What Do A Billion Observations Say About Distance and Relationship Lending?*

Haoyu Gao
Central University of Finance and Economics
E-mail: gaohaoyu@cufe.edu.cn

Hong Ru
Nanyang Technological University
E-mail: ruhong@ntu.edu.sg

Xiaoguang Yang
Chinese Academy of Sciences
Email: xgyang@iss.ac.cn

October 1, 2018

* Haoyu Gao, Central University of Finance and Economics, 39 South College Road, Haidian Dist., Beijing 100081, China; gaohaoyu@cufe.edu.cn. Hong Ru, Nanyang Technological University, 50 Nanyang Avenue, Singapore, 639798; ruhong@ntu.edu.sg. Xiaoguang Yang, Chinese Academy of Sciences, Beijing, China; xgyang@iss.ac.cn.

We thank Jules van Binsbergen, Kaiji Chen, Tarun Chordia, Zhi Da, Ruediger Fahlenbrach, Fangjian Fu, Shiyang Huang, Xing Huang, Yi Huang, Jun-koo Kang, Oğuzhan Karakas, Weikai Li, Dong Lou, Xiaorong Ma, Lin Peng, Yaxuan Qi, Wenlan Qian, Shikhar Singla, Rong Wang, Shang-jin Wei, and Liyan Yang. This paper benefited hugely from conference discussants and participants at NTU Finance Conference 2018, TFA 2018 Conference, China's Economy and Financial Markets 2018, International Workshop on Financial Stability in a New Era, and seminar participants at SMU Brown Bag. We thank the financial support from the Nanyang Technological University, the research grant of National Science Foundation of China (No. 71532013, 71431008, and 71702207), and the in-kind support to the data access from China Banking Regulatory Commission. The views are our own, and we are solely responsible for any errors.

What Do A Billion Observations Say About Distance and Relationship Lending?

Abstract

Using billions of observations on the locations of bank branches and firms in China, we measure lender-borrower distance by geographic information system (GIS) and find a non-trivial amount of distant lending. Distant borrowers are more likely to be connected to banks' local borrowers. We use novel data of monthly internal loan rating changes to directly measure soft information by tracing whether banks downgrade ratings before delinquency. For connected borrowers, banks have better soft information and predict delinquent events more accurately. This effect is more pronounced for distant borrowers. Consequently, connected borrowers' delinquent rate is lower. Our findings show that the inter-firm network facilitates banks to collect soft information and manage risks, especially for distant borrowers.

Keywords: Big Data; Distance; Firm Network; Soft Information

1. Introduction

Banks rely heavily on relationship lending and invest substantially in soft information collection from borrowers to manage default risks. A major challenge in the literature is how to measure the degrees of soft information. Due mainly to data limitation, researchers can't observe borrower soft information directly and use various proxies to extrapolate it.[2] For example, borrower proximity could reduce banks' costs of loan screening and monitoring which facilitates the collection of soft information (e.g., Degryse and Ongena (2005); Dass and Massa (2010)). In contrast, researchers have shown that the geographic proximity becomes less relevant in relationship lending (e.g., Petersen and Rajan (2002)). It is a long-standing puzzle on whether borrower proximity and other commonly used relationship measurements could fully capture soft information. This is a major obstacle to study the economic consequences of relationship lending.

In this paper, we discover a substantial amount of distant lending and find that the distant borrowers are more likely to be connected to banks' local borrowers via an inter-firm network. The novel data on internal loan rating changes allow us to directly observe banks' soft information on borrowers, i.e., downgrading the internal loan ratings before the delinquency. We find that banks can predict the delinquent events more accurately for connected borrowers, especially for distant loans. This suggests that soft information could be passed via the firm network to facilitate distant lending. The inter-firm network serves as a novel channel of banks' soft information collection for distant borrowers.

More specifically, we use the big data of seven million loans from the China Banking Regulatory Commission (CBRC) and calculate the physical distance of billions of lender-borrower pairs by using geographic information system (GIS). The CBRC data record the detailed loan-level information (e.g., loan contract terms, borrower and lender information such as ID and location, and delinquency) for seventeen largest commercial banks and two policy banks in China between 2006 and 2013. The data cover over seven

[2] Many studies, such as Bolton et al. (2016) and Hombert and Matray (2016), argue that there is no consensus in the literature on how to identify soft information and the correlations among different relationship measures are low. Karolyi (2017) find that the average correlation coefficients of some well-used relationship proxies are 0.38.

million individual corporate loans which include all borrowers with an annual credit line over US$8 million. The data count for approximately 80% of the total bank loan market in China. We also obtain the population data of all bank branches in China (i.e., over 200 thousand branches of around 2,800 banking financial institutions) between 1949 and 2016. For each branch, the data record the branch name, ID, hierarchy, full addresses, and the exact opening and closing dates. We employ the GIS technology to obtain the geographic coordinates of firms and bank branches and calculate their physical distances.

Our first analysis concerns the borrower proximity and lending intensities. In particular, for each pair of bank-firm, we calculate the share of loan amounts and the share of loan numbers that have been borrowed from the bank, whether the firm borrowed any new loans from the bank in past 12 months, and the number of banks with outstanding loans. These four variables of lending intensities have been used widely in the literature to measure the lender-borrower relationships. [3] We find that the correlations are low among these measurements, especially between distance and others. This suggests that these proxies are either noisy or capture different aspects of borrowers' soft information. Moreover, we perform the OLS regressions of these lending intensity variables on the distance and its quadratic term. We find a robust *U*-shaped pattern; the linear term of distance is negatively associated with lending intensities while the quadratic term of distance is positively associated with lending intensities. Based on the estimated coefficients, banks decrease their lending intensities with distance within a certain range (around 250 km) but start to increase the lending intensities with distance when the firm is beyond this range. There is a substantial amount of distant lending, e.g., 11% of the loans go to distant borrowers.

Next, we explore the underlying channels behind the significant amount of distant lending. We find that the distant loans are mainly from the bank branches located outside of borrowers' cities (i.e., outside city loans). For example, the average lender-borrower distance is 277 km for outside city loans while it is 42 km for inside city loans. [4]

---

[3] See, for example, Schenone (2010), Bharath et al. (2011), Ioannidou and Ongena (2010), Bae, Kang and Lim (2002), and Jiménez, Salas and Saurina (2006).
[4] Some prior studies in US use 250-mile radius to define the local area (e.g., Ivković and Weisbenner (2005); Seasholes and Zhu (2005, 2010); Chhaochharia, Kumar and Niessen-Ruenzi (2012); Brown, Stice and White (2015)).

Furthermore, we construct the inter-firm network from the CBRC loan-level data that record each borrower's connected firms such as sibling firms, related transaction parties, or shareholders. We then define a borrower as connected to the lending bank if any of its connected firms are local borrowers of the same bank (i.e., borrowing inside city loans). We find that a 100% increase in distance is associated with a 2.4% increase in the borrower's likelihood of being connected to the bank via the firm network. This suggests that the inter-firm network could facilitate distant lending.

To further test whether soft information could be passed through via the firm network, we use the novel data of internal loan rating changes to directly measure soft information. Specifically, for each loan, the CBRC records the internal ratings at the monthly frequency.[5] This allows us to observe whether the banks foresee the delinquency and downgrade ratings before the actual late payments (we call it "early prediction" in the paper). Condition on delinquency, we first perform the OLS regressions of whether the banks downgrade the internal ratings before the actual late payments on whether the delinquent borrower is connected or not. We find that, for distant borrowers, one more connection to the borrower is associated with a 13.6% increase in the banks' ability to early prediction of delinquent events. Moreover, we also study how many months in advance that the banks can predict delinquent events. On average, banks can predict the delinquent event 33.6% earlier for one unit increase in borrowers' number of connections in long distance. For borrowers close-by, we don't find significant effects of the firm network on the early prediction. One caveat is that besides soft information, banks could make the downgrade decisions based on public information of the borrowers. We control for the firm×year fixed effects to eliminate all the firm-specific time trends (e.g., firm-specific public information). The variation we exploit is within the same firm-year and cross banks which captures soft information of particular lenders.

To establish the causal effects of borrower proximity on different channels of soft information collection, we use the 2009 bank entry deregulation in China as the

[5] China introduces the international five-category loan classification for risk management, in which there are five levels for standard rating based on internal rating technology. 1 is the highest rating for the "normal" loans, 2 is for the "special mentioned", 3 is for the "substandard", 4 is for the "doubtful" and 5 is for the "loss". In this paper, rating is a dummy variable for whether the rating is 1 or others (i.e., ratings from 2 to 5). A loan is classified as a "non-performing loan" if it is classified as "sub-standard", "doubtful" or "loss".

instrument for the physical distance between banks and firms. The banking sector in China is heavily regulated. For example, the CBRC issued a bank entry regulation in 2006 to restrict the branch opening. This restriction was partially lifted in 2009 deregulation for joint equity banks. In particular, after April 2009, joint equity banks are allowed to open branches freely in a city where they have already had branches in this city or the provincial capital of this city.[6]

In the first stage regressions, the bank-firm distances decrease significantly after the shock in the deregulated cities. On average, due mainly to the increased number of new branches, the bank-firm distances decrease significantly by 1.4% in deregulated cities after April 2009. In the second stage, consistent with OLS results, the coefficients of *LogDistance* are significantly negative while the interactions of *LogDistance* and inter-firm network connection variables have significantly positive coefficients for the early prediction dummy. On average, for connected borrowers, a 100% increase in distance leads to a 9.45 % increase in the likelihood of the rating downgrades before delinquency. Moreover, the number of months of early prediction also increases for connected distant borrowers. Specifically, a 100% increase in distance leads to 13.2%, 18.8%, and 15.5% increases in early prediction length for borrowers connected via their sibling firms, firms with related transactions, and shareholders respectively.

In sum, our findings show that distant borrowers are more likely to be connected to the local borrowers of the bank. On the one hand, consistent with conventional wisdom, borrower proximity could reduce the monitoring and screening costs of obtaining soft information. The firm network doesn't seem to play a role in passing soft information when borrowers are close-by since the bank can cheaply obtain borrowers' soft information by screening and monitoring. On the other hand, the firm network could overcome the increased costs of soft information collection which helps banks discover good quality borrowers in long distance. Banks use different methods to collect soft information in the long vs. short distance. Specifically, the firm network is an essential

[6] Gao et al. (2018) use the same shock to establish the causal effects of increased bank competition. In China, there are three types of banks; the big five state own commercial banks, twelve joint equity commercial banks, and municipal commercial banks. See detailed discussion in Section 4.

channel of soft information collection for distant lending which improves the predictions of loan delinquency and facilitates the risk management.

In a next step, we explore firm networks' heterogenous effects on soft information collection across various borrowers. Relationship lending is mainly for the small and medium enterprises (SME) while the large firms borrow transaction loans primarily. We refer to the official classification of SME in China to stratify our sample into SMEs and large firms.[7] We find that for SMEs, the firm network could significantly improve the early prediction of delinquent events of distant lending. In contrast, this effect is muted for large firms as predicted by the literature on relationship lending. Moreover, 21% of the loans in our sample have third-party guarantees. The guarantors are legally obliged to make loan payments in default events. We find that the firm network facilitates soft information collection only for loans without guarantees which suggests that banks have lower incentives to collect soft information when the loan payments are guaranteed. The findings of SMEs and loan guarantees further strengthen the firm network channel of soft information collection which is novel in the literature.

