collateral. This could affect the financial accelerator mechanism, by which developments in financial markets and asset prices amplify the effects of changes on the real economy. Collateral is often blamed for amplifying the business cycle, through the so-called collateral channel (Bernanke and Gertler, 1989; Kiyotaki and Moore 1997).

The aim of this paper is to address the following three questions.

i. Do big tech and bank credit react differently to collateral value, local economic conditions and firm-specific characteristics?
ii. How could the increased use of big data and machine learning in solving asymmetric information problems, in lieu of collateral, impact the financial accelerator?
iii. Do big tech platforms matter? Are there differences between credit granted to firms that operate in the ecommerce platform (online) and credit granted to firms that operate on traditional business channels (offline)?

To answer these questions, we use a unique dataset that compares the characteristics of loans provided by MYbank, one of the brands under Ant Group (one of the most important big tech companies in China) with loans supplied by traditional Chinese banks. In particular, we analyse the credit provided to a random sample of more than 2 million Chinese firms in the period 2017:01-2019:04. Differently from the previous literature (Hau et al, 2018), the sample of firms used in our study contains not only firms on Alibaba's e-commerce platforms (online firms) but also those that use more traditional business channels (offline firms). The latter use the Alipay app for mobile payments, through the so-called Quick Response (QR) code, but are not fully integrated into the e-commerce platform.

From Ant Group, we obtain access to detailed information on credit supplied by MYbank[2] and firm characteristics on a monthly frequency. In particular, we have access to credit data (quantity and price), and specific information used to model firms' creditworthiness, such as vendor transaction volumes and their network score. The latter measures users' centrality in the network and is based on their payments history and social interactions in

---

[2] All the data remained located at the Ant Group headquarters and the regression analysis was conducted onsite without the need to export the raw data.

the Alipay ecosystem. These pieces of information are then combined with the bank credit history of the client, where we distinguish between secured (backed by collateral) and unsecured (without collateral) bank loans.

The comparison between big tech and unsecured bank credit is particularly relevant. Although both MYbank credit and bank unsecured credit are not backed by collateral, the risk assessments of these two types of credit are different. MYbank credit scoring is based on granular firm-specific information collected in the Alipay ecosystem, while traditional banks rely more on soft information about the client or the business conditions of the region in which the firm is headquartered. To our knowledge, this is the first study that compares the characteristics of big tech credit with bank credit for the same set of firms.

The main results of the paper are the following.

i. Big tech credit does not correlate with local business conditions and house prices when controlling for demand factors, but reacts strongly to firm characteristics, such as transaction volumes and the network score that are used to calculate firm credit ratings. By contrast, both secured and unsecured bank credit react significantly to local house prices, which incorporate useful information on the business conditions in which clients operate and on their creditworthiness.

ii. An increased use of big tech credit – granted on the basis of big data analysis rather than the use of collateral – could weaken the financial accelerator mechanism, which amplifies the effects of shocks to the real economy by means of loan supply shifts caused by changes in collateral values.

iii. Big tech credit to online firms, fully integrated in the e-commerce platform, is more strongly correlated with transaction volumes and network scores than it is in the case of offline firms. Big tech credit to offline firms shows some sign of correlation with local demand conditions.

The rest of the paper is organised as follows. Section 2 provides a concise literature review and discusses the contribution of our paper. Section 3 presents the data and describes some stylised facts. Section 4 explains our empirical strategy and how we tackle identification issues. Section 5 presents the main results and robustness tests. Section 6 summarises the main conclusions. Annex A reports some facts on Ant Group.

2. Related literature

We contribute mostly to three broad strands of literature. First, we provide new supportive evidence on the characteristics of big tech credit, the way it could contribute to increasing financial inclusion and how it could improve risk assessment. Overall, the evidence suggests fintech is growing where the current financial system is not meeting demand for financial services. For the case of China, Hau et al (2018) show that fintech credit mitigates supply frictions (such as a large geographic distance between borrowers and the nearest bank branch), and allows firms with a lower credit score to access credit. In the United States, Tang (2019) finds that fintech credit complements bank lending for small-scale loans. Jagtiani and Lemieux (2018b) find that Lending Club has penetrated areas that are underserved by traditional banks. In Germany, De Roure et al (2016) find that fintech credit serves a slice of the consumer credit market neglected by German banks. Frost et al (2019) use data for Mercado Crédito that provides credit lines to small firms on the e-commerce platform Mercado Libre in Argentina. They find that credit-scoring techniques based on big data and machine learning have so far outperformed credit bureau ratings in terms of predicting loss rates. Cornelli et al (2020) find that fintech and big tech credit are higher where banking sector mark-ups are higher, where there are fewer bank branches and where banking regulation is less stringent. These papers do not analyse the specific role of data in substituting for collateral in credit provision nor the implications for the monetary transmission mechanism.

Second, we contribute to the empirical literature that studies asymmetric information problems in credit markets. In this stream of the literature, collateral plays a key role in mitigating the financial constraints for the development of economic activity (Bernanke and Gertler, 1989; Kiyotaki and Moore, 1997; Bernanke et al 1999). Gan (2007) shows that the value and the redeployment ability of collateral affect real estate prices and corporate investment. Schmalz et al (2016) find that an increase in collateral value (proxied by house price) leads to a higher probability of becoming an entrepreneur. More recently, some papers find that transaction volume-based lending is also relevant. Lian and Ma (2019) supplement Dealscan data on commercial loans with a variety of data sources and present detailed evidence that the borrowing of US non-financial firms correlates with their transaction volumes, as measured by earnings. Different forms of constraints also have different implications for credit allocation and efficiency, responses to monetary policy, economic recovery and the rise of intangible capital (see, among others, Lorenzoni, 2008; Dávila et al, 2017; and Diamond et al, 2018). Our paper investigates a new mechanism that could reduce financial constraints for SMEs: the use of big data and the presence of network effects rather than collateral could provide a different solution to solve agency problems between the lender and the borrower.

Third, our paper contributes also to the empirical literature that has studied how the collateral channel could affect the macroeconomy (Gertler and Gilchrist, 1994; Jeenas, 2018; Cloyne et al., 2018). The use of data instead of collateral for the analysis of creditworthiness could have important implications for the credit channel and the macroeconomy. One example is the link between asset prices and the business cycle. A rise in collateral values during the expansionary phase of the business cycle fuels a credit boom, while their subsequent fall in a crisis weakens both the demand and supply of credit, leading to a deeper recession. The collateral channel is considered as one of main driver of the Great Depression (Bernanke, 1983), and as an important factor behind the more recent great financial crisis (GFC) (Mian and Sufi, 2011; Ottonello and Winberry 2018; Bahaj et al., 2019). Adelino et al. (2015) and Doerr (2019) further show the importance of real estate collateral for small business employment. Indeed, the GFC has shown that the most serious consequences of the drop in the value of collateral were for SMEs that do not have well-diversified funding conditions (Lian and Ma, 2019; OECD, 2019). Using a structural model Ioannidou et al (2019) shows that a 40% drop in collateral values would lead almost a quarter of loans to become unprofitable, a reduction of average demand by 16% and a drop in banks' expected profits of 25%. Our paper contributes to this stream of the literature by analysing how big techs' use of big data for credit scoring could attenuate the link between collateral value (house price) and credit supply.

3. Data and stylised facts

The empirical analysis in this paper considers Chinese SMEs that obtained credit from MYbank, one of the brands under Ant Group.[3] For these firms we also observe all loans provided by traditional banks, and distinguish between collateralised credit (secured bank credit) and uncollateralised credit (unsecured bank credit).

The database is constructed at the firm-month level over the period 2017:01 to 2019:04. The sample includes more than 2 million firms and has been randomly selected from all firms that had transaction records every month and obtained bank credit since January 2017.

Table 1 presents the summary statistics, divided into three panels: A) big tech credit; B) secured bank credit; C) unsecured bank credit. For big tech credit, we have more than 7 million firm-month observations. Most of the 2 million MYbank borrowers have access only to big tech credit and do not have a bank relationship. However, around 47,000 borrowers also have access to secured bank credit and 120,000 to unsecured bank credit, for around 95,000 and 399,000 observations, respectively. Each panel in Table 1 includes information on: i) firms' characteristics; ii) entrepreneurs' characteristics; and iii) economic and financial conditions where the firm is headquartered. We winsorised all firm and entrepreneur variables at the 1% and 99% level to eliminate outliers.

i. Firms' characteristics

The enterprise data include transaction volumes and credit data. The latter is the actual credit used by the enterprise in a given month. For robustness, we also run some regressions using the overall amount of credit granted by the big tech. Unfortunately, this information is not available for bank credit.

