IMF Working Paper

Fintech Credit Risk Assessment for SMEs: Evidence from China

by Yiping Huang, Longmei Zhang, Zhenhua Li,
Han Qiu, Tao Sun, and Xue Wang

INTERNATIONAL MONETARY FUND

IMF Working Paper

Asia Pacific Department

Fintech Credit Risk Assessment for SMEs: Evidence from China

Prepared by Yiping Huang, Longmei Zhang, Zhenhua Li,
Han Qiu, Tao Sun, and Xue Wang[1,2]

Authorized for distribution by Helge Berger

September 2020

Abstract

Promoting credit services to small and medium-size enterprises (SMEs) has been a perennial challenge for policy makers globally due to high information costs. Recent fintech developments may be able to mitigate this problem. By leveraging big data or digital footprints on existing platforms, some big technology (BigTech) firms have extended short-term loans to millions of small firms. By analyzing 1.8 million loan transactions of a leading Chinese online bank, this paper compares the fintech approach to assessing credit risk using big data and machine learning models with the bank approach using traditional financial data and scorecard models. The study shows that the fintech approach yields better prediction of loan defaults during normal times and periods of large exogenous shocks, reflecting information and modeling advantages. BigTech's

proprietary information can complement or, where necessary, substitute credit history in risk assessment, allowing unbanked firms to borrow. Furthermore, the fintech approach benefits SMEs that are smaller and in smaller cities, hence complementing the role of banks by reaching underserved customers. With more effective and balanced policy support, BigTech lenders could help promote financial inclusion worldwide.

JEL Classification Numbers: G17, G18, G23, G32

Keywords: Fintech lending, Credit risk assessment, Big data, Machine learning

Authors' e-mail addresses: yhuang@nsd.pku.edu.cn; lzhang2@imf.org; sunny.lzh@antfin.com; qiuh@pku.edu.cn; tsun@imf.org; xue.wang@pku.edu.cn

Table of Contents

I. INTRODUCTION

Promoting financial inclusion for vulnerable households and smaller firms has been a perennial challenge for policy makers globally (Abdulsaleh and Worthington 2013; Freel et al. 2012; Vos et al. 2007). An essential element in financial inclusion, access to credit for small and medium-size enterprises (SMEs) remains quite limited, especially in developing countries (Demirgüç -Kunt et al. 2018). The main barriers include high cost, physical distance, and lack of proper documentation (Agarwal and Hauswald 2010; Demirgüç-Kunt and Klapper 2013), which deter banks from managing the risks in servicing SMEs. Since SMEs are numerous and scattered across locations, commercial banks face a high fixed cost in establishing business connections. Moreover, many SMEs lack high-quality financial data and collateral assets for banks to identify and manage credit risk. As a result, banks typically resort to personal guarantees or relationship lending, relying on local connections and soft information to reduce information asymmetry (Berger and Udell 2002, 2006). Nonetheless, relationship building is also costly, and it is difficult to reach a large business scale, as evidenced by the encouraging but still limited coverage of lending services by Muhammad Yunus's Grameen Bank.[3]

In recent years, the surge of fintech has opened new possibilities for expanding access to credit for SMEs. Some big technology (BigTech) companies, such as Alibaba and Tencent in China, Mercado Credito in Argentina, Paytm in India, and Amazon Lending in the United States, have extended loans to millions of small borrowers (Frost et al. 2019; Agarwal et al. 2019). In China, each of the three leading virtual banks—MYbank (affiliated to Alibaba), WeBank (affiliated to Tencent), and XW Bank (affiliated to tech giant Xiaomi)—provides loans to millions of small firms annually, more than 80 percent of which have no credit history. Compared with traditional banks, the loans provided by BigTech lenders are much smaller, shorter in duration, and mainly used for operational rather than long-term investment purposes. Hence, fintech lending has so far played a complementary role to traditional banking by reaching underserved customers.