Lastly, we explore the economic consequences of the firm network on distant lending regarding default risks. We regress the loan defaults (i.e., 90-days delinquency) on whether the borrowers are connected or not. We find that the one more connection for borrowers in the firm network is associated with a 10% decrease in default rate. Specifically, when the borrowers' sibling firms and related transaction partners also borrower from the same bank, the default rates are 20% and 10% lower. This suggests that the banks can indeed use the soft information collected from the firm network to choose the good quality borrowers for the risk management.

This paper contributes to the literature in two folds. First, despite the large body of empirical literature on distance and relationship lending, it is still inconclusive on the consequences of borrowers' geographical proximity. On the one hand, the conventional wisdom argues that borrower proximity lowers screening and monitoring costs which facilitates the collection of soft information and leads to lower borrowing costs and risks

[7] We classify the SMEs based on the "Standards for Classification of Small and Medium-sized Enterprises" issued by the State Council of China.

(e.g., Calomiris (2000); Degryse and Ongena (2005); DeYoung, Glennon, and Nigro (2008); Agarwal and Hauswald (2010); Hollander and Verriest (2016); Bolton et al. (2016); Hombert and Matray (2016)).[8] On the other hand, distant lending is prevalent in many countries since the geographic proximity becomes less relevant in recent years due to various reasons such as technology advancements, developments of transportation system, and religions (e.g., Petersen and Rajan (2002); Giroud (2013); Beck, Ongena, and Şendeniz-Yüncü (2018); Da et al. (2018)). Moreover, the studies on borrower proximity and relationship lending explore only the monotonic patterns between distance and lending activities. This paper, for the first time, show the non-monotonic pattern of distance and lending activities and the different underlying mechanisms of extracting soft information in short vs. long distance. Especially, the firm network serves as a novel channel to overcome the physical distance and facilitate the information flow from borrowers to lenders in distant lending. This also sheds light on the non-trivial distant lending in many countries worldwide and on the mixed evidence on how geographic distance affect the relationship lending. [9]

Our second contribution is to measure soft information directly by using novel data on internal loan rating changes. The key of relationship lending leans on banks' costly monitoring and screening in exchange for soft information, especially for default risks (e.g., Allen (1990); Diamond (1984); Winton (1995); Boot and Thakor (2000)). Due mainly to data limitation, researchers can't directly observe lenders' soft information. Most previous studies extrapolate the degrees of soft information from indirect measurements (e.g., distance, lending frequency, and lending share) which do not correlate strongly with each other. This suggests that these measurements are either noisy or captures different parts of soft information (Bolton et al. (2016); Hombert and Matray

[8] Besides the studies on proximity and bank lending, people also find the important role of proximity in various economic activities (e.g.,; Coval and Moskowitz (1999, 2001); Butler (2008); Bae, Stulz, and Tan (2008);).

[9] The literature has documented non-trivial amount of distant lending. See Table A2 in Appendix for the details. It is also a long debate in the literature on how lender-borrower relationship affect the lending consequences. On the one side, firms can have better access to finance when they have relationships with banks, which mitigate the issues of asymmetric information (e.g., Peterson and Rajan (1995); Berger and Udell (1995); Bharath et al. (2011); Carbo-Valverde, Rodriguez-Fernandez, and Udell (2009); Ayyagari, Demirguc-Kunt, and Maksimovic (2010)). On the other side, firms could be hold up by the relationship lenders and suffer from the price discrimination (e.g., Santos and Winton (2008); Ioannidou and Ongena (2010)).

(2016)). We can observe directly whether banks have soft information to better predict the delinquency by tracing out the dynamics of delinquency and rating changes at the loan level.[10] This is a big step forward in the literature. Giannetti, Liberti, and Sturgess (2017) find that rating changes can be strategic since they would be shared in public. Our internal rating data from the CBRC are not shared among banks which could mitigate the concern of rating manipulation. Moreover, the firm×year fixed effects absorb the public information of individual borrowers which helps tease out the variation of soft information.

The rest of this paper is organized as follows. In Section 2, we describe the institutional background of the banking system in China. In Section 3 we present our data and summary statistics. Section 4 provides the empirical strategies and results. Section 5 concludes.

2. Background

2.1. Banking System in China

There are mainly three types of banks in China: big five national commercial banks, twelve joint equity banks, and municipal commercial banks. The first tier is the big five commercial banks that are state-owned national banks. In 1983, as part of Den Xiaoping's economic reform, big four commercial banks were established to take charges of commercial businesses in China. In particular, the Industrial and Commercial Bank of China (ICBC) focused on the corporate lending business, the Agriculture Bank of China (ABC) focused on the economic development in the rural areas, the Bank of China (BOC) was specialized in the foreign exchange business, and the China Construction Bank (CCB) was responsible for construction and infrastructure developments.

Besides these big four wholly state-owned banks, joint equity banks started to be formed in the late 1980s. The Bank of Communications (BoCom) is the first joint equity banks in China and was established in 1987. Although BoCom is technically a joint equity bank, it is the same as the big four regarding the regulation and political hierarchy. For example, big four banks and BoCom are under the direct control of the central

[10] Nakamura and Roszbach (2016) use the rating change data in Sweden and find that internal credit ratings indeed include valuable private information from monitoring.

government and are held by the Ministry of Finance and China Investment Corporation.[11] People address them as the big five. The second tier is the twelve joint equity banks that are also mainly state-owned while their number of branches are far less than that of big five banks and their businesses (e.g., corporate lending) focus on the local markets. By the end of the year 2013, as reported by CBRC's annual reports, the big five commercial banks dominate the market and control for approximately 43.3% of the market share. On the other hand, joint equity banks are much smaller and control for about 17.8% of the market share. The rest of the financial institutions belong to the third tier such as municipal commercial banks.

2.2. Bank Branches Development

Compared to the big five banks, the developments of the twelve joint equity banks have been suppressed due to the historical reasons and regulations. On the one hand, the big five banks were established earlier and have higher political rankings than joint equity banks. For example, the big five have the priority to open branches across the country. In 2002, the Chinese government announced the plan for the shareholding system reform of big four state-owned commercial banks to improve the efficiency in the banking system for the economic development.[12] Along with this reform, the big four were opening branches to expand their businesses dramatically all over the country. In 2006, big five bank branches had already covered approximately 90% of the cities in China while the 12 joint equity banks had covered only approximately 7% of the cities.

On the other hand, the joint equity banks' branch openings have been strictly regulated. For example, in 2006, the CBRC announced that all the local commercial banks including twelve joint equity banks could only apply to open one branch in each city.[13] Specifically, joint equity banks were not allowed to apply multiple branches in one application. Besides this restriction on numbers, joint equity banks need to apply to CBRC's local office first for the initial approval. After that, they need the approval from

[11] China Investment Corporation is a sovereign wealth fund which manages the foreign exchange reserves of China.
[12] For example, the BOC and CCB were listed in stock market in 2004 and the ICBC was listed in 2005. The ABC is the last one of big four going public and was listed in 2009.
[13] Please refer to CBRC Order [2006] No.2, titled "The implementation of administrative licensing items on Chinese commercial banks"

CBRC's central office. This dual approval process takes more than a year to finish. The number of branches for joint equity banks increased from 3,351 in 2006 to 4,700 in 2009. In contrast, in the end of 2009, the big five have 52,707 branches in total.

3. Data and Summary Statistics

Our paper includes two proprietary datasets for our empirical analyses: CBRC bank branch data and CBRC corporate loan data.

3.1. CBRC Branch Data and Loan Level Data

Our primary dataset is from the CBRC that covers all bank branches in China. In particular, this population dataset records over 200 thousand branches of around 2,800 banking financial institutions between 1949 and 2016. For each branch, we observe the branch name, ID, hierarchy, full addresses, and the exact opening and closing dates.[14] Since we have the history of all bank branches in China, we can observe the full dynamic of individual branch opening and closing across the entire country.

The second dataset is also from the CBRC which records the loan level information from all major banks in China. In particular, this loan-level data cover over 7 million loan contracts granted by 19 largest Chinese banks. The borrowers in our sample have the unique 9-digit organization codes. All borrowers with an annual credit line over RMB 50 million (approximately US$8 million) are included in our sample between October 2006 and June 2013. The data have an excellent representation of China's debt market which accounts for over 80% of the total bank credit in China. In total, there are over 160,000 borrowing firms located in all 31 provinces in China across all 20 different sectors by the Economic Industrial Classification Code in China. Besides the comprehensive coverage, the data also contain detailed loan-level information, i.e., the unique firm identifier, firm-level fundamentals (e.g., size, leverage and location), banks' information (e.g., the names and location of branches), and loan-level characteristics (e.g., loan amount, loan maturity, credit guarantee providers, internal ratings, issuing date, maturity date on contracts, and

[14] Gao et al. (2018) use the same data of bank branches and cross check the CBRC data with the numbers of branches in banks' annual reports. Over 96% of the branches in CBRC dataset could be matched with the branches listed in banks' website. The quality of CBRC bank branch dataset is very good.

loan delinquency status).[15] Specifically, for each loan application, the borrower needs to disclose its most related firms (e.g., shareholders, sibling firms with the common shareholder, and firms with other relationships such as related party transactions). Thus, we build the inter-firm network based on these data. The CBRC also records the unique ID of these connected firms of the borrowers so that we can trace the borrowing activities of these connected firms as well. The loan information is mandatorily updated at month frequency during its whole life cycle. In this way, we can trace the dynamics of internal loan ratings and examine whether the banks early downgrade their borrowers before the actual late payments.

3.2. GIS and Lender-Borrower Distance

Based on these two datasets, for each loan, we can identify the city where the loan was issued and have the address information of all the branches of the bank in this city. This, thus, allows us to locate the lending bank branches accurately. For borrowers, the CBRC loan-level data record the 9-digit firm ID (i.e., the unique identifier for corporations by the General Administration of Quality Supervision, Inspection and Quarantine of China) and the exact addresses of the firm. For the firms without address information, we obtain it from the National Company Credit Information System launched by China's State Administration of Industry and Commerce by using the firm ID.[16] We also manually check the location data which cover approximately 99% of the firms in the cleaned version.

Next, we use the GIS technology to locate the exact longitude and latitude of each borrower and bank branch on the map. In particular, we input the address information to the application programming interface (API) of Autonavi (Amap), the leading company of mapping service provider in China. The outputs from the API are the geographic coordinates of individual firms and bank branches (i.e., four-digit latitudes and longitudes). We restrict the sample to 17 commercial banks (i.e., big five banks and twelve joint equity banks) and their borrowers.

[15] However, the data do not record loan interest rates. In China, the lending rate was fully liberalized after July 20, 2013. During our sample period, the bank lending rates were still highly regulated.
[16] More information can be accessed via http://www.gsxt.gov.cn/index.html.