The median credit volume for big tech borrowers is RMB 6,900 (USD 975), reflecting the micro nature of MYbank credit and the short maturity of the contract. Big tech credit is typically granted for short periods (from 1 month to one year) and then renewed several times, as far as the credit approval remains in place. Often big tech credit assumes the form of a credit line. The median unsecured bank credit is of RMB 60,000 (USD 8,500). The

[3] More information is provided in Annex A.

larger size of the loan could reflect a greater length of the contract (from 1 to 3 years). By contrast, the differences in firm size between big tech and bank credit users are not large. The median monthly transaction volume of firms that use big tech credit is RMB 3,000 (USD 425), while that for firms that also use unsecured bank credit is RMB 4,300 (USD 610). Interestingly, the median firm that uses big tech credit is less connected in the big tech ecosystem (the network score is 58, against an average of 63 for firms that use unsecured bank credit).[4]

Firms that use secured bank credit are slightly larger; the median transaction volume is RMB 5,400 (USD 770) and with a higher network score (66). The median bank secured credit is RMB 300,000 (USD 42,400). Given the presence of collateral, this also reflects the fact that these loans are typically associated with more important investment decisions by the firm.

ii. Entrepreneur information

For SMEs, information about the entrepreneur (typically the owner of the firm or the store) is very important for risk assessment. On the one hand, SMEs have a short life cycle, so the firm information might not be adequately accumulated. On the other hand, the financial situation of SMEs tends to relate very closely to those of the owner. One of the advantages of MYbank in providing risk control measures for SMEs is that Ant Group can obtain the firm information as well as the information of owners. In this paper, we are able to merge these two different sets of information. Borrowers who access big tech credit are slightly younger (the median age is 31 years) than the owners of firms that use unsecured bank credit (36 years) or secured bank credit (38 years).

Another relevant information is the borrowers' level of income. This information is not directly observed by Ant Group, but can be inferred by the total amount of deposits into the Alipay wallet. In particular, we have used this proxy to split the borrowers into three

[4] To mitigate concerns about differences in contractual characteristics and firm size we have run regressions for firms that use all forms of credit, including in the models both time*credit type fixed effects and borrower*credit type fixed effects (see Section 5.5). This allows us to control for the possibility that the relationship between a firm and the big tech is different with respect to the relationship between the same borrower and the bank.

groups: 1 = low income; 2 = medium income; 3 = high income. It should be noted that all entrepreneurs' characteristics are taken at the date of the issuing of the loan.

The main variables of interest used in this paper are firms' transaction volumes and their network scores, which have a crucial role in the credit scoring analysis of MYbank. These variables are time-varying and can be used in our preferred econometric model with borrower fixed effects.

Figure 1 reports the unconditional elasticity between credit and transaction volume from a random sample of 100,000 firms served by both MYbank and traditional commercial banks. The figure is divided into three panels: the left-hand panel plots big tech credit, the middle panel plots secured credit, while the right hand panel plots unsecured bank credit. Linear trend lines are reported in each graph, together with 95% confidence bands. Interestingly, the elasticity is 0.15 for big tech credit, 0.09 for secured bank credit and 0.12 for unsecured bank credit, in line with the intuition that big tech credit is more responsive to changes in a borrower's business conditions. Banks observe transaction volumes with less precision and with a lag.

Figure 2 evaluates the elasticity between big tech credit and the transaction volume, distinguishing between online and offline borrowers. The yellow dots and the yellow line indicate the offline borrowers (those with a QR code, but not trading in the e-commerce platform), while the blue dots indicate online borrowers (those integrated in the e-commerce platform). The elasticity is 0.090 for offline borrowers and 0.407 for online borrowers. The difference reflects the fact that the big tech firms are able to efficiently collect and process information from online lenders that are integrated in the big tech ecosystem. Therefore, they have access to a rich set of additional data to be combined with traditional transaction volumes obtained from payments.

Figure 3 evaluates the unconditional elasticity between the big tech credit used (ie the value of credit drawn down) and the network score. The network score is calculated to measure users' centrality in the big tech ecosystem on the base of payment data, users' financial investments and social interactions. It is worth stressing that both offline and online vendors have a network score because payment data and social interaction information are obtained from Alipay. A user with more connections in the big tech ecosystem has a higher network score. Here, too, the left-hand panel plots big tech credit, the middle panel plots secured credit and the right hand panel plots unsecured bank credit. The elasticity is 0.83 for big tech credit, 0.28 for secured credit and 0.30 for unsecured credit. This is not surprising because the network score is not directly observed by the bank and proxies (other) soft information obtained by the bank credit officer on the firm.

Figure 4 plots the correlation between credit and the network score, but distinguishing between offline and online firms. The yellow dots and the yellow line indicate the offline borrowers (those with a QR code, but not integrated in the e-commerce platform), while the blue dots indicate the online borrowers (those perfectly integrated in the ecommerce platform). Credit reveals a positive and significant elasticity with network effect that is approximately same for online and offline borrowers (respectively 1.120*** and 1.187***). This preliminary evidence shows that the network measure is extremely important for credit scoring, also for those firms that do not conduct their main activity in the e-commerce platform.

iii. House prices, GDP and monetary policy

The data source on house prices is the 100-city housing prices published by China Index Academy and included in the WIND database. The data covers 100 samples of new houses for sale in China, including commercial housing, villas, and affordable housing, and all the houses for sale with a sales license were included in the calculation. According to the available data, the 100-city housing price data is the database with the largest coverage of monthly housing prices in China. China Index Academy has published the *China Real Estate Statistical Yearbook* for 16 consecutive years with State Statistics Bureau.

Figure 5 indicates the unconditional elasticity between the different credit forms and house price. The dots in the figures indicate the average logarithm credit use (y-axis) and the average logarithm of housing price (x-axis) at the city-year level. The left-hand panel plots big tech credit, the middle panel plots bank secured credit and the right hand panel plots bank unsecured credit. Linear trend lines are reported in each graph, together with 95% confidence bands. The (unconditional) elasticity of big tech credit with respect to house prices is 0.09, while that of unsecured bank credit is twice as high (0.184) and that of secured bank credit is five times higher (0.488).

The different elasticities of the three credit types with respect to house prices remain quite stable even controlling for different local GDP conditions and including a complete set of time fixed effects. Figure 6 reports the different elasticities and associated standard errors for this simple model. Interestingly, the elasticity of unsecured bank credit with respect to GDP at the city level is more than three times higher (0.147) than that of big tech credit (0.041). The elasticity of bank secured credit with respect to local GDP is not statistically different from zero.

4. Econometric strategy

Our analysis starts with a simple model that analyses the main determinants of credit. We consider the following baseline model:

$$ln\left(credit_{i,j,t}\right) = \mathrm{A}'X_{i,j,t} + \Gamma'Y_{j,t} + \mu_j + \mu_T + \varepsilon_{i,t} \quad (1)$$

where $ln\left(credit_{i,j,t}\right)$ is the logarithm of the credit granted by MYbank or traditional banks (secured and unsecured) to firm *i*, headquartered in city *j*, in time *t*. $X_{i,j,t}$ is a vector that contains time-variant firm characteristics (transaction volume, network score) and owners' time-invariant characteristics (age, income). $Y_{j,t}$ are the city-level indicators to capture regional conditions, including log of house price and local GDP.

The model (1) includes time (μ_T) and city (μ_j) fixed effects and $\varepsilon_{i,t}$ is an error term. Following Chaney et al (2012), we cluster the standard errors at the city-month level.

We also consider a model that includes firm fixed effects (μ_i).

$$ln\left(credit_{i,j,t}\right) = \mathrm{A}'X_{i,j,t} + \Gamma'Y_{j,t} + \mu_i + \mu_T + \varepsilon_{i,t} \quad (2)$$

This model is able to control for all firm (unobserved) invariant characteristics but at the cost to limit the analysis to firms that received at least two loans in the period under investigation. Due to the inclusion of firm fixed effects, we cannot include in vector $X_{i,j,t}$ the time-invariant firm characteristics (age and income).

We run model (2) for the three different types of credit and compare the coefficients of housing price and local GDP to evaluate if big tech and bank credit (secured and unsecured) react differently to business and asset price conditions. Moreover, we can distinguish the effects between offline firms and online firms in order to verify the additional effects if any for firms that are fully integrated in the big tech ecosystem. As big tech credit could take different contractual forms, we also run model (2) within homogenous credit products categories to check for the possible existence of aggregation biases.

Another identification challenge derives from the fact that the customers of the big tech company could be very different from those of a traditional bank and it could be in principle very difficult to compare them. To address this concern, we select the firms who both have big tech credit and traditional bank credit and use a difference-in-difference approach that follows Khwaja and Mian (2008). This approach allows us to compare the characteristics of credit from different sources used by the same customer. In particular, we used a nested model in which big tech credit and bank credit (secured or unsecured) are jointly analysed. In this case, we include both time*credit type (μ_{TC}) fixed effects and borrower*credit type (μ_{iC}) fixed effects. The inclusion of these fixed effects controls for the fact that the relationship between a borrower and the big tech firm could be quite different with respect to the relationship between the same borrower and the bank. The inclusion of these fixed effects necessitate for each firm to have at least two big tech credits and two bank credits over the sample horizon. In particular, we run the following model:

$$ln\left(credit_{i,j,t}\right) = A'X_{i,j,t} + \Gamma'Y_{j,t} + B'X_{i,j,t} * credit\_type + \\ \mathrm{K}'Y_{j,t} * credit\_type + \mu_{iC} + \mu_{TC} + \varepsilon_{i,t} \qquad (3)$$

The different reaction of bank credit with respect to big tech credit is evaluated by interacting a $credit\_type$ dummy that takes the value of one for bank secured (or bank unsecured) credit and 0 for big tech credit. The test for the difference between the coefficients is given directly by the sign and the significance of the interaction terms (B' for the borrower specific characteristics and K' for the local economic conditions).