The wide reach of BigTech lending to SMEs reflects a confluence of factors:

- First, digital technologies enable BigTech companies to use e-commerce, social networks, and mobile payment services to connect to millions of customers at very low marginal costs. This could help overcome barriers to large-scale customer acquisition. In addition, BigTech companies can monitor borrowers' activities, most of which occur on the BigTech platforms.

---

[3] Grameen Bank is a microfinance organization and community development bank founded in Bangladesh. It makes small loans to impoverished individuals without requiring collateral. Grameen Bank originated in 1976, through the work of Professor Muhammad Yunus, who launched a research project to study how to design a credit delivery system to provide banking services to the rural poor. As of November 2019, it had 9.6 million members, 97 percent of whom were women. With 2,568 branches, Grameen Bank provides services in 81,678 villages, covering more than 93 percent of the total villages in Bangladesh.
(http://www.grameen.com/introduction/Grameen)

- Second, big data, including traditional and proprietary information, could help in credit risk assessment, in the absence of financial history and collateral assets (Holmström 2018). Traditional data include basic information about individuals or firms, such as gender, age, location, profession, and business. Such information is mostly available offline, but it is costly to collect without the support of BigTech. Proprietary information refers to broadly defined digital footprints, such as messages exchanged, payments made, and websites browsed on the BigTech platforms. Such information could be deployed to assess borrowers' financial conditions and behavioral characteristics (Berg et al. 2019; Gambacorta et al. 2019; Jagtiani and Lemieux 2019). To some extent, these data are similar to soft information in relationship lending (Cornée 2019).

- Third, digital technologies, such as cloud computing and artificial intelligence, allow BigTech companies to process massive numbers of loan applications quickly and update risk assessment dynamically, based on real-time data. Furthermore, the big data and machine learning approach enables BigTech lenders to restructure loans quickly at large volumes, which significantly reduces operating costs.[4] The "contact-free feature" of the business process, from customer acquisition to loan underwriting and restructuring, makes the BigTech lending model particularly robust during the COVID-19 pandemic.

Despite the achievements of BigTech lending so far, a central question that remains is whether its credit risk assessment approach, that is, big data plus machine learning, is more reliable in predicting loan defaults than the traditional method, which features financial data and scorecard models. In addition, given the short history of BigTech lending, it is yet to be tested whether such an approach can withstand business cycles and various economic shocks. Moreover, it is critical to understand the contribution of data versus methodology in credit risk assessment. Can traditional models also fully utilize the information value of big data? Can artificial intelligence (AI) algorithms extract more information from traditional data? In particular, for borrowers with no bank credit history and incomplete financial data, is BigTech's proprietary information sufficient to substitute credit registry information in risk assessment?

To shed light on these questions, we employ a unique data set of 1.8 million MYbank SME loans to replicate and compare the fintech approach with the traditional banking approach in risk assessment. A leading virtual bank in China, MYbank was established in May 2015, with its BigTech lending business inherited from its parent company, Ant Group ("Ant" hereafter), which, in turn, is an affiliated company of the e-commerce giant Alibaba Group ("Alibaba" hereafter). To our knowledge, Alibaba and Ant were the first groups of companies globally to create the BigTech lending business

---

[4] Loan forbearance and restructuring are resource-intensive and time consuming for traditional banks. Reflecting the small size of SME loans, the unit costs of providing financing services to SMEs are typically high, which deters traditional banks from reaching out proactively to SMEs.

model in 2010.[5] Borrowers in this data set are all online vendors on Alibaba's e-commerce platforms and Ant's Alipay users, mostly small retailers and self-employed. The data set features a large volume of proprietary information, such as business transactions, payments, customer ratings, consumption patterns, and importance in the ecosystem, along with other traditional data. The data set also contains information on borrowers' credit histories with traditional banks, if they have ever borrowed in the past.