We then combine the data of firm locations and bank branch locations to calculate the physical distance between the lenders and borrowers. In particular, we follow the formula proposed by Coval and Moskowitz (1999) to calculate the distance between a firm and the bank branch at the monthly frequency. The formula is: $\text{Distance} = \text{r} \times \arccos[\sin(firmlat) \times \sin(banklat) + \cos(firmlat) \times \cos(banklat) \times \cos(banklon - firmlon)]$, where *firmlat* and *banklat* are the latitudes of the firm and the bank branch locations, respectively. The *firmlon* and *banklon* are the longitudes of the firm and the bank branch locations, respectively. r is the radius of the earth in kilometers (i.e., approximately 6,378 kilometers). Moreover, the latitude and longitude numbers are converted into radians by means of a division by $180/\pi$. This methodology based on spherical law estimations has been used in many prior studies, such as Baik, Kang and Kim (2010), Chhaochharia, Dass and Massa (2011), Kumar and Niessen-Ruenzi (2012).

For each loan, we calculate the distance between the borrower and all branches in the city where the loan is issued. We then follow the prior literature to choose the shortest one (i.e., the distance between the borrower and the closest bank branch in a given city) as the lender-borrower distance. Moreover, we also trace the dynamics of branch opening over time and update the lender-borrower distance at the monthly frequency. In total, there are over five billion observations of lender-borrower distance at the loan-branch-month level.

3.3. Summary Statistics

Table 1 presents summary statistics of the data. Panel A reports the variables for lender-borrower relationships (i.e., soft information) at the firm-bank-month level, including the distance and other four commonly used measures, i.e., *Bank_Num, LoanAmount_Share, LoanFreq_Share,* and *Recent_Lending*, respectively. We construct these conventional relationship measurements following the prior studies in the literature.[17] Table A1 in Appendix shows the detailed definition of these variables. The

[17] Schenone (2010) introduce an intensity-based relationship variable: the number of historical loans by a bank divided by the total number of loans to date. Bharath et al. (2011) construct two variables: (1) the ratio of total amounts of loans by bank j to borrower i in the last 5 years over the total amounts borrowed by the borrower i in the last 5 years; (2) the ratio of total number of loans by bank j to borrower i in the last 5 years over the total number of loans borrowed by the borrower i in the last 5 year. Regarding our sample period,

average distance between firms and their closet bank branches is 59.7 km while the median is about 2.4 km. This is consistent with the statistics in other countries.[18] For *Bank_Num*, on average, the borrowers have 3.2 lending banks (e.g., Bae, Kang and Lim (2002); Jiménez, Salas, and Saurina (2006)). For *LoanAmount_Share* and *LoanFreq_Share*, firms borrow around 45% of their loans (either in loan amount or loan number) from a particular bank (e.g., Schenone (2010); Bharath et al. (2011)). For *Recent_Lending*, among over 7.5 million firm-bank-month observations, there are 80.8% of them having borrowing experience from CBRC sample banks in the past 12 months (e.g., Ioannidou and Ongena (2010)).[19] This number turns to be 77.6% for new loan issuance sample, which suggests that 77.6% new loans in our sample flow into "existing" borrowers that have got loans from CBRC banks in the past 12 months.

[Place Table 1 about here]

Panel B presents the summary of loan contract terms and borrower characteristics. The average amount of loan is around 15 million RMB with short-term maturity. Approximately 21% of loans have third-party guarantee, which provides a credit enhancement scheme for lenders since the guarantors are legally liable to pay back the loans under default event. The default rate defined as over 90 days delinquency is 1.1% that is comparable to the non-performing loan rate disclosed in banks' annual reports in China. The average size of firms equals 4.0 billion RMB while the standard deviations are large. On average, the firm leverage level is 61.3%. Moreover, we also construct the borrower connection variables regarding various firm inter-links. In particular, *Sibling Firms Connected, Other Related Firms Connected*, and *Stock Holder Connected* are the dummies defined at individual loan level for whether the borrower has any sibling firms (i.e., firms with the same controlling parent company), firms with related party

---

we trace back one year to define our intensity-based relationship variables. Our results are robust to different windows.

[18] For example, Agarwal and Hauswald (2010) document that the mean and median values of bank-borrower distance in their sample are 9.9 and 2.6 miles, respectively. In Petersen and Rajan (2002), although the distance between a firm and its lender varies across lender type and the year the relationship began, the median value ranges from 2.0 to 5.0 miles. Using the Turkey data, the distance number calculated in Beck, Ongena, and Şendeniz-Yüncü (2018) is smaller than ours. The median value of distance between the borrower and the closest branch ranges from 0.2 km to 1.5 km.

[19] This data set is based on outstanding loans, which is different from the new loan sample when we examine the distance effect on loan contract terms and default risk.

transactions, and shareholders that are located in the same city as the bank and also has been borrowing from this bank. *Connection#* is the total number of these three types of connections. On average, 16.6%, 18%, and 39.8% of the borrowers are connected to their borrowing banks via sibling firms, firms with related transactions, and shareholders that have also been borrowing from the same banks, respectively.

Panel C summaries the early downgrade activities condition on delinquent loans. As shown in this Panel, around 60% of delinquent loans have no rating changes at all while 41.8% of the default loans were downgraded before the delinquent date. Moreover, on average, banks downgrade these delinquent loans 3.3 months before the actual delinquent time.

4. Empirical Analysis and Results

4.1. Non-monotonic Lending over Lender-Borrower Distance

We start our analysis by comparing various commonly used measures of lender-borrower relationships in the literature. In particular, for each loan, we examine the distance between the bank branch and firm vs. the variables for lending intensities. The conventional wisdom is that banks have better soft information when they have been lending more intensively to the firms (e.g., more frequent, more shares, and fewer lenders). We construct four additional relationship measurements following previous studies in the literature. For each pair of bank-firm-month, we calculate the number of banks with loans outstanding (*Bank_Num*), two variables of relationship strength measures, i.e., one is the share of loan amounts that have been borrowed from the bank (*LoanAmount_Share*), and the other is the share of loan numbers that have been borrowed from the bank (*LoanFreq_Share*), and the dummy for whether the firm borrowed any new loans from the bank in the past 12 months (*Recent_Lending*).

Based on these variables, we first calculate the correlation matrix in Panel A of Table 2. We find that the correlations between *LogDistance* and other four lending intensity variables are near zero. The variation of distance is almost orthogonal to the other proxies of bank-firm relationships. Moreover, even among the four traditional lending relationship variables (e.g., *Bank_Num, LoanAmount_Share, LoanFreq_Share, Recent_Lending*), not all the correlations are very strong. For example, the correlations between *Recent_Lending* vs. *Bank_Num, LoanAmount_Share, and LoanFreq_Share* are only 0.05, 0.53, and 0.54, respectively. These numbers are also low which shows that these commonly used relationship measurements might not be good proxies. In other words, if these variables can measure the lender-borrower relationship (i.e., soft information) by large, they should have correlated with each other closely since they measure the same factor. This low correlation phenomenon is consistent with studies in other countries (e.g., Bolton et al. (2016); Hombert and Matray (2016); Karolyi (2017)). This suggests that either the distance and other proxies are noisy measurements of bank-firm relationships or these proxies capture the different aspects of soft information.

In Panel B, we calculate the correlation matrix among the five relationship variables and the loan characteristics. We find that the unconditional correlations are again very low. For example, the correlations between ex-post loan performance *Default* and *LogDistance*, *Recent_Lending, Bank_Num, LoanAmount_Share, and LoanFreq_Share* are -0.01, -0.04, -0.02, -0.02, and -0.01, respectively.

[Place Table 2 about here]

Next, we explore the non-monotonic relationships between the distance and lending intensity. We perform the OLS regressions of the four lending intensity variables on the quadratic forms of distance (i.e., $LogDistance$ and $Log^2Distance$). The regression equation is:

$$Y_{i,j,t} = \beta_1 \times LogDistance_{i,j,t} + \beta_2 \times Log^2Distance_{i,j,t} + \text{Firm}\times\text{Year FE} + \text{BankFE} + \varepsilon \quad (1),$$

where $Y_{i,j,t}$ is the lending intensity measures (i.e., *Recent_Lending* and *Bank_Num, LoanAmount_Share, and LoanFreq_Share*) for firm i from bank j at month t, $LogDistance_{i,j,t}$ is the natural logarithm of the shortest distance between firm i and bank j's branches at month t when the loan is issued, and $Log^2Distance_{i,j,t}$ is the quadratic term of $LogDistance_{i,j,t}$ (i.e., $LogDistance_{i,j,t} \times LogDistance_{i,j,t}$). We control for the firm×year fixed effects and bank fixed effects to absorb any variation attributable to constant characteristics within firm years and banks, respectively. The firm-year fixed effects absorb all demand factors of firm borrowing which allows us to examine the variation between distance and lending intensity within the firm-year across banks. The robust standard errors are clustered at the firm level.

Table 3 shows the regression results. In columns (2) to (4), for *LoanAmount_Share, LoanFreq_Share, and Recent_Lending,* the linear terms of *LogDistance* have significantly negative coefficients while the quadratic terms of *LogDistance* have significantly positive coefficients. Moreover, in column (1), distance also shows the *U*-shaped relationships on *Bank_Num* (i.e., the firms borrow from more other banks when distance increases but significantly fewer other banks when the distance is very long). These *non-monotonic* relationships suggest that the lending intensities are significantly higher for firms which are either closer to the bank branches or more distant from the bank branches than the firms in the middle. We perform the back-to-the-envelope calculation to predict the patterns between distance and lending intensity based on the estimated coefficients in Table 3. Figure 1 shows very strong predicted *U*-shaped patterns. For example, consistent with the results in Table 3, the *LoanFreq_Share* decrease with distance between bank branches and firms within 250 km and starts to increase with distance when it is beyond 250 km.

[Place Table 3 and Figure 1 about here]

Furthermore, we find that the distant loans are mainly from the bank branches located outside of borrowers' cities (i.e., outside city loans). For example, the average lender-borrower distance is 277 km for outside city loans while it is 42 km for inside city loans (i.e., borrowers and banks are in the same city). [20] This is consistent with Figure 1 with the parabola vertex at 250 km. Thus, we use 250 km as the cut-off to define the short- vs. long-distance loans. In un-reported robustness checks, our main results remain when we try different cut-offs (e.g., 100 km, 200 km, 300 km, and 400 km). Overall, 11% of the loans go to the distant borrowers. The amount of distant lending in China is substantial and consistent with other countries which have significant amounts of distant lending as well (see Table A2 in Appendix for the summary). The rationale of these intensive lending to long-distance borrowers is beyond the traditional wisdom that the borrower proximity

[20] In Table A3, outside city loans are for bigger firms with larger assets and more employees. For other loan and borrower characteristics, we find that long distance borrowers are larger in size, smaller percentage of SOEs, more efficient in terms of Assets Turnover Ratio (ATRs) or Total Factor Productivity (TFP).

could lower down banks' costs of loan monitoring and screening to extract soft information of the borrowers.