Another identification challenge is the necessity to properly control for demand shifts. Model (3) can be further enriched to control for specific changes in the economic conditions at the city level that could affect the credit market. One possibility is to integrate the model with city*time fixed effects (μ_{LT}) to control for changes in local conditions over time. In doing so, however, the local economic indicators are subdued $Y_{j,t}$ by the city*time fixed effect. The model becomes:

$$ln\left(credit_{i,j,t}\right) = \mathrm{A}'X_{i,j,t} + \mathrm{B}'X_{i,j,t} * credit\_type + \mathrm{K}'Y_{j,t} * credit\_type + \mu_{iC} + \mu_{TC} + \mu_{LT} + \varepsilon_{i,t} \quad (4)$$

In this model we can focus our attention on the significance of the interaction terms B' (or K') to evaluate a different reaction of credit types to borrower specific characteristics (local economic conditions).

Following Jimenez et al (2014), an alternative way to control for shifts in demand is to progressively saturate model (3) with Time*Borrowers (μ_{TB}) fixed effects, together with City*$credit\_type$ (μ_{LC}) and Time*$credit\_type$ (μ_{TC}) fixed effects. In this way, we use borrower*time to absorb all time-varying, observed and unobserved firm heterogeneity. City*$credit\_type$ and Time*$credit\_type$ control the location-varying and time-varying heterogeneity of different credit types. This specification allows us to control more precisely for borrower specific demand shifts but necessitates a further restriction of the sample to consider only firms that have in place both big tech credit and bank credit in one month. We estimate the following regression:

$$ln\left(credit_{i,j,t}\right) = \mathrm{A}'X_{i,j,t} + \Gamma'Y_{j,t} + \mathrm{B}'X_{i,j,t} * credit\_type + \mathrm{K}'Y_{j,t} * credit\_type + \mu_{TB} + \mu_{LC} + \mu_{TC} + \varepsilon_{i,t} \quad (5)$$

A final concern is reverse causality. In principle, credit expansion by (large) firms may also affect house prices; however, we argue that this concern is unlikely to affect our results because the firms in our sample are relatively small. Following Chaney et al (2012), we also instrument the house price by using a hand-collected monthly measure for land supply by the Government and its interaction with mortgage rates in order to insulate price movements that are (exogenously) driven by supply changes.

5. Results

5.1 Baseline model

Table 2 reports the results of model (1) that includes time-invariant borrower characteristics (age, income), excluding borrower fixed effects. The dependent variable in the first column is the Log of MYbank credit, while the other two columns report the results for the Log of secured bank credit and the Log of unsecured bank credit, respectively.

It is worth stressing that we consider as explanatory variables for bank credit also transaction volumes and network score that are variables obtained from MYbank database. Transaction volumes, however, are also visible to banks. Each vendor has indeed the possibility to print and report to the bank credit officer a detailed documentation of the activity developed using the Alipay payment system or the e-commerce platform, but the reporting activity could be subject to a lag and cannot be directly updated by the bank. The network score is not observed by the bank but it is probably correlated with bank soft information available to the bank.

The results in Table 2 indicate that big tech credit is not correlated with house prices and local economic conditions. By contrast, bank credit is significantly correlated with house prices (especially secured credit) and more weakly with local economic activity (only unsecured bank credit). The correlation of unsecured bank credit to local business conditions could reflect supply factors (local GDP indicates a higher credit worthiness by firms) or demand factors (GDP reflects higher local activity). In the following analysis we will try to disentangle these two effects.

Big tech lending is strongly correlated with firm-specific variables (transaction volumes and network score) reflecting the nature of the financial services provided by big techs that is tailored towards the characteristics of their client. The correlation of bank credit (secured and unsecured) with respect to firm-specific characteristics is weaker than that displayed by big tech credit, especially when considering the network score.

The other control variables indicate that the provision of all three forms of credit, other things equal, is positively correlated with income and age. The results do not change considering a non-monotonic (concave) relationship with age.

Table 3 focuses on the drivers of big tech credit using model (2) that includes borrower-specific fixed effects. This allows us to control for unobservable (time invariant) client's characteristics. The table also splits the borrower in two groups: offline borrowers (those with a QR code, but not trading on the ecommerce platform) and online borrowers (those integrated in the ecommerce platform).

The left hand panel of the table reports the results when the log MYbank credit used by the borrowers is considered as dependent variable. Comparing the first column of Table 2 and the first column of Table 3, we can notice that the adjusted R^2 increases from 0.242 to 0.635. This means that around 40% of MYbank credit variability is captured by borrower fixed effects.

Also using this different model, big tech credit is not correlated with house prices but becomes significantly correlated with local economic conditions. Interestingly, the positive correlation between big tech credit and city-level GDP is significant only for firms that work offline, while it is not different from zero for firms that work on the e-commerce platform. This could be explained by the fact that that while offline firms (ie a restaurant or a shop) depend on local business conditions, the activity of firms that sell their products on the e-commerce platform does not necessarily depend on the economic activity where the firm is headquartered. The positive correlation between big tech credit and local business conditions for offline firms should reflect mostly demand factors.

Big tech credit continues to be highly correlated with borrower-specific characteristics (transaction volumes and network score). The correlation is lower for firms that work offline than for firms that work online. The latter are indeed more integrated in the big tech ecosystem and MYbank could get more information on them.

The analysis presented so far has considered as dependent variable the credit used by the firm. In order to grasp more information on the credit lines supplied by MYbank we consider in the right panel of Table 3, the value of the credit line granted by MYbank as dependent variable. Credit lines represents around 20% of total credit contracts and indeed the number of observations drops from 7.1 to 1.3 million.

The results remain qualitatively similar, with the notable exception that now local GDP and house prices are not significantly correlated with big tech credit lines granted, both for offline and online borrowers. As now the dependent variable represents big tech credit supply, the result reinforces the interpretation that big tech credit assessment focuses more on firm-specific information rather than on local economic conditions (which more closely reflect changes in the demand for credit).

5.2 Homogenous big tech credit contracts and interest rates

The results presented in the first three columns of Table 3 refer to the overall value of big tech credit used, aggregating credit contractual forms that could be quite different. It is important therefore to replicate the analysis considering specific and more homogenous forms of credit contracts. This will allow us to test for the possible existence of biases related with the aggregation of credit with different contractual conditions.

In Table 4, therefore, we report the regressions for model (2) for two particular products offered by MYbank to firms. The left hand side of the table considers a very popular credit product (Product 1) that is directly accessible to the firm in very simple steps, using a smartphone for instance. The application for the credit is completed with a few taps on the screen and no collateral is required. This contractual form is particularly used by offline firms (QR code merchants). MYbank offers a credit line for each merchant that is based on her specific risk profile. As long as the (offline) merchants use Alipay QR code to collect payments, they will have the opportunity to obtain and renew the credit.

The second credit contract (Product 2) is offered by MYbank to firms on the basis of the overall value of their orders and receivables in the Taobao platform. As the information is obtained on the e-commerce platform, this contractual form is used by online firms. This product is a trade credit product. Every vendor can access a credit line from MYbank, but the amount of credit granted to each customer is determined by accounts receivable.

Table 4 is divided into two parts. The first two columns consider as dependent variables the quantity of credit, while the last two columns report results for regressions where the interest rate is the dependent variable. Neither form of big tech credit is correlated with house prices nor with local economic conditions. By contrast, they are highly correlated with borrower-specific characteristics (both transaction volumes and network score). The correlation is higher for Product 2, used by firms that work online, while the correlation for Product 1, used by firms that work offline, is lower.

When considering the interest rate as the dependent variable (see the third and fourth columns of Table 4) the results parallel those obtained on quantities. Interest rates do not react to the evolution of house prices and local economic conditions. By contrast, price conditions react to borrowers' specific characteristics, more strongly for Product 2 that is available for firms that operate online on the big tech e-commerce platform.

5.3 Main drivers of bank lending: collateralised vs uncollateralised contracts

Table 5 reports the result of model (2) for bank credit. The left hand side of the table reports the results for secured bank credit. As expected, the latter is highly correlated with house prices, while it is not correlated with local economic conditions. Moreover, secured bank credit is not correlated with borrower-specific characteristics (transaction volumes and network score). Collateralised bank credit shows some signs of correlation only with respect to transaction volumes for firms that work online. However, we will see later that this result vanishes using more complete specifications that control for demand shifts.

The right hand side of Table 5 shows that unsecured bank credit is also correlated with house prices but the elasticity is significantly lower than for secured bank credit (the elasticities are 0.21 and 0.59, respectively). The positive correlation between unsecured bank credit and house prices could reflect higher demand in cities with higher asset prices, with the latter reflecting in general better economic conditions. We will try to filter out this effect in Section 5.6. Another explanation may be that banks do not have enough granular information on the firm, so local house price dynamics turn out to be one relevant indicator to identify a firm's creditworthiness. Interestingly, the correlation is not statistically significant for firms that operate online and for which local conditions are less relevant. Unsecured bank credit is instead positively correlated with local GDP conditions for reasons similar to those discussed above for house prices. Unsecured bank credit is weakly correlated with borrower-specific characteristics, especially for online borrowers.

This could reflect the fact that online vendors' activity is less visible to banks than that of offline vendors (for example, a restaurant or a shop). The physical presence in the territory could indeed be relevant for a bank credit officer who could observe more directly firms' characteristics.