Our study shows that the fintech approach of combining big data and machine learning significantly improves the accuracy of loan default prediction, compared with the traditional approach. This reflects a combination of information and modeling advantages. BigTech's proprietary information is utilized by AI algorithms to reveal the financial condition (i.e., capacity to repay) and behavioral characteristics (i.e., willingness to repay) of the borrower. Applying the traditional scorecard model to big data could also provide reasonably good risk assessment, yet some borrower characteristics would not be fully captured. Similarly, the AI approach improves traditional information's predictive power, reflecting its capacity to model more complex interactions among the variables. For borrowers without a bank credit history, we find that BigTech proprietary information can effectively substitute credit registry information in risk assessment, hence enhancing financial inclusion. Finally, this study shows that the fintech risk assessment model is robust when BigTech companies face exogenous policy shocks, and its advantages are more pronounced for firms that are smaller and in smaller cities. Hence, the model naturally complements traditional banks in broadening firms' access to credit.

The remainder of the paper is structured as follows. Section II reviews the literature and outlines the key analytical steps of the study. Section III briefly introduces the virtual bank MYbank, from which we obtained the data set, and presents the stylized facts of its lending business. Section IV describes the data set, explains the analytical strategy, and discusses the empirical results. Section V summarizes the main findings and draws policy implications for China and other countries.

II. LITERATURE REVIEW

The most fundamental challenge for financial transactions, including lending, is information asymmetry, which may cause ex-ante adverse selection and ex-post moral hazard problems. In a way, financial institutions (e.g., banks), market mechanisms (e.g., rating agencies), and regulatory policies (e.g., information disclosure requirements) are all devised to deal with the information asymmetry problem. Commercial banks normally adopt three methods to analyze and mitigate credit risk: *financial history*, *collateral assets*, and *soft information*. Financial history relies on detailed analyses of borrowers' financial data, especially balance sheets, income statements, and cash flow information, to predict probabilities of default. This method is often most useful for large corporations, which can typically provide comprehensive information. Collateral or guarantees are more frequently used in

---

[5] Appendix A provides more information on MYbank.

lending to SMEs. Several studies find that pledging collateral may help resolve adverse selection phenomena (Besanko and Thakor 1987a, 1987b; Stiglitz and Weiss 1981; Cerqueiro, Ongena, and Roszbach 2016) and moral hazard actions (Berger, Frame, and Ioannidou 2016). However, the fact that many SMEs do not have collateral assets constrains lending to SMEs, especially micro firms and self-employed businesses. Therefore, an alternative model of relationship banking uses soft information about the firm (such as information on its owner and the local community) to predict the probability of default (Berger and Udell 2002, 2006; Agarwal and Hauswald 2010). This approach can reduce banks' dependence on financial data and collateral assets in credit risk management, but it requires significant investment in physical capital and human resources. Banks need to have broad networks to be close to their potential customers and maintain regular interactions. Therefore, the costs of relationship lending may be high, making it difficult for SMEs to cover these high costs to access financing.

Recent developments in BigTech lending reflect advances in credit risk assessment by employing big data and machine learning models. The fintech approach for credit risk assessment is still a new but growing field of research. A few papers analyze the role of digital footprints in enhancing credit risk assessment. Using a customer data set from a German online store, Berg et al. (2019) demonstrate that the easily accessible variables from the digital footprints customers leave on the e-commerce platform complement credit bureau information, affect access to credit offered by the online shop, and reduce default rates. The authors show that even simple information on mobile phone operating systems, iOS or Android, reveals a lot about customers' creditworthiness. Berg et al. (2019) speculate that by carefully analyzing digital footprints, financial services might be able to cover the world's 200 million unbanked individuals.

Using unique and proprietary loan-level data from a large fintech lending firm in India, Agarwal et al. (2019) find that mobile and social footprints have significantly more predictive power for loan approvals and defaults, outperforming the traditional credit scores used by banks. The authors also show that this new approach can expand credit as well as reduce the overall default rate. Similarly, based on a case study in Argentina, Frost et al. (2019) provide evidence that BigTech lenders have an information advantage in credit risk assessment relative to a traditional credit bureau. By analyzing a transaction data set from Lending Club, Jagtiani and Lemieux (2019) prove that the use of alternative information sources allows some borrowers classified as subprime by traditional criteria to be slotted into "better" loan grades and therefore obtain lower-priced credit.