4.2. *Distant Lending and Inter-firm Network Connection*

Next, we explore the underlying channels behind the significant amount of distant lending. In particular, we study the role of the inter-firm network in banks' soft information collection. To formally test it, we regress various connection measures (i.e., *Sibling Firms Connected, Other Related Firms Connected*, *Stock Holder Connected,* and *Connection#*) on the logarithm distance between bank branches and borrowers. We control for loan- level characteristics such as internal rating, guarantee status, loan size, and loan maturity, as well as firm-level characteristics such as firm size and leverage. We also control for the firm×year fixed effects and bank fixed effect.

Table 4 shows the regression results. We find that the coefficients of *LogDistance* are significantly positive for the borrowers' connection. For example, in column (1), the coefficient of *LogDistance* is 0.034 at 1% significant level. This means that a 100% increase in *LogDistance* is associated with a 3.2% (i.e., log(2)×0.034/0.744) increase in borrowers' number of connections. More specifically, column (2) and (3) show that the distant borrowers are significantly more likely to have sibling firms and firms with related transactions that are borrowing from the same banks in the same city.

[Place Table 4 about here]

The results in Table 4 provides an underlying mechanism for the *U*-shaped patterns and distant lending documented in Table 3. On the one hand, under the conventional wisdom, banks issue more relationship loans to firms close-by due to the low monitoring and screening costs of soft information. This force dominates when borrowers are relatively close to the lenders. Consequently, the shorter the distance is associated with the higher the lending intensities since banks can cheaply identify the good quality borrowers close-by. On the other hand, when the lender-borrower distance passes a certain point, the costs of loan monitoring and screening are mostly fixed and don't vary too much by distance. To find good borrowers, banks need more soft information passed via the firm network that became the primary method of soft information collection. In other words, among distant borrowers, banks prefer the ones that are connected to banks local borrowers for the soft information collection. Our findings in Table 4 suggest that the firm network could overcome the increased costs of soft information collection and, thus, facilitate the distant lending.

4.3. Inter-firm Network Connection and Soft Information

Furthermore, we exploit the novel internal rating change data to measure the soft information available to banks directly. In particular, our data record the monthly internal ratings of individual loans. This allows us to trace whether banks downgrade the internal ratings before the actual delinquent events. We assume that better soft information could help banks predict and manage the default risks. Specifically, condition on delinquent loans, we define the variable "*Early Downgrade*" as the dummy for whether the bank downgrades the internal ratings before the month of the delinquency. Compared with the other soft information measurements (e.g., distance and lending intensity), the early downgrading is a more direct measure of banks' soft information on borrowers. This is a huge step forward in the relationship banking literature.

To understand the role of the firm network in soft information collection, we stratify our sample into short- vs. long-distance loans. We then perform the OLS regressions of dummy *Early Downgrade* on whether delinquent borrowers' connections in the firm network. Table 5 shows the regression results where column (1) to (4) are for distant borrowers (i.e., beyond 250km) and column (5) to (8) are for borrowers close by (i.e., within 250km). In column (1) of Panel A, the coefficient of *Connection#* is 0.136 at 5% significant level. This means that, for distant borrowers, one more connection to the borrower is associated with a 13.6% increase in the banks' ability to early prediction of delinquent events. More specifically, in column (2) to (4), all three types of connections (i.e., *Sibling Firms Connected, Other Related Firms Connected*, and *Stock Holder Connected*) have significantly positive coefficients which suggest that the distant borrowers' soft information could be passed via these connections to banks. In contrast, for borrowers close by, we don't find any significant effects of firm connections on early downgrade from column (5) to (8). Banks need the firm network to collect soft information only for distant borrowers.

Moreover, in Panel B of Table 5, we study how many months in advance banks can predict the delinquent events. In particular, the dependent variable is *Log(Months Earlier)* that is the logarithm of one plus the number of months between the date of the first downgrade and the delinquent time. For long distant borrowers in column (1) to (4), coefficients of borrower connections are significantly positive. For example, in column (1), the coefficient of Connection# is 0.336 with 1% significant level. This means that, on average, banks can predict the delinquent event 33.6% earlier for the connected borrowers in distant lending. For borrowers close-by, consistent with Panel A, we don't find significant effects of the firm network on the months of early prediction.

There are two potential concerns about using the early downgrade to measure soft information. First, rating changes can be strategic since they would be shared in public and other banks can free ride on it (e.g., Giannetti, Liberti, and Sturgess (2017)). Second, banks can use both public and private information in their downgrade decisions. We use the internal loan rating changes in the analysis to mitigate the concern about free-rider since they are not shared among banks. Moreover, we control for the firm×year fixed effects to absorb any variation at firm-year level including firms' public information. The variation we exploit is within the same firm-year but across lenders banks which captures only the soft information of specific lenders. Our findings suggest that firm network is an important channel of soft information collection, especially for distant lending.

[Place Table 5 about here]

4.4. Interaction of Distance and Firm Network in Soft Information

Besides the estimations on two subsamples in Table 5, we further analyze the interaction between distance and firm network connections for soft information collection. In particular, we perform the regressions of the early downgrade on the interaction terms of distance and various borrower connections. The Panel A of Table 6 shows the OLS regression results. On the one hand, in column (1) to (4), all the coefficients of *Log(Distance)* are significantly negative which means that the longer lender-borrower distance makes it harder for banks to foresee the delinquent events. This is consistent with the conventional wisdom that borrower proximity facilitates soft information collection by lowering the costs of loan monitoring and screening. On the other hand, the coefficients of interaction terms between *Log(Distance)* and borrower connection are significantly positive in column (1) to (4). For example, in column (1), the coefficient of *Log(Distance)×Connection#* is 0.057 at 5% significant level. This means that, for distant lending, the firm network connection could overcome the increased costs of loan monitoring and screening for borrowers' soft information.

[Place Table 6 about here]

4.4.1. Identification of Distance's Effect on Soft Information Collection

In Panel A of Table 6, we have shown the different underlying mechanisms of soft information collection in short- vs. long-distance lending. One caveat is that banks don't choose their branch locations randomly and the borrower proximity might be correlated with other fundamental factors such as local investment opportunities. To establish the causal effects of borrower proximity on different channels of soft information collection, we use the 2009 bank entry deregulation in China as the instrument for physical distance between banks and firms. As described in Section 2, the bank entry restriction introduced in 2006 put a huge restriction on joint equity banks' branch expansion. This entry barrier was partially lifted in 2009.

In April 2009, the CBRC initialed a series of reforms in banking system including the deregulation on branch entry restriction. Specifically, after April 2009, the joint equity banks can open branches freely in a city if this bank had already established branches in this city or their capital city. For example, when the joint equity bank has existing branches in city A or provincial capital city B, it can apply multiple branch opening requests at once. Moreover, the local CBRC offices have the discretion to approve the branch applications which makes the whole process much faster (i.e., usually within four months). Furthermore, the CBRC also removed the total number caps of branches as well as the requirement on capitals for new branches. For the joint equity banks which didn't have any branches in the city or the provincial capital city, they were still regulated under the entry rules and can't open branches freely as in the deregulated cities. Overall, for joint equity banks, 38.5% of the city-bank pairs are deregulated after 2009, and the other 61.5% are still under the 2006 CBRC bank entry regulation. Moreover, the big five banks are not directly affected by the 2009 deregulation, and they have established branches in almost all cities in China before 2009.

The 2009 deregulation provides an ideal empirical setting to tease out the exogenous variation in the distance between the bank branches and borrowers. In particular, the 2009 deregulation only lifts the entry barrier for joint equity banks in specific regions. Even within the same city, different banks could have different exposure to the shock, depending on whether they have existing branches in the areas or not. The exogenous variation is across banks and cities. It is hardly confounded with the city-specific time trends since different banks could have different exposures to the shock in the same city.[21]

To formally employ this shock, we use the 2009 bank entry deregulation in China as the instrument for physical distance between banks and firms. We first perform the regression as follows:

$$\begin{aligned} &LogDistance_{i,j,t} \\ &\quad = \beta_1 \times Shock2009_{i,j,t} + \beta_2 \times Treatment_{i,j,t} + \beta_3 \times After2009_t + \text{Firm} \\ &\quad \times \text{YearFE} + \text{BankFE} + \varepsilon, (2) \end{aligned}$$

where $LogDistance_{i,j,t}$ is the logarithm of the physical distance between firm i from bank j at month t. $Treatment_{i,j,t}$ is the dummy for whether bank j can freely open branches in firm i's city (i.e., deregulated) after April 2009. $After2009_t$ is the dummy for whether it is after April 2009 or not at monthly frequency. $Shock2009_{i,j,t} = Treatment_{i,j,t} \times After2009_t$. We control for the firm×year fixed effects and bank fixed effects. Table A4 shows the first stage regression results. Column (1) is for 3 months before and after the shock (e.g., between Jan 2009 and June 2009) and the coefficient of *Shock2009* is -0.003 at 1% significance level. Moreover, for longer term effects in Column (2) to (5), *Shock2009* has significantly negative coefficients and the magnitude increases when the regression time window is longer. For example, when we include all years available (i.e., between Oct 2006 and June 2013), the regression coefficient of *Shock2009* is -0.014 at 1% significance level. This suggests that the bank branch-firm distance

[21] Gao et al. (2018) use the same entry deregulation shock and find that there are no significant pre-trends. This further support that this deregulation was not expected by the markets and was not driven by other underlying demand forces.

decreased by 1.4% after the 2009 bank entry deregulation when joint equity banks can open branches freely in deregulated cities which shorten the distance significantly.

We then follow the standard approach in Wooldridge (2002) by using the predicted $LogDistance$ (i.e., $Log\widehat{Distance}$) from equation (2) and its interaction form (i.e., $Log\widehat{Distance} \times \widehat{Connection\ Variables}$) as the instruments for both $LogDistance$ and $LogDistance \times Connection\ Variables$ to perform the standard 2SLS. Table 6 Panel B shows the second stage regression results. In column (1) to (4), consistent with the patterns in OLS regressions, the coefficients of interaction terms between $LogDistance$ and borrower connections are significantly positive while the coefficients of $LogDistance$ are significantly negative. Column (5) to (8) are for *Log(Months Earlier)*, and show similar patterns. Specifically, a 100% increase in distance leads to 13.2%, 18.8%, and 15.5% increases in early prediction length for borrowers connected via their sibling firms, firms with related transactions, and shareholders respectively.

In sum, these results show the different underlying mechanisms to collect soft information in short- vs. long-distance borrowers. On the one hand, consistent with the conventional wisdom, when borrowers are geographically close to the bank, the firm network doesn't seem to play a role in passing soft information since the bank can cheaply obtain borrowers' soft information by screening and monitoring. On the other hand, for distant lending, banks mainly collect soft information via the firm network which overcomes the increased costs of loan screening and monitoring and helps bank choose good quality distant borrowers. The soft information improves the predictions of loan delinquency and facilitates the risk management.