5.4 Endogeneity issues

In principle, there are potential sources of endogeneity in model (2) and house prices could affect credit through channels other than rising collateral values. This could happen for three reasons. The first one is a simple reverse causality argument: large firms may have a non-negligible impact through the demand for local labour and locally produced goods on local activity, so that an increase in credit demand for such large firms could trigger also a housing price appreciation. This would lead us to overestimate the coefficient on housing price. Second, it could be that our measure of housing prices proxies for local demand shocks that are not fully captured by local GDP conditions. Third, expansion in credit may also have effects on house prices (Favara and Imbs, 2015).

The first issue is unlikely to affect our results because the firms analysed in this study are of small dimension and their credit decisions are unlikely to affect local output via increase in local labour and/or increase in produced goods. Furthermore, we have winsorised all the firms and entrepreneurs' variables at the 1% and 99% level to eliminate the effects of outliers. On the second issue, we have used city*time fixed effect and borrower*time fixed effect to control for shifts in the demand side. The inclusion of borrower*$credit\_type$ in nested models will allow us to control for a heterogeneous demand schedules for big tech and bank credit for the same client (see below Section 5.5).

To address the third issue we instrument the housing price. Table 6 presents the results of the first stage regression where we use one-year lagged land supply and its interaction with mortgage rates as instrumental variables. The local government has a great influence on housing prices through the land supply in China (Glaeser et al, 2017). The literature on the determinants of house prices for China indeed uses information about land as an instrumental variable to model housing price. For example, Hau and Ouyang (2018) use the lagged value of the surface of newly useable residential land scaled by the size of the existing housing stock and local population density. Andrew et al (2020) use the lagged volume (in square kilometres) of cumulative land sales in each city.

In line with these papers, we use a hand-collected monthly measure for land supply. In particular, we have calculated for each month the annual cumulative measure for land supply for each local government scaled by urban construction land. Our measure represents an improvement over the other measures indicated above, such as land sales or the proxies for local government land supply. In our empirical model, following Chaney et al (2012), we also include the interaction between the land supply measure and mortgage rate as an instrumental variable. This should control for differential price effects in different cities caused by a different sensitivity to monetary policy conditions. The results in Table 6 indicate that, as expected, land supply has a negative effect on house prices. When mortgage rates decrease, house prices of cities with higher land supply increase by less. There may be one concern about the endogeneity of the mortgage rate. In particular, mortgage rates could be correlated to local conditions. However, the mortgage rate used in our first stage regression is nationwide and highly correlated to the benchmark interest rate controlled by the People's Bank of China and hence, for practical purposes, exogenous to local conditions.

Table 7 presents the results of model (2) on big tech credit, secured bank credit and unsecured bank credit, using the log house prices instrumented in Table 6. Only bank credit is significantly correlated with house prices; secured bank credit is significant at the 95% level, while unsecured bank credit is significant at the 90% level. This result underscores that in the case of an (exogenous) increase in the value of collateral triggered by an expansion in the supply of land by the government, there is no positive effect on big tech credit. This result is interesting because it indicates a reduction of the effectiveness of the financial accelerator in case of big tech credit. Big tech credit also remains less correlated to local economic conditions than unsecured bank credit. The other results on firm-specific characteristics hold.

5.5 Nested models.

As discussed in Section 4, the comparison between the coefficients of models (1) and (2) across the different credit types is difficult because the estimations are derived from groups of firms with different characteristics.

Table 8 presents the results of the nested model (3) in which big tech credit and secured bank credit are jointly analysed. To check for unobservable characteristics we include in this case both Time*$credit\_type$ fixed effects and borrower*$credit\_type$ fixed effects. The different reaction of each form of credit with respect to the explanatory variables is evaluated by interacting the latter with a dummy variable ("Bank secured") that takes the value of one for bank secured credit and 0 for big tech credit. The test for the difference in the coefficients between the two different forms of credit is given directly by the sign and the significance of the interaction term. For example, big tech credit does not correlate with house price (the coefficient is -0.061 with a standard error of 0.149), while secured bank credit does (-0.061+0.633=0.572), with the difference between fintech credit and secured bank credit that is statistically significant (0.633***).

The other results are confirmed. Big tech credit and bank secured credit are not correlated with local economic conditions. Big tech credit is highly correlated with borrower-specific characteristics (transaction volumes and network score), more for e-commerce firms that work online. By contrast, bank secured credit is not correlated with borrower-specific characteristics (transaction volumes and network score).

Table 9 presents the results of the nested model (3) in which big tech credit and unsecured bank credit are jointly analysed. In this case, as well, the different reaction of each form of credit with respect to the explanatory variables is evaluated by interacting the latter with a dummy variable ("Bank unsecured") that takes the value of one for bank unsecured credit and 0 for big tech credit. The test for the difference in the coefficients between the two different forms of credit is given directly by the sign and the significance of the interaction term. For example, fintech credit does not correlate with house price (the coefficient is -0.038 with a standard error of 0.071), while unsecured bank credit does (-0.038+0.188=0.150), with the difference between fintech credit and unsecured bank credit that is statistically significant at the 10% level (0.188*). The other results are qualitatively similar to those already reported.

5.6 Additional controls for changes in local conditions and demand shifts

Table 10 presents the comparison between big tech credit and bank credit including additional controls for local condition or borrowers' demand shifts. Indeed, one concern for our results is that the evolution of quarterly GDP at the city level is not sufficient to fully capture the effects on firms' demand. This could be particularly important for offline firms that are more affected by local economic conditions. We report therefore in the first two columns of Table (8) the results using equation (4) that includes Time*city fixed effect (together with Borrower*$credit\_type$ and Time*$credit\_type$ fixed effects) to control for unobserved (to the econometrician) change in local conditions.

The right hand part of Table 10 considers instead equation (5) with a complete set of Time*borrower fixed effects (and also City*$credit\_type$ and Time*$credit\_type$ fixed effects). These controls are more stringent and their inclusion does not allow us to keep in the specification time-varying macroeconomic and borrower characteristics. In this case, we simply focus on the interaction terms between each variable and the credit type dummy. Moreover, using this specification, we need to further restrict the number of observations as the analysis can only be carried out for borrowers who have both big tech credit and bank credit in one month.[5]

Even after controlling more appropriately for demand shifts, (secured and unsecured) bank credit is more correlated than big tech credit with respect to house prices, and the difference is particularly high for the more restrictive model (5) that includes Time*borrower fixed effects. Controlling for demand shifts, the correlation of unsecured bank credit with local economic condition is more similar to that of big tech. By contrast, the latter remains significantly more correlated with borrower specific characteristics (transaction volumes and network score) than the two forms of bank credit.

---

[5] It is worth remembering that in the first and third columns the dummy variable Bank credit takes the value of 1 for bank secured credit and 0 for big tech credit. Vice versa, in the second and fourth column the dummy variable Bank credit takes the value of 1 for bank unsecured credit and 0 for big tech credit.

6. Conclusions

The use of massive amounts of data by large technology firms (big techs) to analyse the creditworthiness of borrower firms could replace the role of collateral in solving asymmetric information problems, with significant implications for the macroeconomy and the conduct of monetary policy.

Using a unique dataset of more than 2 million Chinese firms that received credit from both an important big tech firm (Ant Group) and traditional banks, this paper investigates how these different forms of credit correlate with house prices, local business conditions and firm characteristics. We find that big tech credit does not correlate with house prices, but reacts strongly to firm-specific characteristics, such as transaction volumes and a network score used to calculate firm credit ratings. By contrast, both secured and unsecured bank credit react significantly to local house prices, which likely reflect useful information on the business conditions in which firms operate and on their creditworthiness.

These results could have important macroeconomic implications. They indicate that the provision of big tech credit tends to reduce the effectiveness of the financial accelerator mechanism, because the provision of credit is detached from the movement of asset prices. Big tech credit is indeed less dependent on the financial cycle than is traditional bank credit. This could have relevant effects for the stability of SME financing. For example, bank credit could be tightened or made more expensive in response to a negative shock to asset prices, but big tech credit to SMEs should be less affected. Moreover, big tech credit seems to be correlated with local GDP conditions only for offline vendors, while firms operating on the e-commerce platform are not influenced by local economic activity and are less subject to local demand shocks. On the other hand, big tech credit (especially those of online firms) is more sensitive to the recent performance of SME borrowers.

A credit supply that is based on data analysis rather than use of collateral could have significant implications for the monetary transmission mechanism. While the financial accelerator mechanism implies that monetary policy impulses are transmitted to bank credit supply via changes in the value of collateral, this channel no longer operates in the case of big tech credit.

References

Adelino, M., A. Schoar, F. Severino (2015). House prices, collateral, and self-employment. Journal of Financial Economics, 117(2), 288-306.

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

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

Bahaj, S., G. Pinter, A. Foulis, P. Surico (2019). Employment and the collateral channel of monetary policy. Bank of England Staff Working Paper 827

Bahaj S., A. Foulis, G. Pinter (2020). Home Values and Firm Behavior, American Economic Review, 110(7), 2225-70.

Bank for International Settlements (2019). Big techs in finance: Opportunities and Risks, BIS Annual Economic Report, June.

Bazarbash, M. (2019). FinTech in Financial Inclusion – Machine Learning Applications in Assessing Credit Risk. IMF Working Paper, 109.

Berg, T., V. Burg, A. Gombović, M. Puri (2018). On the rise of FinTechs–Credit scoring using digital footprints. NBER Working Paper 24551.

Bernanke, B. (1983). Nonmonetary effects of the financial crisis in the propagation of the great depression. American Economic Review 73(3), 257–276.