In addition to its use of big data, another salient feature of fintech risk assessment is the machine learning approach, in contrast to traditional models. Financial institutions have established many traditional methods for decision making, and credit scoring is the most widely used technique (Kithinji 2010; Ahmed and Rajaleximi 2019). The retail banking business, such as credit cards, mortgages, and personal loans, makes heavy use of predictive statistical models called scorecards (Thomas, Oliver, and Hand 2005), which are built using data from past customers and credit histories. The same method can be applied to corporate risk models for SMEs (Miller and Rojas 2004). The

purpose is for banks to create a scorecard with as many borrower characteristics as possible, to enable better decision making (Siddiqi 2012, 2017). However, the scorecard models have several obvious weaknesses (Hand and Crowder 2005). Many researchers argue that they require a training period of at least one year. Further, updating the scorecard is time consuming and resource intensive (Hopper and Lewis 2002). In contrast, fintech lenders often automate their credit assessment process by continuously training machine learning models with updated big data. Frost et al. (2019) show that BigTech lenders process mortgage applications 20 percent faster than other lenders, without incurring higher default cost.

The machine learning models adopted by BigTech lenders capture interactions among various explanatory variables. The models turn various digital footprints into data, ranging from social network activities to the physical locations of applicants' activities (Bazarbash 2019; Sirignano, Sadhwani, and Giesecke 2018). Khandani, Kim, and Lo (2010) find that machine learning forecasts are considerably more adaptive and capture the dynamics of credit cycles as well as absolute levels of default rates. Butaru et al. (2016) provide evidence that decision trees and random forests, the two most popular machine learning methods, outperform logistic regression in out-of-sample and out-of-time forecasts of credit card delinquencies. Analyzing data from a Chinese fintech firm, Gambacorta et al. (2019) show that a model based on machine learning and nontraditional data is better at predicting losses and defaults, compared with traditional models, by capturing nonlinearities and including additional information. They also find that this conclusion is robust to an exogenous shock to the supply of credit. Fuster et al. (2020) find that new technologies such as machine learning could help increase predictive accuracy from the improved use of information.

Our study contributes to the literature in the following respects. First, to our knowledge, this is the first study to carry out a horse race analysis between the fintech approach and the traditional approach for credit risk assessment. By controlling the data and models, we quantify the information and model advantages. Second, this study is the first to analyze a virtual bank's lending decisions by focusing on big data. In contrast, previous studies only looked at the summary indicator of the credit score, without analyzing the underlying data. Third, we examine the key question of whether fintech risk assessment remains robust when faced with a large policy shock.

III. STYLIZED FACTS

Compared with traditional banks, MYbank mainly offers online loans to micro firms that have no access to credit. The loans are small, short in duration, and used primarily for operational purposes. The loan quota is linked to a firm's cash flows more than its assets, and the historical average nonperforming loan (NPL) ratio has been about 2 percent.

MYbank loans are granted with a "contact-free feature." MYbank's lending business is fully digitalized, from customer acquisition to loan underwriting and restructuring. It operates on the so-called 3-1-0 model, which promises user registration and application within 3 minutes, money

transferred to an Alipay account within 1 second, and 0 human intervention. The contactless feature has dramatically reduced operating costs and ensured business continuity during the pandemic.

MYbank mainly serves customers with no access to traditional bank lending. Since its establishment in mid-2015, MYbank has provided loans to more than 20 million SMEs, about 80 percent of which are micro firms with annual sales of less than RMB 1 million and no access to bank lending.[6] In our sample of 1.8 million loans, 99.8 percent of the borrowers are micro firms (Figure 1), among which only 7.5 percent have borrowed from traditional banks. Hence, MYbank has played a complementary role to traditional banks, mainly serving unbanked customers in the left tail of the distribution.