4.5. SME vs. Large Firms

In a next step, we explore firm networks' heterogenous effects on soft information collection across various borrowers. In particular, the relationship lending is mainly for the SMEs while the large firms borrow more transaction loans since the information asymmetry is more pronounced for SMEs than large firms. We expect the firm networks plays a bigger role for SMEs in soft information collection. We use the official classification of SME in China to stratify our sample into SMEs and large firms. Specifically, the State Council issued a document, "Standards for Classification of Small and Medium-sized Enterprises", to classify the SMEs and we use firms' total sales in 2008 to separate SMEs and large firms. We then repeat the analysis in Table 6 in two subsamples, respectively. Table 7 shows the regression results. In Panel A, column (1) to (4) are for SMEs, and the coefficient of interaction terms between *LogDistance* and borrower connections are positive. For example, in column (1), the coefficient of *LogDistance×Connection#* is 0.051 at 5% significant level while the coefficient of *LogDistance* is significantly negative.

In contrast, column (5) to (8) are for large firms, and the coefficient of interaction terms between *LogDistance* and borrower connections are not positive. Moreover, Panel B shows the 2SLS regression results by using the 2009 deregulation as the IV for distance. Consistent with Panel A, we find significant positive coefficients of interaction terms between *LogDistance* and borrower connections only for SMEs. This effect is muted for large firms. Furthermore, we also use *Log(Months Earlier)* as the dependent variable in Table A5 in Appendix and find similar patterns as in Table 7. These findings support the hypothesis that the relationship lending is mainly for the SMEs.

[Place Table 7 about here]

Moreover, 21% of the loans in our sample have third-party guarantees. The guarantors are legally liable to make loan payments under default events which lower down banks' incentives to collect soft information for the risk management. In Table A6, we stratify the sample into loans with and without third-party guarantees and repeat the analysis in Table 6. We find that the firm network facilitates soft information collection only for loans without guarantees which suggests that banks have lower incentives to collect soft information when the loan payments are guaranteed. The findings of SMEs and loan guarantees further strengthen the firm network channel of soft information collection which is novel in the literature.

4.6. Inter-firm Network Connection and Loan Performance

Lastly, we study the economic consequences of the firm network on distant lending regarding the default risks. Specifically, we run OLS regressions of defaults on the inter-firm network connections. Table 8 shows the regression results. In particular, the coefficients of borrower connections have significantly negative coefficients. For example, in column (1), the coefficient of *Connection#* is -0.001 at 1% significant level. This means that one more connection for borrowers in the firm network is associated with a 0.1% decrease in default rate. The average default rate in China is 1.1% so that this reduction equals approximately a 10% decrease in default rate which is both statistically and economically significant. These results suggest that the banks can indeed use the soft information collected from the firm network to choose the good quality borrowers. Subsequently, the banks can better manage risks and improve the loan performance. This echoes Norden and Weber (2010) who find that public credit ratings and account information is related to the default of short-distance borrowers but not for long distance borrowers. Banks need the soft information to facilitate their distant lending.

[Place Table 8 about here]

5. Conclusion

In this paper, we document a novel channel of soft information collection. In particular, we find that soft information could be passed through the inter-firm network to facilitate lending, especially for distant borrowers. On the other hand, borrower proximity lowers down the costs of monitoring and screening to collect soft information which helps the banks to choose the better quality borrowers close-by. We use big data in China's banking sector to disentangle these two different underlying mechanisms of extracting soft information in short- vs. long-distance lending. This sheds light on the non-trivial distant lending in many countries worldwide.

Furthermore, by tracing out the monthly internal loan rating changes, we can observe whether the banks can predict the delinquency in advance by lowering down the internal ratings before the actual late payments. This is a big step forward in the literature since the previous studies use indirect proxies to extrapolate the degrees of lender-borrower relationships. The novel dataset of rating changes allows us to observe banks' monitoring dynamics better and to explore the different underlying forces behind these conventional proxies such as distance. This sheds light on the long-standing puzzle of low correlations among conventional proxies of soft information.

China is the second largest economy worldwide, and it has developed the largest banking market across the globe. It is essential to understand how banks collect soft information and issue relationship loans which is vital for the SMEs. This, in turn, would equip us with better tools to study the economic consequences of the relationship lending not only in China but also in other countries.

References

Agarwal, S, and Hauswald, R., 2010. Distance and private information in lending. Review of Financial Studies 23, 2757–2788.

Allen, F., 1990. The market for information and the origin of financial intermediation. Journal of Financial Intermediation 1, 3-30.

Ayyagari, M., Demirgüç-Kunt, A., and Maksimovic, V., 2010. Formal versus informal finance: Evidence from China. Review of Financial Studies 23, 3048-3097.

Bae, K. H., Kang, J. K., and Lim, C. W., 2002. The value of durable bank relationships: Evidence from Korean banking shocks. Journal of Financial Economics 64, 181-214.

Bae, K. H., Stulz, M. R., and Tan, H. P., 2008, Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts. Journal of Financial Economics 88, 581-606.

Beck, T., and Ongena, S., and Sendeniz-Yuncu, I., 2018, Keep walking? Geographical proximity, religion, and relationship banking. Journal of Corporate Finance forthcoming.

Bellucci, A., Borisov, A., and Zazzaro, A., 2013. Do banks price discriminate spatially? Evidence from small business lending in local credit markets. Journal of Banking & Finance 37, 4183-4197.

Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G., and Stein, J. C., 2005. Does function follow organizational form? Evidence from the lending practices of large and small banks. Journal of Financial Economics 76, 237-269.

Berger, A.N. and Udell, G.F., 1995. Relationship lending and lines of credit in small firm finance. Journal of Business 68, 351-381.

Bharath, S., Dahiya S., Saunders A., and Srinivasan A., 2011. Lending relationships and loan contract terms. Review of Financial Studies 4, 1141-1203.

Bolton, P., Freixas, X., Gambacorta, L., and Mistrulli, P. E., 2016. Relationship and transaction lending in a crisis. Review of Financial Studies 29, 2643-2676.

Boot, A.W. and Thakor, A.V., 2000. Can relationship banking survive competition? Journal of Finance 55, 679-713.

Brown, N. C., Stice, H., and White, R. M., 2015. Mobile communication and local information flow: Evidence from distracted driving laws. Journal of Accounting Research 53, 275-329.

Butler, A.W., 2008. Distance still matters: Evidence from municipal bond underwriting. Review of Financial Studies 22, 763–784,

Calomiris, C., 2000. U.S. Bank deregulation in historical perspective. Cambridge University Press, New York.

Carbo-Valverde, S., Rodriguez-Fernandez, F., and Udell, G. F., 2009. Bank market power and SME financing constraints. Review of Finance 13, 309-340.

Chhaochharia, V., Kumar, A., and Niessen-Ruenzi, A., 2012. Local investors and corporate governance. Journal of Accounting and Economics 54, 42-67.

Coval, J. D., and Moskowitz, T. J., 1999. Home bias at home: Local equity preference in domestic portfolios. Journal of Finance 54, 2045-2073.

Coval, J. D., and Moskowitz, T. J., 2001, The geography of investment: Informed trading and asset prices. Journal of Political Economy 109, 811–841.

Da, Z., Gurun, U.G., Li, B., and Warachka, M., 2018, Investment in a smaller world: The implications of air travel for investors and firms. Working Paper.

Dass, N. and Massa, M., 2011. The impact of a strong bank-firm relationship on the borrowing firm. Review of Financial Studies 24, 1204-1260.

Degryse, H. and Ongena, S., 2005. Distance, lending relationships, and competition. Journal of Finance 60, 231-266.

DeYoung, R., Glennon, D., and Nigro, P., 2008. Borrower-lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. Journal of Financial Intermediation 17, 113-143.

Diamond, D., 1984. Financial intermediation and delegated monitoring, Review of Economic Studies 51, 393-414.

Gao, H., Ru, H., Townsend, R., and Yang X., 2018. Rise of bank competition: Evidence from banking deregulation in China. Working Paper.

Giannetti, M., Liberti, J., and Sturgess, J., 2017. Information sharing and rating manipulation. Review of Financial Studies 30, 3269–3304.

Giroud, X., 2013, Proximity and investment: Evidence from plant-level data. Quarterly

Journal of Economics 128, 861-915.

Herpfer, C., Mjøs, A., and Schmidt, C., 2018. The causal impact of distance on bank lending. SSRN Working Paper.

Hollander, S., and Verriest, A., 2016. Bridging the gap: The design of bank loan contracts and distance. Journal of Financial Economics 119, 399-419.

Hombert, J., and Matray, A., 2016. The real effects of lending relationships on innovative firms and inventor mobility. Review of Financial Studies 30, 2413-2445.

Ioannidou, V. and Ongena, S., 2010. "Time for a change": Loan conditions and bank behavior when firms switch banks. Journal of Finance 65, 1847-1877.

Ivković, Z., and Weisbenner, S., 2005. Local does as local is: Information content of the geography of individual investors' common stock investments. Journal of Finance 60, 267-306.

Jensen, M.C., 1986. Agency costs of free cash flow, corporate finance, and takeovers. American Economic Review 76, 332–329.

Karolyi, S. A., 2018. Personal lending relationships. Journal of Finance 73, 5-49.

Nakamura, L.I., and Roszbach, K., 2016. Credit ratings, private information, and bank monitoring ability. FRB of Philadelphia Working Paper No. 16-14.

Norden, L., and Weber, M., 2010. Credit line usage, checking account activity, and default risk of bank borrowers. Review of Financial Studies 23, 3665-3699.

Petersen, M.A., and Rajan, R.G., 1994. The benefits of lending relationships: Evidence from small business data. Journal of Finance 49, 3-37.

Petersen, M. A., and R. G. Rajan. 2002. Does distance still matter? The information revolution in small business lending. Journal of Finance 57, 2533–70.

Santos, J. and Winton, A., 2008. Bank loans, bonds, and informational monopolies across the business Cycle. Journal of Finance 63, 1315-1359.

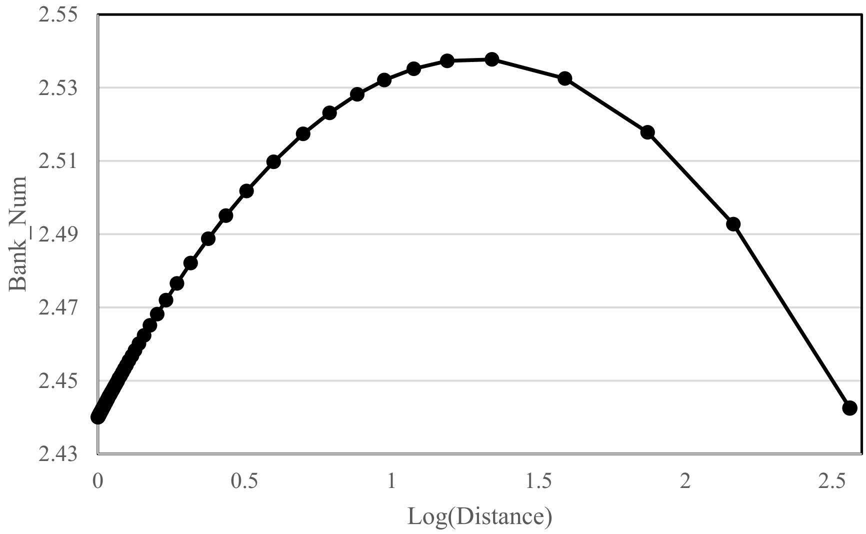
Schenone, C., 2009. Lending relationships and information rents: Do banks exploit their information advantages? Review of Financial Studies 23, 1149-1199.