Bernanke, B., M. Gertler (1989). Agency costs, net worth, and business fluctuations. American Economic Review, 79(1), 14–31.

Bernanke, B. S., M. Gertler, S. Gilchrist (1999). The financial accelerator in a quantitative business cycle framework. Handbook of macroeconomics, 1, 1341-1393.

Besanko, D., A.V. Thakor (1987). Competitive equilibrium in the credit market under asymmetric information. Journal of Economic Theory, 42(1), 167–182.

Bester, H. (1985). Screening vs. rationing in credit markets with imperfect information. American Economic Review, 75(4), 850-855.

Boyd, J. H., B.D. Smith (1994). How good are standard debt contracts? Stochastic versus non-stochastic monitoring in a costly state verification environment. Journal of Business 67(4), 539–561.

Chan, Y.S., G. Kanatas (1985). Asymmetric valuations and the role of collateral in loan agreements. Journal of Money, Credit and Banking, 17(1), 84–95.

Chaney, T., D. Sraer, D. Thesmar (2012). The collateral channel: How real estate shocks affect corporate investment. American Economic Review, 102(6), 2381-2409.

Cloyne, J., C. Ferreira, M. Froemel, P. Surico (2018). Monetary Policy, Corporate Finance and Investment. NBER Working Paper 25366.

Cooley, T., R. Marimon, V. Quadrini (2004). Aggregate consequences of limited contract enforceability. Journal of Political Economy 112(4), 817–847.

Cornelli, G., J. Frost, G. Gambacorta, R. Rau, R. Wardop, T. Ziegler (2020). Fintech and big tech credit: a new database. BIS Working Papers, forthcoming.

Dávila, E. and A. Korinek (2017). Pecuniary externalities in economies with financial frictions. The Review of Economic Studies, 85(1), 352-395.

De Roure, C, L. Pelizzon, P. Tasca (2016). How does P2P lending fit into the consumer credit market?. Deutsche Bundesbank Discussion Papers 30.

Diamond, D. W., Y. Hu, R. G. Rajan (2017). Pledgeability, industry liquidity, and financing cycles. NBER Working Paper 23055.

Doerr, S. (2019). Unintended side effects: stress tests, entrepreneurship, and innovation. BIS Working Papers 823.

Favara, G., J. Imbs (2015). Credit Supply and the Price of Housing. The American Economic Review, 105(3), 958-992.

FSB (2019), Evaluation of the effects of financial regulatory reforms on small and medium-sized enterprise (SME) financing, Consultative document. June.

Frost, J., L. Gambacorta, Y. Huang, H. S. Shin, P. Zbinden (2020). Big tech and the changing structure of financial intermediation. Economic Policy, forthcoming.

Gale, D., M. Hellwig (1985). Incentive-compatible debt contracts: The one-period problem. Review of Economic Studies, 52(4), 647–663.

Gambacorta L., Y. Huang, H. Qiu, J. Wang (2019). How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm. BIS Working Papers 834.

Gan, J. (2007). Collateral, debt capacity, and corporate investment: Evidence from a natural experiment. Journal of Financial Economics, 85(3), 709-734.

Gertler, M., S. Gilchrist (1994). Monetary policy, business cycles, and the behavior of small manufacturing firms. The Quarterly Journal of Economics, 109(2), 309-340.

Glaeser, E. L., W. Huang, Y. Ma, A. Shleifer, (2017). A Real Estate Boom with Chinese Characteristics. Journal of Economic Perspectives, 31(1), 93-116.

Hau, H., Y. Huang, H. Shan, Z. Sheng (2018). Fintech credit, financial inclusion and entrepreneurial growth. Mimeo.

Hau, H., D. Ouyang (2018). Capital Scarcity and Industrial Decline: Evidence from 172 Real Estate Booms in China. Swiss Finance Institute Research Paper, 38.

Holmström, B., J. Tirole (1997). Financial intermediation, loanable funds and the real sector. Quarterly Journal of Economics, 112(3), 663-691.

Ioannidou, V., N. Pavanini, Y. Peng (2019). Collateral and Asymmetric Information in Lending Markets. CEPR Discussion Paper Series, 13905.

Jagtiani, J., C. Lemieux (2018a). The roles of alternative data and machine learning in fintech lending: evidence from the Lending Club consumer platform. Federal Reserve Bank of Philadelphia Working Papers, 18–15, April.

Jagtiani, J., C. Lemieux (2018b). Do fintech lenders penetrate areas that are underserved by traditional banks. Federal Reserve Bank of Philadelphia Working Papers 13.

Jeenas, P. (2018). Monetary policy shocks, financial structure, and firm activity: A panel approach. Mimeo, New York University

Jimenez, G., S. Ongena, J.L. Peydro, J. Saurina (2014). Hazardous Times for Monetary Policy: What Do Twenty‐Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk‐Taking?. Econometrica, 82(2), 463-505.

Kiyotaki, N., J. Moore (1997). Credit cycles. Journal of political economy, 105(2), 211-248.

Khwaja A., A. Mian (2008). Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market. The American Economic Review, 98(4), 1413-1442.

Lian, C., Y. Ma, (2019). Anatomy of corporate borrowing constraints. Mimeo.

Luohan Academy (2019). Digital Technology and Inclusive Growth. Hangzhou.

Lorenzoni, G. (2008). Inefficient credit booms. The Review of Economic Studies, 75(3), 809-833.

Mian, A., A. Sufi (2011). House prices, home equity-based borrowing, and the US household leverage crisis. American Economic Review, 101(5), 2132-56.

Mian, A., A. Sufi (2014). What explains the 2007–2009 drop in employment?. Econometrica, 82(6), 2197-2223.

OECD (2019). Financing SMEs and Entrepreneurs 2019: an OECD scoreboard. Paris.

Ottonello, P., T. Winberry (2018). Financial heterogeneity and the investment channel of monetary policy. NBER Working Paper 24221.

Schmalz, M. C., D.A. Sraer, D. Thesmar (2017). Housing collateral and entrepreneurship. The Journal of Finance, 72(1), 99-132.

Tang, H. (2019). Peer-to-peer lenders versus banks: substitutes or complements? Review of Financial Studies, 32(5), 1900-1938.

Waxman, A., Y. Liang, S. Li, P.J. Barwick, M. Zhao (2019). Tightening belts to buy a home: Consumption responses to rising housing prices in urban China. Journal of Urban Economics, 103190.

Annex A. Some facts about Ant Group

The *Alibaba Group* is one of the biggest tech companies in the world. It was publicly listed on the New York Stock Exchange in September 2014, and has a market capitalisation of USD 640 billion as of July 2020. *Alipay* is a third-party mobile and online payment platform, established by the Alibaba Group that was subsequently rebranded as *Ant Financial Services Group* in October 2014 and *Ant Group* in June 2020. Initially, Alipay provided financial service to online business on Alibaba Group's e-commerce platforms. Today, the business of Ant Group includes Alipay, Ant Fortune, MYbank, ZHIMA Credit and Ant Group Cloud, serving millions of small and micro-sized enterprises (SMEs), both online and offline, and retail customers. Our paper focuses on the credit to SMEs, so our data is obtained from Alipay and MYbank.

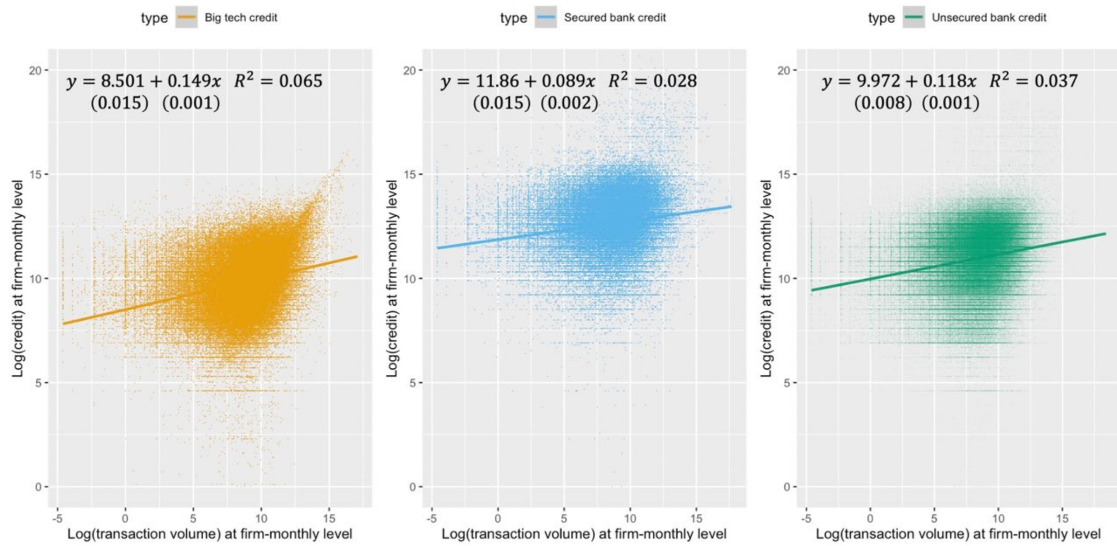
Operated by Ant Group, Alipay is a payment and lifestyle platform. Launched in 2004, Alipay currently serves over 1 billion users with its local e-wallets partners. Alipay is thus the world's largest mobile and online payments platform with a market share of over 50 percent in China. Ant Group has detailed information on enterprises and customers based on Alipay. MYbank is a private online bank established on June 25, 2015 by Ant Group with a mission to serve SMEs, to support the real economy and to practice inclusive finance. MYbank provides online, unsecured loan to SMEs based on a credit-scoring algorithm. The provision of credit is very fast and completely automated based on the so-called "310 model": 3 minutes to apply for credit, 1 second to approve and 0 people involved in the decision.