---

[6] https://www.caixinglobal.com/2020-04-28/alibaba-backed-online-bank-lends-to-over-20-million-small-and-micro-firms-101548039.html.

Figure 1: Distribution of Firm Size
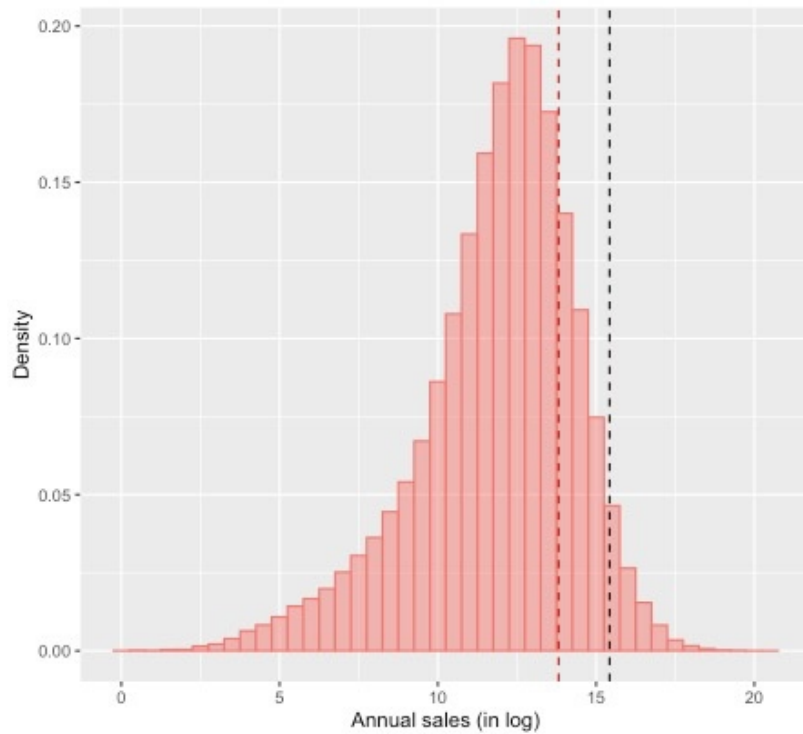
Source: Calculations using data from MYbank.
Note: The graph shows the distribution of firm size proxied by the logarithm of annual sales, since most small and medium-size enterprises have no information on assets. The red dashed line marks the threshold (RMB 1 million) for micro firms, and the black dashed line marks the threshold (RMB 5 million) for small firms. The measures of firm size in the retail sector are from the Classification of Large, Medium, Small and Miniature Enterprises for Statistics (2017), China's National Bureau of Statistics.