Seasholes, M. S., and Zhu, N., 2005. Is there information in the local portfolio choices of individuals? Journal of Finance 65, 1987-2010.

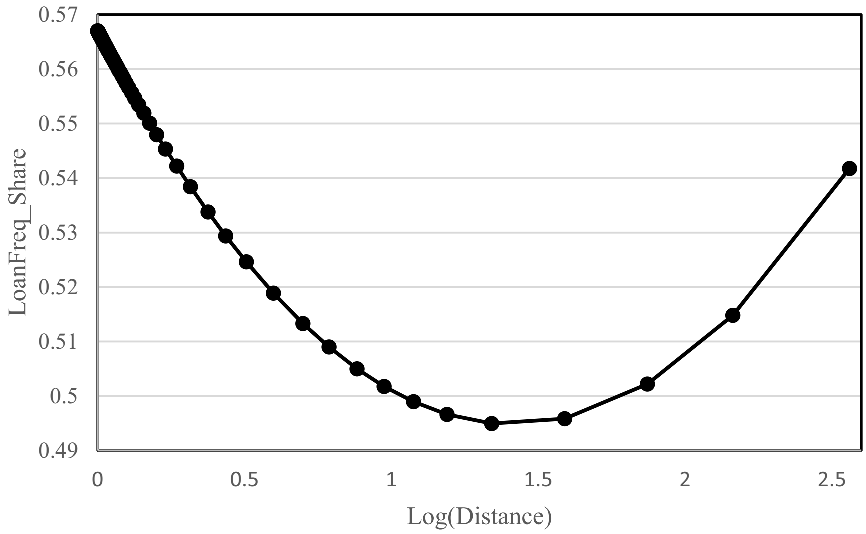
Seasholes, M. S., and Zhu, N., 2010. Individual investors and local bias. Journal of Finance 65, 1987-2010.

Winton, A., 1995, Delegated monitoring and bank structure in a finite economy. Journal of Financial Intermediation 4, 158-187.

Wooldridge, J., 2002, Econometric Analysis of Cross Section and Panel Data, MIT Press.



Panel A: Distance and Number of Banks



Panel B: Distance and Lending Intensity

Figure 1. Distance and Lending Relationship. The figure plots the *U*-shape patterns between geographical lender-borrower distance and lending intensities. The horizontal axis represents the logarithm of one plus the distance between bank branches and borrowers. In Panel A, the vertical axis represents the predicted number of banks with loans outstanding based on the regression coefficients estimated in Table 3. In Panel B, the vertical axis represents the predicted share of loans that have been borrowed from the bank based on the regression coefficients estimated in Table 3.

Table 1: Summary Statistics

Table 1 presents summary statistics of main variables in the empirical analysis. Panel A consists of 7,623,883 firm-bank-month observations. The sample is from Oct 2006 to Jun 2013 and covers 17 commercial banks (i.e., the big five commercial banks and the twelve joint equity banks). Panel B consists of 6,755,227 loans and reports the descriptive statistics of the loan- and borrower-characteristics. Panel C consists of 46,661 delinquent loans. Mean, standard deviation, P25, median and P75 are reported. Definitions for these variables are described in Appendix Table A1.

	N	Mean	Median	Std. Dev.	P25	P75
	Panel A: Firm-Bank-Month Sample					
Distance (100KM)	7,623,859	0.597	0.024	1.972	0.007	0.117
Bank_Num	7,623,883	3.194	2.000	2.735	1.000	4.000
LoanAmount_Share	6,161,111	0.446	0.500	0.386	0.167	1.000
LoanFreq_Share	6,161,111	0.448	0.495	0.389	0.184	1.000
Recent_Lending	7,623,883	0.808	1.000	0.394	1.000	1.000
	Panel B: Loan Issuance Sample					
Default	5,581,460	0.012	0.000	0.108	0.000	0.000
Loan Amount	6,824,375	14.921	4.000	30.889	0.600	13.000
Rating	6,824,375	0.978	1.000	0.146	1.000	1.000
Maturity	6,824,375	11.552	6.000	21.019	4.000	12.000
Guarantee	6,824,375	0.210	0.000	0.408	0.000	0.000
Assets (Billion RMB)	6,824375	4.055	0.799	9.361	0.273	3.049
Leverage	6,824,331	0.613	0.611	0.187	0.490	0.734
SOE	5,613,464	0.104	0.000	0.306	0.000	0.000
Employees	5,613,464	1777.710	374.000	3985.735	70.000	1471.000
Connection #	6,823,955	0.744	0.000	0.931	0.000	1.000
Sibling Firms Connected	6,823,955	0.166	0.000	0.372	0.000	0.000
Other Relation Connected	6,823,955	0.398	0.000	0.489	0.000	1.000
Stock Holder Connected	6,823,955	0.180	0.000	0.384	0.000	0.000
	Panel C: Delinquent Loan Sample					
Early Downgrade	19,051	0.418	0.000	0.493	0.000	1.000
Months Earlier	19,051	3.299	0.000	5.578	0.000	5.000

Table 2: Correlation Matrix

Table 2 reports the correlation matrix of relationship lending measurements and loan characteristics. Panel A reports the correlation matrix of the five commonly used variables for borrower-lender relationships (i.e., measures of soft information) at the firm-bank-month level. Panel B reports the correlation matrix of the loan contract terms and the relationship measurements at the loan level. Definitions for these variables are described in Appendix Table A1.

Panel A: Firm-Bank-Month Sample

		1	2	3	4	5
LogDistance	1	1.00				
Bank_Num	2	0.04	1.00			
LoanAmount_Share	3	-0.05	-0.51	1.00		
LoanFreq_Share	4	-0.04	-0.52	0.97	1.00	
Recent_Lending	5	0.00	0.05	0.53	0.54	1.00

Panel B: Loan Sample

		1	2	3	4	5	6	7	8	9	10
Loanamt	1	1.00									
Rating	2	-0.03	1.00								
Maturity	3	0.30	-0.03	1.00							
Guarantee	4	0.10	-0.08	0.07	1.00						
Default	5	0.05	-0.17	0.02	0.03	1.00					
LogDistance	6	-0.03	0.03	-0.04	-0.05	-0.01	1.00				
Bank_Num	7	0.19	0.04	0.04	-0.04	-0.02	-0.05	1.00			
LoanAmount_Share	8	-0.18	0.00	-0.07	-0.03	-0.02	-0.01	-0.31	1.00		
LoanFreq_Share	9	-0.13	-0.01	-0.05	0.00	-0.01	-0.01	-0.39	0.92	1.00	
Recent_Lending	10	-0.06	0.02	-0.08	-0.02	-0.04	-0.05	0.17	0.64	0.60	1.00

Table 3: Distance and Lending Intensities

Table 3 reports the OLS regressions of various lending relationship proxies on the quadratic terms of distance between bank branches and borrowers. The sample consists of 7,623,883 firm-bank-month observations from Oct 2006 to Jun 2013. The dependent variables are *Bank_Num*, *LoanFreq_Share*, *LoanAmount_Share,* and *Recent_Lending*, respectively. The main independent variables are the linear and quadratic terms of *LogDistance* that is the natural logarithm of one plus the physical distance between banks and borrowers. We control for firm×year fixed effects and bank fixed effects in all regressions. The constants are omitted for brevity. The *t*-statistics based on standard errors clustered at firm level are reported in parentheses. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	Bank_Num	LoanFreq_Share	LoanAmount_Share	Recent_Lending
Log^2(Distance)	-0.029***	0.033***	0.016***	0.004**
	(-3.61)	(15.97)	(8.71)	(2.14)
Log(Distance)	0.054***	-0.091***	-0.037***	-0.030***
	(3.17)	(-20.86)	(-9.32)	(-7.25)
Firm×Year FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
Observations	7,615,119	6,154,977	6,154,977	7,615,119
Adj. R-squared	0.869	0.700	0.717	0.225

Table 4: Distance and Connections via Firm-Network

Table 4 reports the OLS regressions of various borrower connection measurements on the logarithm of borrower-lender distance. *Sibling Firms Connected*, *Other Related Firms Connected*, and *Stock Holder Connected* are the dummies defined at individual loan level for whether the borrower has any sibling firms (i.e., firms with the same controlling parent company), firms with related party transactions, and shareholders that are located in the same city as the bank and also has been borrowing from this bank. *Connection #* is the total number of these three types of connections. The main independent variable is *LogDistance*, the natural logarithm of one plus the borrower-lender distance. We control for loan- level characteristics such as internal rating, guarantee status, loan size, and loan maturity, and firm-level characteristics such as firm size and leverage in all columns. The coefficients of constants and control variables are omitted for brevity. We control for firm×year fixed effects and bank fixed effects in all regressions. The *t*-statistics based on standard errors clustered at firm level are reported in parentheses. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	Connection #	Sibling Firms Connected	Other Relation Connected	Stock Holder Connected
Log(Distance)	0.034***	0.015***	0.020***	0.000
	(23.26)	(26.08)	(24.72)	(0.11)
Loan Controls	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes
Firm×Year FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
Observations	6,784,908	6,784,908	6,784,908	6,784,908
Adj. R-squared	0.678	0.624	0.627	0.648

Table 5: Firm Network Connections and Soft Information

Table 5 reports the OLS regression of bank's early downgrade warning on the borrower connections in firm network. The sample consists of 19,051 delinquent loans from Jan 2007 to Jun 2013. Panel A reports the results of OLS regressions of *Early Downgrade*, a dummy variable indicating whether this delinquent loan's internal rating is downgraded before the delinquency. Panel B reports the results of OLS regressions of *Log(Months Earlier)* that is the logarithm of one plus the months between date of the first downgrade and the delinquent date. In each panel, columns (1) to (4) report the subsample of long-distance loans (i.e., the distance is longer than 250 km). Columns (5) to (8) report the subsample of short-distance loans (i.e., the distance is shorter than 250 km). *Sibling Firms Connected*, *Other Related Firms Connected*, and *Stock Holder Connected* are the dummies defined at individual loan level for whether the borrower has any sibling firms (i.e., firms with the same controlling parent company), firms with related party transactions, and shareholders that are located in the same city as the bank and also has been borrowing from this bank. *Connection#* is the total number of these three types of connections. The coefficients of constants and control variables are omitted for brevity. We control for firm×year fixed effects and bank fixed effects in all regressions. The *t*-statistics based on standard errors clustered at firm level are reported in parentheses. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Early Downgrade

	Early Downgrade							
	Long Distance Borrowers				Short Distance Borrowers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Connection #	0.136**				0.001			
	(2.09)				(0.04)			
Sibling Firms Connected		0.270*				0.031		
		(1.90)				(0.42)		
Other Relation Connected			0.272*				0.017	
			(1.69)				(0.26)	
Stock Holder Connected				0.633***				-0.047
				(5.32)				(-0.72)
Firm×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	949	949	949	949	14,651	14,651	14,651	14,651
Adj. R-squared	0.627	0.627	0.627	0.626	0.695	0.695	0.695	0.695