Alibaba Group owns three major trading e-commerce platforms, Alibaba (B2B), Tmall (B2C) and Taobao (C2C). Tmall and Taobao have the largest market shares in China at more than 50 percent. It is easier for firms fully integrated into the Alipay/Ant Group ecosystem to obtain financial services. This is for the following three reasons. First, the information on these firms is very rich. The big tech company can collect and process the data of these companies more comprehensively, such as those on business operations and scoring. Second, as discussed above, for firms in the ecosystem it is strategically more difficult to default, as big techs can use the receivables of these companies in their accounts to repay their debts. Third, given network effects and high switching costs, big techs could also enforce loan repayments by the simple threat of a downgrade or exclusion from their ecosystem if in default. Overall, the provision of credit to online borrower can be done with a careful credit scoring assessment and the credit was (at least initially) less risky that that provided to offline borrowers, operating out of the platform.

The use of QR code and offline vendors. In the second half of 2017, Ant Group promoted a campaign to offer to offline vendors a QR code technology for payments. Many small stores only needed to place a QR code sticker for their customers to scan and complete their payments. Larger stores also installed scanners of Ant Group to directly scan the QR code of the customer in Alipay. The QR code expanded the services of MYbank from the online firms to offline stores, which greatly expanded their business. As payments were done by means of Alipay, the data could be collected and used to analyse the evolution of the vendor's activity. Most of the offline stores are small (micro enterprises) and could therefore receive a credit score evaluation for the first time. The data were also used to calculate a network score to evaluate the position of the offline vendor in the big tech ecosystem and their connection with other vendors.

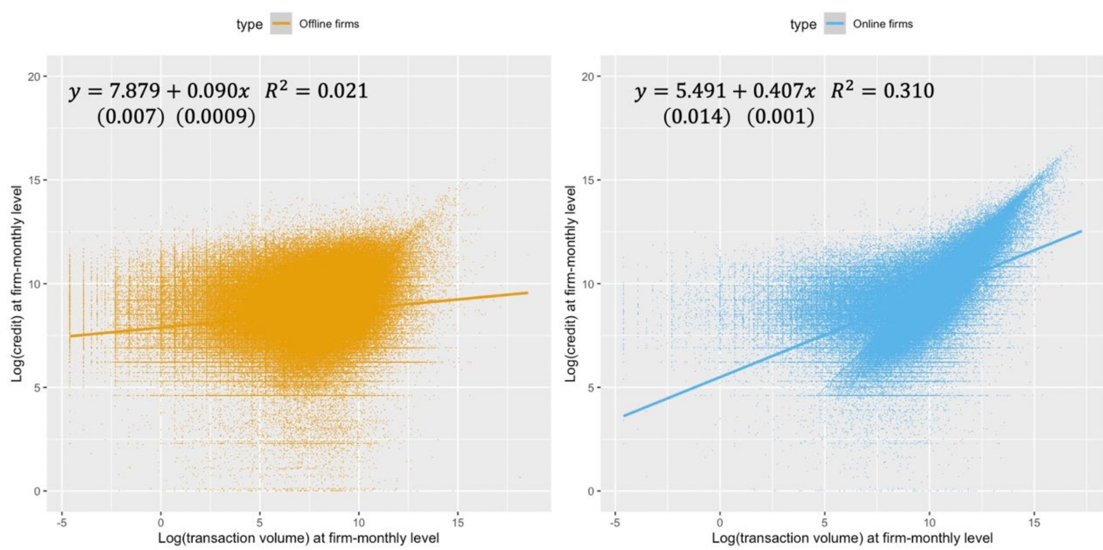
Ant Group Credit Scoring technique. The risk control model of MYbank is implemented through a credit scoring that use machine learning techniques and big data. The latter include transaction information, entrepreneur information, credit information and third-party information (client reviews and network score).

Figure 1. Elasticity between credit and transaction volumes



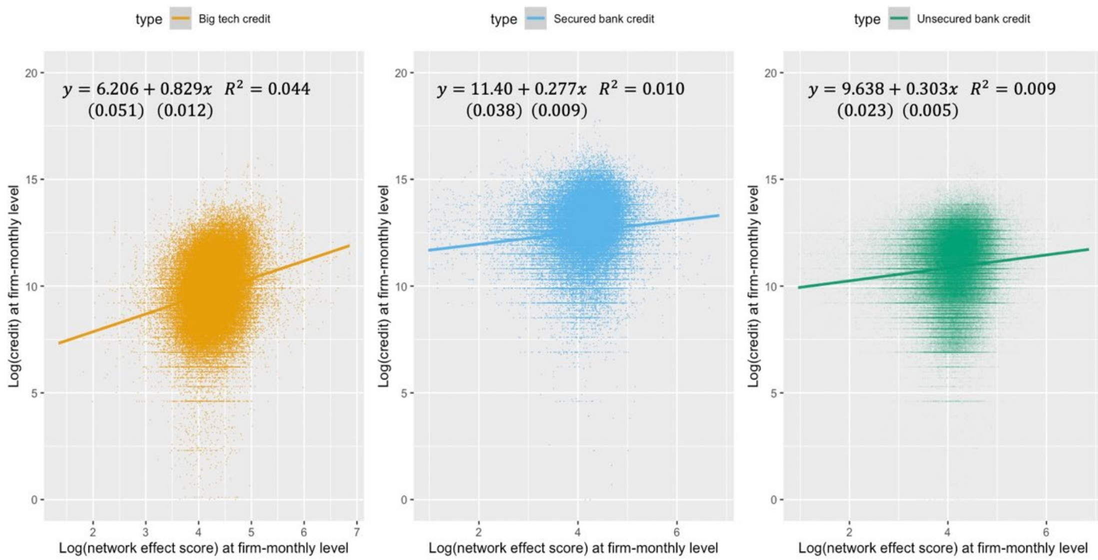
Note: Based on a 100,000 random sample of firms served by both MYbank and traditional banking. The dots in the figure indicates the log of credit use (y-axis) and the log of transaction volume (x-axis) at the firm-month level. The left-hand panel plots big tech credit, the middle panel plots secured credit and the right hand panel plots unsecured bank credit. Linear trend lines are reported in the graphs, together with 95% degree confidence bands. Standard errors in brackets.

Figure 2. Elasticity between big tech credit and transaction volume: offline firms vs online firms



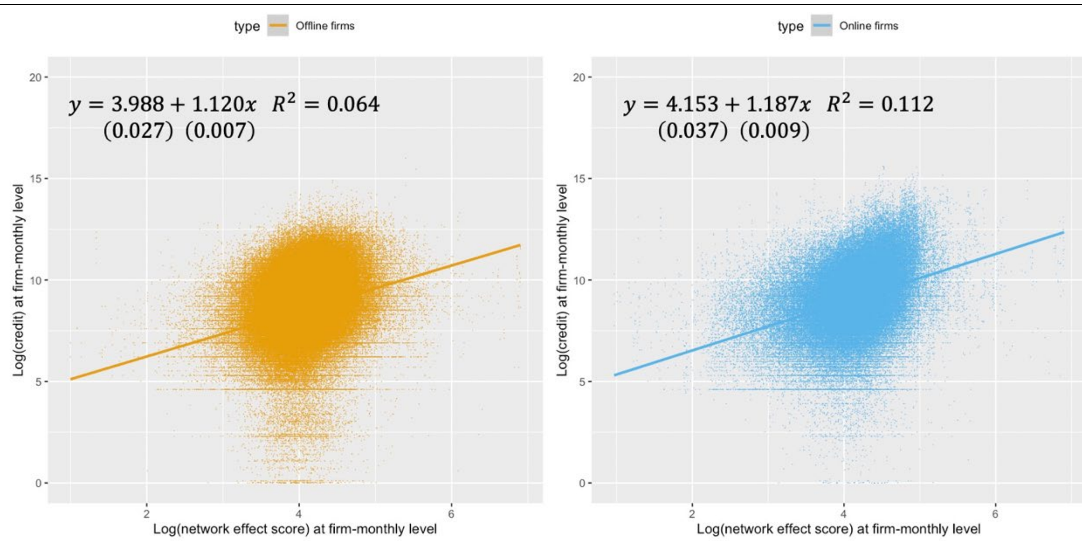
Note: Based on a 100,000 random sample of firms that received credit by MYbank. The dots in the figure indicates the log of credit use (y-axis) and the log of transaction volume (x-axis) at the firm-month level. The left-hand panel plots credit to offline firms and the right hand panel plots credit to online firms. Linear trend lines are reported in both graphs, together with 95% degree confidence bands. Standard errors in brackets.

Figure 3. Correlation between credit and the network score



Note: Based on a 100,000 random sample of firms served by both MYbank and traditional banks. The dots in the figure indicates the log of credit use (y-axis) and network score (x-axis) at the firm-month level. The left-hand panel plots big tech credit, the middle panel plots secured credit and the right hand panel plots unsecured bank credit. Linear trend lines are reported in the graphs, together with 95% degree confidence bands. Standard errors in brackets.