MYbank's loans are smaller and shorter in duration compared with traditional banks' loans. In our sample, MYbank's average loan size is RMB 2,600 ($367) (Figure 2), while the average loan size for SMEs from banks is RMB 1 million ($150,000), more than 30 times larger than that of MYbank. The duration of MYbank loans is much shorter. For instance, the duration of over 60 percent of MYbank's loans is less than one month, compared with nearly 57 percent of traditional bank loans for SMEs being longer than one year (inferred from MYbank borrowers who have also borrowed from traditional banks) (Figure 3). Reflecting the short duration, these loans are often used as working capital for operational purposes rather than longer-term investment.
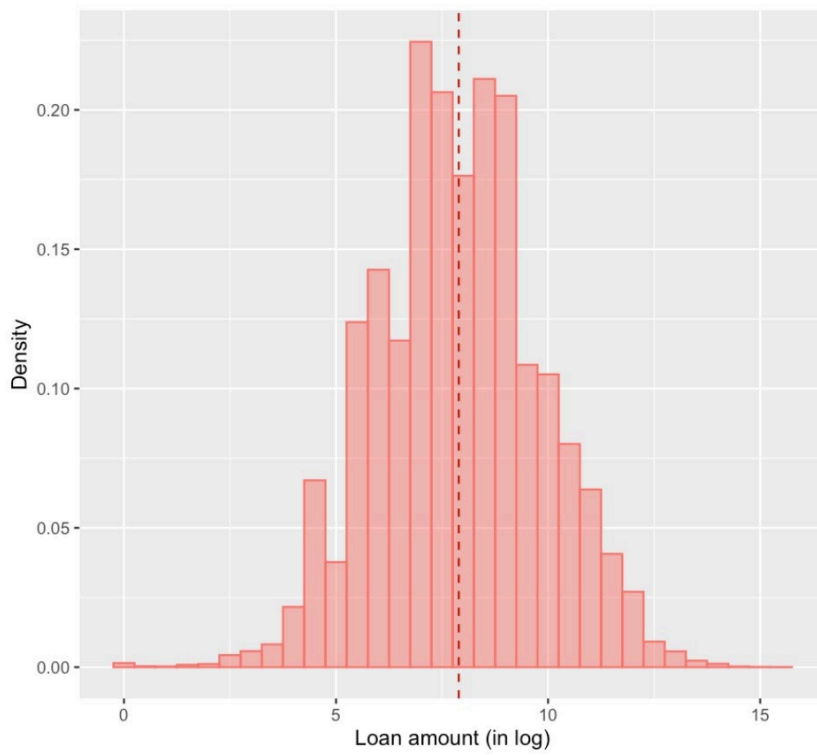
Figure 2: Distribution of MYbank Loans

Source: Calculations using data from MYbank.
Note: The graph shows the distribution of the logarithm of MYbank's loan amounts. The red dashed line marks MYbank's average loan size (RMB 2,600, equivalent to $380).

Figure 3: Distribution of Loan Duration: MYbank versus Traditional Banks
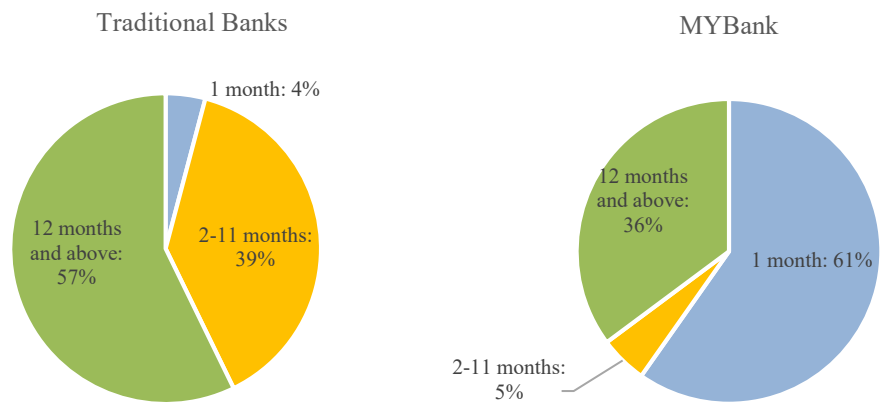
Source: Calculations using data from MYbank.

MYbank loans are more closely associated with cash flows, compared with traditional bank loans to SMEs in the same sample, which are more associated with firm assets. Simple correlations show that MYbank loans are more sensitive to firms' cash flows, with a 1 percent increase in cash flow being associated with a 0.45 percent increase in the loan amount, compared with 0.40 percent for traditional bank loans. In addition, about 30 percent of the loan quota variation can be explained by cash flows, compared with around 10 percent for traditional bank loans (Figure 4). In contrast, traditional banks rely more on collateral assets in lending decisions (Figure 5). Housing ownership significantly increases borrowers' loan quota from traditional banks, while there is no significant impact on MYbank loans.