(*To be continued*)

Table 5: Firm Network Connections and Soft Information-*continued*

Panel B: Log(Months Earlier)

	Log(Months Earlier)							
	Long Distance Borrowers				Short Distance Borrowers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Connection #	0.336***				-0.025			
	(5.82)				(-0.57)			
Sibling Firms Connected		0.733***				-0.061		
		(4.96)				(-0.54)		
Other Relation Connected			0.791***				-0.068	
			(5.23)				(-0.57)	
Stock Holder Connected				-0.153				-0.028
				(-1.17)				(-0.24)
Firm×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	949	949	949	949	14,651	14,651	14,651	14,651
Adj. R-squared	0.771	0.771	0.772	0.766	0.801	0.801	0.801	0.801

Table 6: Interaction of Distance and Firm Network Connections

Table 6 reports the interaction effects of distance and firm network connections on early downgrades. Panel A reports the results of OLS regressions and Panel B reports the results of 2SLS regressions. We follow the standard approach in Wooldridge (2002) by using the predicted $Log(Distance)$ (i.e., $Log(\widehat{Distance})$) from regression of $Log(Distance)$ on the 2009 bank entry deregulation shock (i.e., equation (2)) and its interaction term with connection variables (e.g.,$Log(\widehat{Distance}) \times Connection\ \#$) as the instruments for both $Log(Distance)$ and $Log(Distance) \times Connection\ \#$. The sample consists of 19,051 delinquent loans from Jan 2007 to Jun 2013. There are two dependent variables that characterize the early downgrade activities, i.e., *Early Downgrade*, a dummy for whether this delinquent loan's internal rating is downgraded before the delinquency and *Log(Months Earlier)*, the logarithm of one plus the months between date of the first downgrade and the delinquent date. We include all main effects of $Log(Distance)$ and connection variables. The coefficients of constants and control variables are omitted for brevity. We control for firm×year fixed effects and bank fixed effects in all regressions. The t-statistics based on standard errors clustered at firm level are reported in parentheses. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively.

Panel A: OLS

	Early Downgrade				Log(Months Earlier)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Distance) ×Connection #	0.057**				0.104***			
	(2.29)				(2.83)			
Log(Distance) ×Sibling Firms Connected		0.111*				0.191**		
		(1.73)				(2.15)		
Log(Distance) ×Other Relation Connected			0.129**				0.271***	
			(2.29)				(3.16)	
Log(Distance) ×Stock Holder Connected				0.150*				0.224*
				(1.91)				(1.80)
Log(Distance)	-0.271***	-0.239***	-0.278***	-0.251***	-0.357***	-0.297***	-0.387***	-0.313***
	(-4.86)	(-4.50)	(-4.70)	(-4.70)	(-3.46)	(-3.02)	(-3.51)	(-3.13)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,605	15,605	15,605	15,605	15,605	15,605	15,605	15,605
Adj. R-squared	0.687	0.687	0.687	0.687	0.797	0.797	0.797	0.797

(*To be continued*)

Table 6: Interaction of Distance and Firm Network Connections -*continued*

Panel B: IV

	Early Downgrade				Log(Months Earlier)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Distance) ×Connection #	0.223***				0.507***			
	(3.99)				(5.40)			
Log(Distance) ×Sibling Firms Connected		0.705***				0.852***		
		(4.32)				(3.35)		
Log(Distance) ×Other Relation Connected			0.747***				1.693***	
			(4.12)				(5.13)	
Log(Distance) ×Stock Holder Connected				0.150				1.086***
				(0.92)				(3.95)
Log(Distance)	-0.343**	-0.346**	-0.401***	-0.347**	-0.223	-0.254	-0.329	-0.174
	(-2.11)	(-2.04)	(-2.94)	(-2.05)	(-0.69)	(-0.77)	(-1.17)	(-0.52)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,605	15,605	15,605	15,605	15,605	15,605	15,605	15,605
Wald F-stat	339.0	386.3	247.7	338.2	403.4	378.7	294.8	402.5

Table 7: Interaction of Distance and Firm Network Connections

Table 7 reports the interaction effects of distance and firm network connections on early downgrades for SMEs and large firms. Panel A reports the results of OLS regressions and Panel B reports the results of 2SLS regressions. We follow the standard approach in Wooldridge (2002) by using the predicted $Log(Distance)$ (i.e., $Log(\widehat{Distance})$) from regression of $Log(Distance)$ on the 2009 bank entry deregulation shock (i.e., equation (2)) and its interaction term with connection variables (e.g.,$Log(\widehat{Distance}) \times Connection\ \#$) as the instruments for both $Log(Distance)$ and $Log(Distance) \times Connection\ \#$. The sample consists of 19,051 delinquent loans from Jan 2007 to Jun 2013. Columns (1) to (4) report the subsample of SME borrowers. Columns (5) to (8) report the subsample of large borrowers. The dependent variables that characterize the early downgrade activities, i.e., *Early Downgrade*, a dummy for whether this delinquent loan's internal rating is downgraded before the delinquency. We include all main effects of $Log(Distance)$ and connection variables. The coefficients of constants and control variables are omitted for brevity. We control for Firm×Year fixed effects and Bank fixed effects in all regressions. The t-statistics based on standard errors clustered at firm level are reported in parentheses. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively.

Panel A: OLS

	Early Downgrade							
	SMEs				Large Firms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Distance) ×Connection #	0.051**				0.151			
	(2.02)				(0.70)			
Log(Distance) ×Sibling Firms Connected		0.111*				-1.024		
		(1.71)				(-1.41)		
Log(Distance) ×Other Relation Connected			0.113**				0.330	
			(1.99)				(1.41)	
Log(Distance) ×Stock Holder Connected				0.125				-1.146*
				(1.52)				(-1.75)
Log(Distance)	-0.282***	-0.256***	-0.280***	-0.265***	-0.036	0.070	-0.159	0.056
	(-4.55)	(-4.31)	(-4.40)	(-4.36)	(-0.15)	(0.52)	(-0.62)	(0.41)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,343	9,343	9,343	9,343	6,270	6,270	6,270	6,270
Adj. R-squared	0.649	0.649	0.649	0.649	0.744	0.744	0.744	0.745

(*To be continued*)

Table 7: Interaction of Distance and Firm Network Connections-*continued*

Panel B: IV

	Early Downgrade							
	SMEs				Large Firms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Distance) ×Connection #	0.182***				-1.638			
	(3.19)				(-0.95)			
Log(Distance) ×Sibling Firms Connected		0.707***				-15.401		
		(4.16)				(-0.71)		
Log(Distance) ×Other Relation Connected			0.505***				-8.9	
			(2.73)				(-0.87)	
Log(Distance) ×Stock Holder Connected				0.076				45.705
				(0.41)				(1.31)
Log(Distance)	-0.596***	-0.608***	-0.593***	-0.629***	5.246**	-13.586	13.120	3.324
	(-10.35)	(-10.44)	(-9.98)	(-10.73)	(2.16)	(-0.54)	(1.00)	(1.52)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,343	9,343	9,343	9,343	6,270	6,270	6,270	6,270
Wald F-stat	114.4	245.0	56.43	104.5	6.039	0.171	0.679	1.563

Table 8: Firm Network Connections and Loan Default

Table 8 reports the OLS regression of ex-post loan default on the firm network connections. The dependent variable is ex-post loan default dummy, indicating whether the loan is overdue 90 days or not. The sample consists of 5,563,124 loans from Jan 2007 to Jun 2013. *Sibling Firms Connected*, *Other Related Firms Connected*, and *Stock Holder Connected* are the dummies defined at individual loan level for whether the borrower has any sibling firms (i.e., firms with the same controlling parent company), firms with related party transactions, and shareholders that are located in the same city as the bank and also has been borrowing from this bank. *Connection#* is the total number of these three types of connections. We control for loan- level characteristics such as internal rating, guarantee status, loan size, and loan maturity, and firm-level characteristics such as firm size and leverage in all columns. The coefficients of constants and control variables are omitted for brevity. We control for firm×year fixed effects and bank fixed effects in all regressions. The *t*-statistics based on standard errors clustered at firm level are reported in parentheses. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Connection #	-0.001***			
	(-8.51)			
Sibling Firms Connected		-0.002***		
		(-8.08)		
Other Relation Connected			-0.001***	
			(-9.23)	
Stock Holder Connected				0.000
				(0.01)
Loan Controls	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes
Firm×Year FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
Observations	5,563,124	5,563,124	5,563,124	5,563,124
Adj. R-squared	0.267	0.267	0.267	0.267

Table A1: Variables' Definition and Construction

Variables	Definitions
Distance (100KM)	For each loan, we calculate the physical distances between the firm and all the branches of the issuing bank (approximately one billion observations) then choose the shortest one (i.e., the distance between the borrower and the closest bank branch). We divide it by 100 and take natural logarithm in regressions.
Bank_Num	The number of banks with loans outstanding at borrower-month level.
LoanAmount_Share	The share of loan amounts that have been borrowed from the specific bank in last 12 months at borrower-bank-month level.
LoanFreq_Share	The share of loan numbers that have been borrowed from the specific bank in last 12 months at borrower-bank-month level.
Recent_Lending	A dummy indicating whether the borrower gets any new loans from the bank in the last 12 months at borrower-bank-month level.
Loan Amount	The balance of each loan contract. The unit is in a million RMB.
Rating Dummy	The credit score placed by the loan officers in the bank. The larger the number, the worse the credit quality of the obligor. It takes the value of one if the rating is in the first category and zeroes otherwise.
Maturity	The term of each loan contract. The unit is months.
Guaranteed	A dummy variable that equals one if the bank requires third-party guarantee protections and zero otherwise.
Default	A loan performance measure that equals one if the loan is not repaid over three months after due date and zero otherwise.
Rating Change	A dummy variable indicating whether the delinquent loan has any rating changes in the loan lifecycle.
Early Downgrade	A dummy variable indicating whether the delinquent loan is downgraded before the delinquent date.
Connection #	The aggregate number of different types of connections, i.e. sibling firms, other related firms and stock holders, ranging from 0 to 3.
Sibling Firms Connected	A dummy variable indicating whether the borrower has any sibling firms (i.e., firms with the same controlling parent company) that has been borrowing from the same bank in the borrower's city.
Other Relation Connected	A dummy variable indicating whether the borrower has any other connected firms that has been borrowing from the same bank in the borrower's city, such as related party transactions, subsidiary firms, and so on.
Stock Holder Connected	A dummy variable indicating whether any of the borrower's stock holders are located in the same city as the lending bank branch and these stock holders have been borrowing from this bank.
Months Earlier	The number of months between the initial rating downgrade date and the delinquent date. We take natural logarithm in regressions.
Assets	The total assets of firms. The unit is in a million RMB. We take natural logarithm in regressions. For brevity, we report the assets in unit of billion RMB in Table 1 and Table 6.
Leverage	Book leverage, measured as the ratio of total liabilities over total assets.
SOE	A dummy variable based on borrower's registration type, indicating whether the borrower is state owned.
Employees	The amount of employment.
TFP	A measure of firm-level efficiency, i.e., total factor productivity.
ATR	Asset turnover ratio is defined as the total operating income divided by total assets.
After2009	A dummy variable that equals one if it is after the deregulation shock and zeroes otherwise.
Treatment	A dummy variable at city-bank level takes the value of one if the joint-equity bank has outstanding branches in this city or its capital city of the province before the bank expansion policy shock and zero otherwise.