Figure 4. Elasticity between big tech credit and the network score: offline firms vs online firms

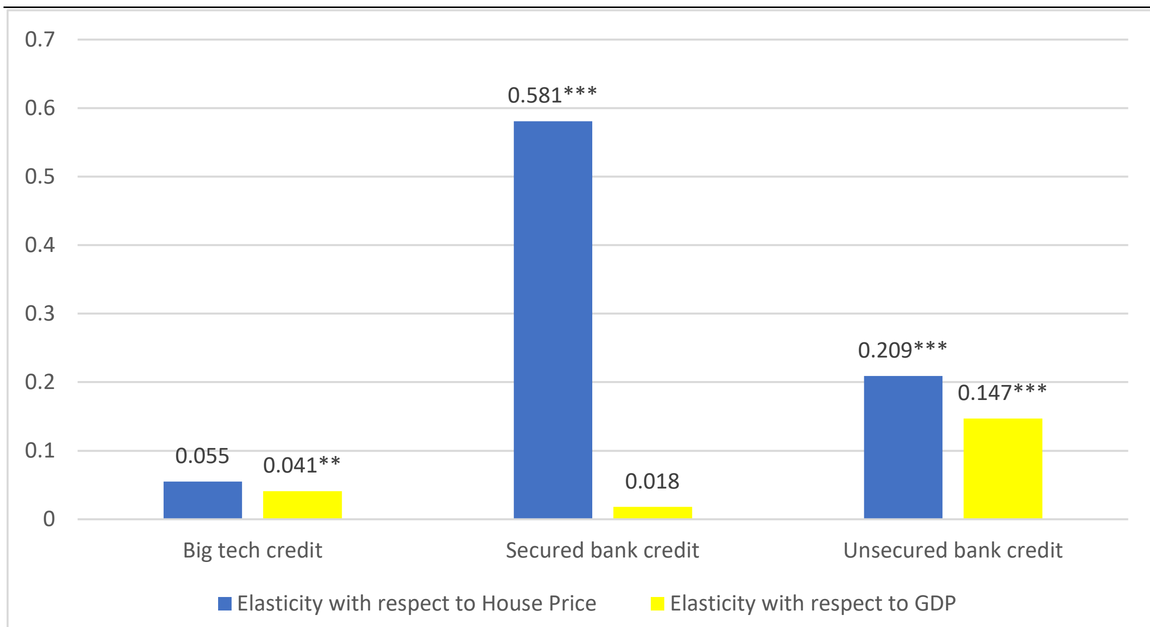


Note: Based on a 100,000 random sample of firms that received credit by MYbank. The dots in the figure indicates the log of credit use (y-axis) and the log of network effect score (x-axis) at the firm-month level. The left-hand panel plots credit to offline firms and the right hand panel plots credit to online firms. Linear trend lines are reported in both graphs, together with 95% degree confidence bands. Standard errors in brackets.

Figure 5. Elasticity of credit with respect to house prices

Note: Based on a 100,000 random sample of firms served by both MYbank and traditional banking. The dots in the figures indicate the average logarithm credit use (y-axis) and the average logarithm of housing price (x-axis) at the city-year level. Growth rates are approximated using first differences of log values. The left-hand panel plots big tech credit, the middle panel plots bank secured credit and the right hand panel plots bank unsecured credit. Linear trend lines are reported in each graph, together with 95% degree confidence bands. Standard errors in brackets.

Figure 6. Elasticity of credit with respect to house prices and GDP



Note: The figure reports the coefficient of three different regressions (one for each credit types) in which the log of credit is regressed with respect to the log of house prices at the city level, the log of GDP at the city level and a complete set of time dummies. Significance level: ** p<0.05; *** p<0.01.

Table 1. Summary statistics

A) Big tech credit	N	Mean	St. Dev.	P25	Median	P75
i) Firms' characteristics						
MYbank credit used (RMB)	7,096,381	17,998	26,280	1,950	6,900	20,000
Transaction volume monthly (RMB)	7,096,381	19,611	39,168	488	2,973	15,000
Network Score	7,096,381	62	22	45	58	75
Online	7,096,381	0.292	0.454	0	0	1
ii) Entrepreneurs' characteristics						
Age	6,299,630	32	8	27	31	36
Income level	6,299,630	2.010	0.818	1	2	3
iii) Economic and financial conditions						
House prices (RMB)	7,096,381	18,146	13,746	9,491	11,914	21,383
GDP (100 million RMB)	7,096,381	7,103	6,836	2,351	4,692	9,458
Land supply (%)	7,096,381	2.316	1.894	0.767	2.139	3.150
Mortgage rate (%)	7,096,381	5.431	0.380	5.260	5.600	5.720

B) Secured bank credit	N	Mean	St. Dev.	P25	Median	P75
i) Firms' characteristics						
Secured bank credit (RMB)	94,948	1,215,024	28,497,431	130,000	300,000	700,000
Transaction volume (RMB)	94,948	21,156	35,731	1,002	5,421	21,259
Network score	94,948	69	26	50	66	86
Online	94,948	0.126	0.332	0	0	0
ii) Entrepreneurs' characteristics						
Age	91,767	39	7	33	38	43
Income level	91,767	2.092	0.840	1	2	3
iii) Economic and financial conditions						
House price (RMB)	94,948	14,695	9,613	8,725	11,431	16,302
GDP (100 million RMB)	94,948	5,354	5,159	1,975	3,550	6,622
Land supply (%)	94,948	2.686	1.992	1.470	2.519	3.413
Mortgage rate (%)	94,948	5.319	0.420	5.010	5.420	5.680

C) Unsecured Bank Credit	N	Mean	St. Dev.	P25	Median	P75
i) Firms' characteristics						
Unsecured bank Credit (RMB)	398,789	346,788	13,647,023	20,000	60,000	150,000
Transaction volume (RMB)	398,789	16,485	27,559	830	4,312	16,975
Network score	398,789	67	26	48	63	83
Online	398,789	0.155	0.362	0	0	0
ii) Entrepreneurs' characteristics						
Age	365,344	36	7	31	36	41
Income level	365,344	2.068	0.839	1	2	3
iii) Economic and financial conditions						
House price (RMB)	398,789	17,707	12,558	9,832	13,193	20,990
GDP (RMB 100 million)	398,789	6,181	5,899	2,129	4,189	7,857
Land supply (%)	398,789	2.205	1.676	0.872	2.056	2.952
Mortgage rate (%)	398,789	5.208	0.447	4.690	5.260	5.680

Table 2. Main drivers of credit

Explanatory variables	Dependent variable: Log (MYbank credit) (I)	Dependent variable: Log (Secured bank credit) (II)	Dependent variable: Log (Unsecured bank credit) (III)
Log House Price (1)	0.052	0.371**	0.239*
	(0.047)	(0.151)	(0.138)
Log GDP (2)	-0.028	0.053	0.229*
	(0.034)	(0.101)	(0.121)
Log Transaction Volume	0.064***	0.036***	0.064***
	(0.001)	(0.002)	(0.001)
Log Network Score (3)	0.465***	0.096***	0.128***
	(0.004)	(0.011)	(0.008)
Age of the borrower	0.026***	0.009***	0.037***
	(0.0003)	(0.001)	(0.001)
Middle income (4)	0.569***	0.129***	0.075***
	(0.003)	(0.012)	(0.007)
High income (5)	1.281***	0.325***	0.312***
	(0.005)	(0.014)	(0.008)
Time FE	Yes	Yes	Yes
City FE	Yes	Yes	Yes
Number of observations	6,299,630	91,767	365,344
Adjusted R-squared	0.242	0.098	0.115

Notes: (1) At the city-month level. (2) At the city-quarter level. (3) Network score measures users' centrality in the network and is based on users' payment and funds information and social interactions. The user who has more connections gets a higher network score. (4) Middle income is a dummy variable that takes the value of one for the second third of the income distribution. (5) High income is a dummy variable that takes the value of one for the last third of the income distribution. Income is proxied by the total amount of funds into the Alipay wallet. Standard errors in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Table 3. Drivers of big tech credit

Explanatory variables	Dependent variable: Log (MYbank credit used)			Dependent variable: Log (MYbank credit line granted)		
	All (I)	Offline (II)	Online (III)	All (IV)	Offline (V)	Online (VI)
Log House Price (1)	0.002	0.046	0.019	-0.041	-0.089	0.154
	(0.038)	(0.047)	(0.042)	(0.065)	(0.066)	(0.096)
Log GDP (2)	0.046**	0.064***	-0.003	0.0004	-0.009	0.018
	(0.019)	(0.022)	(0.027)	(0.017)	(0.017)	(0.030)
Log Transaction Volume	0.046***	0.021***	0.160***	0.0005**	0.0003*	0.001**
	(0.001)	(0.0004))	(0.002)	(0.0002)	(0.0002)	(0.001)
Log Network Score (3)	0.567***	0.092***	0.818***	0.205***	0.193***	0.179***
	(0.012)	(0.009)	(0.010)	(0.010)	(0.010)	(0.024)
Time FE (month)	Yes	Yes	Yes	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	7,096,381	4,885,482	2,210,899	1,354,461	1,254,503	163,025
Adjusted R-squared	0.635	0.642	0.619	0.918	0.917	0.930

Notes: (1) At the city-month level. (2) At the city-quarter level. (3) Network score measures users' centrality in the network and is based on users' payment and funds information and social interactions. The user who has more connections gets a higher network score. Standard errors reported in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Table 4. Drivers of big tech credit: two specific products