Figure 4: Loan Size versus Cash Flow

a. MYbank loans   b. Traditional bank loans
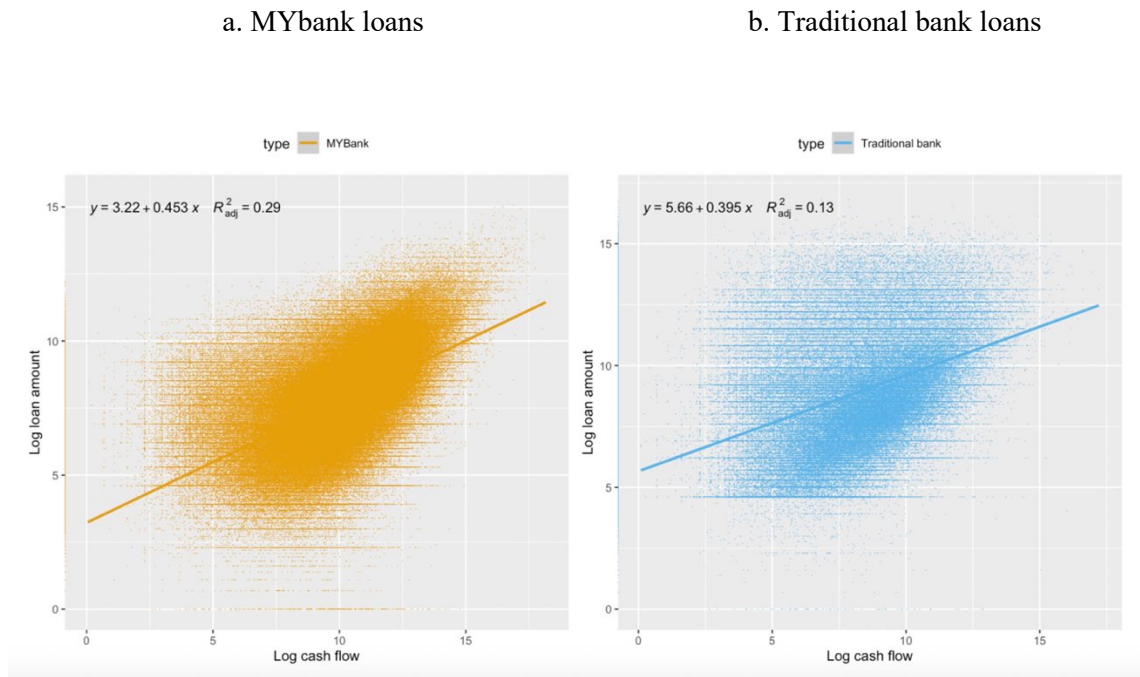
Source: Calculations using data from MYbank.
Note: The graphs show the relationship between loan amounts and cash flows. The dots in the figures indicate the log of loan amount (y-axis) and the log of cash flow (x-axis) for each borrower. Linear trend lines are reported in both graphs.

Figure 5: Loan Amounts and Housing Assets

a. MYbank loans   b. Traditional bank loans
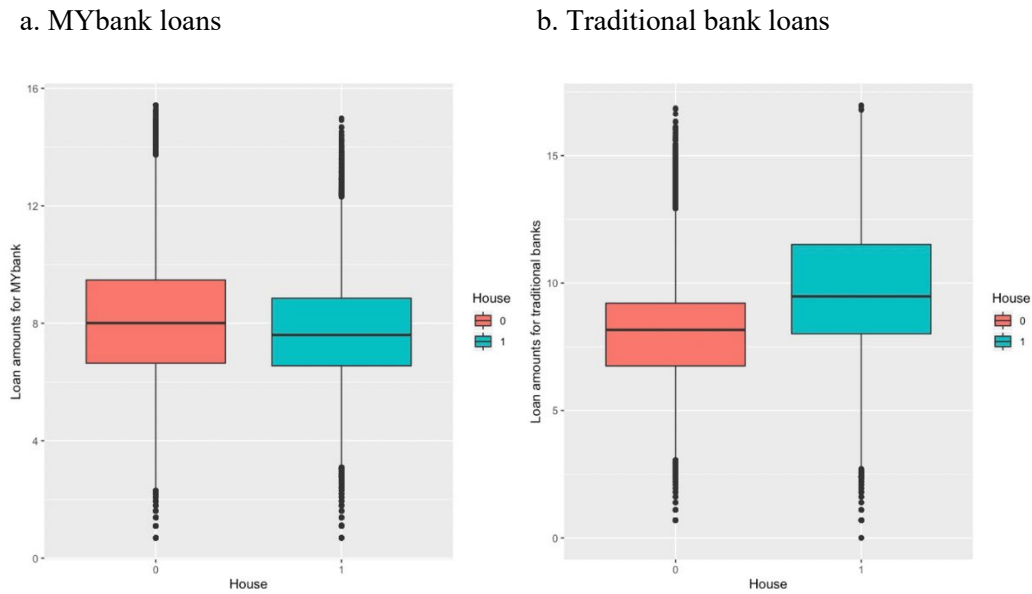
Source: Calculations using data from MYbank.
Note: In the boxplots, the lowest point is the minimum value, and the highest point is the maximum. The box is drawn from the first quartile (25th percentile) to the third quartile (75th percentile), with a horizontal line denoting the median. Red boxes (house ownership equals 0) plot the loan amounts of borrowers who do not own a house, and green boxes (house ownership equals 1) plot the loan amounts of borrowers who own a house.