Table A2: Literature Summary

This table summarizes the descriptive statistics of borrower-lender distance in the literature.

	Sample Country	Unit	Mean	Median	Std. Dev.	Min (P1)	Max (P99)
Herpfer, Mjos, and Schmidt (2018)	Norway	hours	0.58	0.42	0.48	0.03	3.00
Beck, Ongena, and Sendeniz-Yuncu (2018)	Turkey	kilometers	12.60	1.50	54.60	0.00	1286.40
Hollander and Verriest (2016)	U.S.	Ln(miles)	6.36	6.67	1.45	0.02	8.08
Bellucci, Borisov and Zazzaro (2013)	Italy	kilometers	5.02		7.25		
Dass and Massa (2011)	U.S.	miles	25.46	4.89	67.47		
Agarwal and Hauswald (2010)	U.S.	miles	9.91	2.62	21.44	0.00	3102.00
Norden and Weber (2010)	German	kilometers	7.21	0.00			
DeYoung, Glennon, and Nigro (2008)	U.S.	miles	49.97		237.62		
Berger, Miller, Petersen, Rajan, and Stein (2005)	U.S.	miles	26.05	3.00	136.99		
Degryse and Ongena (2005)	Belgian	driving minutes	6.90	4.29	7.30	0.00	51.00
Petersen and Rajan (2002)	U.S.	miles	42.50	4.00			

Table A3: Out-of-City Loans *versus* Within-City Loans

Table A3 reports the loan summary statistics of loan and borrower characteristics. Panel A is for the comparisons between out-of-city loans and within-city loans. Panel B is for the comparisons between distant loans (i.e., beyond 250km) and short-distance loans (i.e., within 250km). The number of loans, the means, medians, and standard deviations are reported. *Distance* is the physical distance between borrower and lending branch. *Assets* is the total amount of assets, in unit of a million RMB. *Leverage* is the total liabilities divided by total assets. *SOE* is a dummy indicating whether the borrower is state-owned based on registration type. *Employees* is the number of employees. ATR is the asset turnover ratio defined as the total operating income divided by total assets. TFP is the total factor productivity.

Panel A: Out-of-City vs. Within City

	Outside City Loan				Inside City Loan			
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.
Assets (Billion RMB)	826,046	4.812	1.024	9.538	5,998,329	3.950	0.777	9.332
Leverage	826,046	0.661	0.670	0.191	5,998,285	0.606	0.604	0.185
SOE	587,307	0.073	0.000	0.261	5,026,157	0.108	0.000	0.310
Employees	587,307	2444.62	565.000	4941.506	5,026,157	1699.78	366.000	3851.108
ATR	543,281	2.672	2.557	2.513	4,665,637	2.541	2.474	2.323
TFP	543,281	2.526	1.353	3.446	4,665,637	1.983	1.228	2.724

Panel B: >250 km vs. <=250km

	>250 km				<=250 km			
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.
Assets (Billion RMB)	571,483	3.689	0.527	8.560	6,252,892	4.088	0.824	9.431
Leverage	571,483	0.657	0.660	0.200	6,252,848	0.609	0.607	0.185
SOE	405,985	0.076	0.000	0.265	5,207,479	0.106	0.000	0.308
Employees	405,985	1768.38	375.000	4268.719	5,207,479	1778.43	374.000	3962.824
ATR	376,296	2.842	2.767	2.372	4,832,622	2.533	2.465	2.341
TFP	376,296	2.769	1.355	3.642	4,832,622	1.983	1.229	2.729

Table A4: Bank Entry Deregulation Impact on Bank-Borrower Distance

Table A4 reports the OLS regressions of bank-borrower distance on the bank entry deregulation in April 2009. The sample consists of 7,623,883 firm-bank-month observations and columns (1) to (5) presents the results based on different subsamples with different event windows. Column (1) presents the results of 6-month window subsample estimations, Column (2) presents the results of 12-month window subsample estimations, Column (3) presents the results of 24-month window subsample estimations, Column (4) presents the results of 48-month window subsample estimation, and Column (5) presents the results using overall sample. The dependent variable is the *LogDistance*, the natural logarithm of one plus the distance between banks and borrowers. The main independent variable is the interaction, *Shock2009=Treatment*After2009*, where *After2009* equals one for observations after the policy shock in Apr, 2009 and zero before and *Treatment* equals one for treated bank-cities and zero for controlled bank-cities. According to the policy, an eligible Joint-equity bank k in city j free of regulation on new-branch entry is a bank that have outstanding branches in this city or in the capital city of the province that the city j is located in prior to the bank expansion policy shock. For the biggest five state-owned banks, i.e. Industrial and Commercial Bank of China (ICBC), China Construction Bank (CCB), Bank of China (BOC), Agricultural Bank of China (ABC) and Bank of Communications (BOCOM), *Treatment* always equals zero. All variables used in the regression are defined in Table A1. The constants are omitted for brevity. We control for firm×year fixed effects and bank fixed effects in all regressions. The *t*-statistics based on standard errors clustered at firm level are reported in parentheses. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively.

	Log (Distance)				
	(1)	(2)	(3)	(4)	(5)
	Window 6M	Window 12M	Window 24M	Window 48M	All
Shock2009	-0.003***	-0.003**	-0.005***	-0.010***	-0.014***
	(-2.59)	(-2.09)	(-2.73)	(-4.33)	(-5.12)
Treatment	-1.060***	-1.041***	-0.974***	-0.848***	-0.590***
	(-21.33)	(-23.26)	(-26.47)	(-31.07)	(-34.22)
After2009	-0.001**	-0.002***	-0.005***	-0.010***	-0.016***
	(-2.27)	(-4.31)	(-7.55)	(-12.06)	(-17.13)
Firm×Year FE	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes
Observations	456,500	909,857	1,855,067	3,821,430	7,623,859
Adj. R-squared	0.848	0.838	0.828	0.815	0.801

Table A5: Interaction of Distance and Firm Network Connections

Table A5 reports the interaction effects of distance and firm network connections on months of early downgrades for SMEs and large firms by 2SLS regressions. We follow the standard approach in Wooldridge (2002) by using the predicted $Log(Distance)$ (i.e., $Log(\widehat{Distance})$) from regression of $Log(Distance)$ on the 2009 bank entry deregulation shock (i.e., equation (2)) and its interaction term with connection variables (e.g.,$Log(\widehat{Distance}) \times Connection\ \#$) as the instruments for both $Log(Distance)$ and $Log(Distance) \times Connection\ \#$. The sample consists of 19,051 delinquent loans from Jan 2007 to Jun 2013. Columns (1) to (4) report the subsample of SME borrowers (i.e., the total assets is lower than 300 million RMB). Columns (5) to (8) report the subsample of large borrowers (i.e., the total assets is lower than 300 million RMB). The dependent variables that characterize the early downgrade activities, i.e., *Log(Months Earlier)*, the logarithm of one plus the months between date of the first downgrade and the delinquent date. We include all main effects of $Log(Distance)$ and connection variables. The coefficients of constants and control variables are omitted for brevity. We control for firm×year fixed effects and bank fixed effects in all regressions. The *t*-statistics based on standard errors clustered at firm level are reported in parentheses. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively.

	Log(Months Earlier)							
	SMEs				Large Firms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Distance)*Connection #	0.404***				-2.045			
	(4.25)				(-0.76)			
Log(Distance)*Sibling Firms Connected		0.756***				-19.113		
		(2.78)				(-0.76)		
Log(Distance)*Other Relation Connected			1.269***				-13.306	
			(3.75)				(-0.86)	
Log(Distance)*Stock Holder Connected				0.915***				92.601
				(2.94)				(1.53)
Log(Distance)	-0.688***	-0.743***	-0.672***	-0.682***	8.238**	-15.238	19.810	4.210
	(-4.09)	(-4.51)	(-3.87)	(-4.08)	(2.14)	(-0.52)	(1.00)	(1.05)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,343	9,343	9,343	9,343	6,270	6,270	6,270	6,270
Wald F-stat	304.0	331.8	180.0	293.7	20.73	1.140	2.926	1.864

Table A6: Explicit Guarantee and Firm Network Effect

Table A6 reports the results of OLS regression estimates on relations between firm network connection and early downgrades. In Panel A, the dependent variable is *Early Downgrade*, a dummy for whether this delinquent loan's internal rating is downgraded before the delinquency. In Panel B, the dependent variable is *Log(Months Earlier)*, the logarithm of one plus the months between date of the first downgrade and the delinquent date. *Sibling Firms Connected*, *Other Related Firms Connected*, and *Stock Holder Connected* are the dummies defined at individual loan level for whether the borrower has any sibling firms (i.e., firms with the same controlling parent company), firms with related party transactions, and shareholders that are located in the same city as the bank and also has been borrowing from this bank. *Connection#* is the total number of these three types of connections. Columns (1) to (4) report the subsample of loans without third party guarantees. Columns (5) to (8) report the subsample of loans with third party guarantees. The coefficients of constants and control variables are omitted for brevity. We control for firm×year fixed effects and bank fixed effects in all regressions. The *t*-statistics based on standard errors clustered at firm level are reported in parentheses. *, **, and *** indicate the significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Early Downgrade

	Early Downgrade							
	Without Guarantee				With Guarantee			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Connection #	0.055***				-0.043*			
	(2.71)				(-1.85)			
Sibling Firms Connected		0.115*				-0.040		
		(1.73)				(-0.60)		
Other Relation Connected			0.140***				-0.107**	
			(3.73)				(-2.10)	
Stock Holder Connected				0.061				-0.112*
				(1.04)				(-1.66)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,954	8,954	8,954	8,954	6,249	6,249	6,249	6,249
Adj. R-squared	0.784	0.784	0.784	0.783	0.658	0.658	0.659	0.658

(*To be continued*)

Table A6: Explicit Guarantee and Firm Network Effect-*continued*

Panel B: Log(Months Earlier)

	Log(Months Earlier)							
	Without Guarantee				With Guarantee			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Connection #	0.065*				-0.106***			
	(1.84)				(-2.81)			
Sibling Firms Connected		0.155				-0.254**		
		(1.45)				(-2.53)		
Other Relation Connected			0.169***				-0.274***	
			(2.81)				(-3.05)	
Stock Holder Connected				0.051				-0.114
				(0.48)				(-1.01)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,954	8,954	8,954	8,954	6,249	6,249	6,249	6,249
Adj. R-squared	0.865	0.865	0.865	0.865	0.800	0.800	0.800	0.799