Explanatory variables	Dependent variable: Log (credit used)		Dependent variable: Interest rate	
	Product 1	Product 2	Product 1	Product 2
Log House Price (1)	-0.005	-0.060	-0.001	-0.0004
	(0.104)	(0.113)	(0.001)	(0.002)
Log GDP (2)	0.022	0.051	0.00001	0.0002
	(0.054)	(0.074)	(0.0003)	(0.001)
Log Transaction Volume	0.012***	0.602***	-0.00001***	-0.001***
	(0.001)	(0.003)	(0.00001)	(0.00004)
Log Network Score (3)	0.091***	0.360***	-0.001**	-0.005***
	(0.030)	(0.021)	(0.0003)	(0.0003)
Time FE (month)	Yes	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes	Yes
Number of observations	743,187	186,190	743,187	186,190
Adjusted R-squared	0.413	0.713	0.915	0.794

Notes: (1) At the city-month level. (2) At the city-quarter level. (3) Network score measures users' centrality in the network and is based on users' payment and funds information and social interactions. The user who has more connections gets a higher network score. Standard errors reported in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Table 5. Drivers of bank credit

Explanatory variables	Dependent variable: Log (Secured bank credit)			Dependent variable: Log (Unsecured bank credit)		
	All	Offline	Online	All	Offline	Online
Log House Price (1)	0.591***	0.480***	1.129***	0.212***	0.249***	0.036
	(0.145)	(0.148)	(0.319)	(0.081)	(0.084)	(0.169)
Log GDP (2)	-0.002	-0.024	0.117	0.144**	0.118*	0.292**
	(0.107)	(0.112)	(0.286)	(0.057)	(0.062)	(0.119)
Log Transaction Volume	0.003	0.001	0.013**	0.004***	0.004***	0.004
	(0.002)	(0.003)	(0.006)	(0.001)	(0.001)	(0.003)
Log Network Score (3)	-0.039	-0.051	-0.001	0.016	0.022	-0.029
	(0.029)	(0.032)	(0.082)	(0.012)	(0.014)	(0.031)
Time FE (month)	Yes	Yes	Yes	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	94,948	81,799	13,149	398,789	369,287	67,103
Adjusted R-squared	0.579	0.578	0.583	0.651	0.651	0.650

Notes: (1) At the city-month level. (2) At the city-quarter level. (3) Network score measures users' centrality in the network and is based on users' payment and funds information and social interactions. The user who has more connections gets a higher network score. Standard errors reported in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Table 6. First stage regression: The impact of local housing supply and interest rate conditions on housing prices

Explanatory variables	Dependent variable: Log House Price
Lagged Land Supply (1)	-0.0370***
	(0.0131)
Lagged Land Supply (1) * mortgage rate (2)	0.00667**
	(0.00269)
Time FE	Yes
City FE	Yes
Number of observations	2,688
Adjusted R-squared	0.9952

Notes: (1) Lagged land supply are calculated using annual land supply scaled by urban construction land lagged by 12 months. (2) Nationwide interest rate at which banks refinance their home loans at the quarterly level. Standard errors in brackets are clustered at the city level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Table 7. Results using instrumented values of the housing prices

Explanatory variables	Dependent variable: Log (MYbank credit used)	Dependent variable: Log (Secured bank credit)	Dependent variable: Log (Unsecured bank credit)
Log House Price IV (1)	0.074	0.605**	0.393*
	(0.074)	(0.292)	(0.211)
Log GDP (2)	0.043**	0.008	0.144***
	(0.017)	(0.087)	(0.053)
Log Transaction Volume	0.039***	0.004**	0.004***
	(0.001)	(0.002)	(0.001)
Log Network Score (3)	0.462***	-0.037	0.015
	(0.011)	(0.024)	(0.014)
Time FE	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes
Number of observations	7,096,381	94,948	398,789
Adjusted R-squared	0.635	0.583	0.641

Notes: (1) Log House Prices are instrumented using the model described in Table 6. (2) At the city-quarterly level. (3) Network score measures users' centrality in the network and is based on users' payment and funds information and social interactions. The user who has more connections gets a higher network score. (4) Dummy variable that takes the value of one for bank unsecured credit and 0 for big tech credit. Standard errors in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Table 8. Big tech credit vs bank secured credit

Explanatory variables	Dependent variable: Log (credit)		
	All	Offline	Online
Log House Price (1)	-0.061	-0.297	0.092
	(0.149)	(0.185)	(0.230)
Log GDP (2)	0.074	0.007	0.178
	(0.092)	(0.110)	(0.173)
Log Transaction Volume	0.037***	0.015***	0.145***
	(0.002)	(0.002)	(0.007)
Log Network Score (3)	0.456***	0.079*	0.830***
	(0.040)	(0.048)	(0.063)
Log House Price * Bank secured (4)	0.633***	0.689***	1.106***
	(0.217)	(0.238)	(0.395)
Log GDP* Bank secured (4)	-0.090	-0.024	-0.217
	(0.157)	(0.178)	(0.336)
Log Transaction Volume* Bank secured (4)	-0.035***	-0.016***	-0.128***
	(0.003)	(0.004)	(0.009)
Log Network Score * Bank secured (4)	-0.505***	-0.148**	-0.840***
	(0.052)	(0.061)	(0.099)
Time*credit type FE	Yes	Yes	Yes
Borrower*credit type FE	Yes	Yes	Yes
Number of observations	168,518	128,163	40,355
Adjusted R-squared	0.737	0.741	0.732

Notes: (1) At the city-month level. (2) At the city-quarter level. (3) Network score measures users' centrality in the network and is based on users' payment and funds information and social interactions. The user who has more connections gets a higher network score. (4) Dummy variable that takes the value of one for bank secured credit and 0 for big tech credit. Standard errors in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Table 9. Big tech credit vs bank unsecured credit

Explanatory variables	Dependent variable: Log (credit)		
	All	Offline	Online
Log House Price (1)	-0.038	-0.101	0.108
	(0.071)	(0.076)	(0.120)
Log GDP (2)	0.114**	0.074	0.175*
	(0.050)	(0.051)	(0.096)
Log Transaction Volume	0.037***	0.014***	0.113***
	(0.001)	(0.001)	(0.003)
Log Network Score (3)	0.389***	0.059***	0.776***
	(0.022)	(0.022)	(0.034)
Log House Price * Bank unsecured (4)	0.188*	0.224*	-0.098
	(0.111)	(0.102)	(0.219)
Log GDP* Bank unsecured (4)	0.043	0.028	0.136
	(0.075)	(0.075)	(0.156)
Log Transaction Volume* Bank unsecured (4)	-0.034***	-0.011***	-0.110***
	(0.002)	(0.002)	(0.004)
Log Network Score * Bank unsecured (4)	-0.371***	-0.047*	-0.782***
	(0.027)	(0.027)	(0.049)
Time*credit type FE	Yes	Yes	Yes
Borrower*credit type FE	Yes	Yes	Yes
Number of observations	699,755	525,118	174,657
Adjusted R-squared	0.679	0.689	0.651

Notes: (1) At the city-month level. (2) At the city-quarter level. (3) Network score measures users' centrality in the network and is based on users' payment and funds information and social interactions. The user who has more connections gets a higher network score. (4) Dummy variable that takes the value of one for bank unsecured credit and 0 for big tech credit. Standard errors in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Table 10. Big tech credit vs bank credit controlling for demand shifts

Explanatory variables	Dependent variable: Log (credit)			
	Big tech credit vs Bank secured credit	Big tech credit vs Bank unsecured credit	Big tech credit vs Bank secured credit	Big tech credit vs Bank unsecured credit
Log transaction volume	0.041***	0.037***		
	(0.002)	(0.001)		
Log network score (3)	0.447***	0.396***		
	(0.040)	(0.022)		
Log house prices (1) * Bank credit (4)	0.583***	0.161	1.551**	0.892***
	(0.217)	(0.119)	(0.602)	(0.292)
Log GDP (2) * Bank credit (4)	−0.100	0.064	0.102	0.017
	(0.167)	(0.080)	(0.375)	(0.203)
Log transaction volume* Bank credit (4)	-0.039***	-0.034***	-0.057***	-0.008**
	(0.004)	(0.002)	(0.006)	(0.003)
Log network score (3) * Bank credit (4)	-0.484***	-0.369**	-0.443***	-0.210***
	(0.052)	(0.027)	(0.043)	(0.022)
Time*City FE	Yes	Yes	No	No
Time*Borrower FE	No	No	Yes	Yes
Borrower*credit type FE	Yes	Yes	No	No
City*credit type FE	No	No	Yes	Yes
Time*credit type FE	Yes	Yes	Yes	Yes
Number of observations	168,518	699,775	48,574	221,082
Adjusted R-squared	0.737	0.680	0.577	0.488

Notes: (1) At the city-month level. (2) At the city-quarter level. (3) Network score measures users' centrality in the network and is based on users' payment and funds information and social interactions. The user who has more connections gets a higher network score. (4) Dummy variable that takes the value of one for bank secured credit (columns 1 and 3) or for bank unsecured cred in the first and third columns the dummy variable Bank credit takes the value of 1 for bank secured credit and 0 for big tech credit. Vice versa, in the second and fourth column the dummy variable Bank credit takes the value of 1 for bank unsecured credit and 0 for big tech credit. Standard errors in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.