MYbank's average NPL ratios have been much lower than those of traditional banks, although SME loans typically have a higher default rate than other loan portfolios. In 2018, the share of NPLs of Chinese banks' SME loans was 3.2 percent, and 5.5 percent for SME loans that were less than RMB 5 million,[7] compared with the overall NPL ratio of 1.9 percent. In contrast, MYbank has kept its average NPL ratio at around 1 percent, which to some extent may reflect the smaller size and shorter duration of its loans. Since the onset of the COVID-19 pandemic, SMEs have been severely affected amid stringent containment measures. As a result, MYbank's average NPL ratio has increased, but it has remained contained at below 2 percent. MYbank also has unique means of addressing moral hazard problems, such as downgrading business ratings in the e-commerce ecosystem. In contrast, traditional banks must resort to lengthy bankruptcy procedures or dispose NPLs to asset management companies.

The financial inclusion feature of MYbank lending is reflected mainly in credit access, rather than price. MYbank's annualized lending rate is between 10 and 17 percent, similar to the prevailing private lending rates in China, such as the Wenzhou composite lending rate, but it is higher than the

---

[7] China Small and Micro Enterprises Financial Services Report 2018, accessed June 5, 2020, http://www.gov.cn/xinwen/2019-06/25/5402948/files/f59aaafc00da4c848a322ac89fdec1e5.pdf.

average bank lending rate of 4.35 percent.[8] The difference reflects several factors. First, SME lending by traditional banks often enjoys preferential regulatory policies, such as cheap central bank funding and explicit government subsidies. Lending rates are also set low under government guidance, which does not necessarily reflect the risk premium. In contrast, MYbank's funding cost is higher, given its disadvantage in attracting retail deposits and ineligibility for preferential policies. Establishing a big data infrastructure also requires high fixed costs. Lending rates are set to ensure an adequate profit margin. Second, MYbank borrowers are smaller than typical SMEs with bank relationships and are therefore associated with higher risks and typically excluded by banks.

IV. EMPIRICAL STUDY

A. Data Description

This study employs a unique data set of 1.8 million SME loans granted by MYbank between March and August in 2017.[9] All the loans have maturity of one year, and 99.8 percent of the borrowers are micro firms, that is, firms with annual sales less than RMB 1 million. As shown in Figure 6, these firms are widely distributed across China, and about half of them are located in less developed cities, that is, Tier-3 and Tier-4 cities. All the borrowers are online vendors on Alibaba's e-commerce platforms and users of Ant's Alipay, which helped generate large volumes of proprietary data on businesses, financing, social networks, and individual behaviors. The data set also includes information on loan repayment through August 2018.

---

[8] Lending rates for MYbank are from the official website: https://render.mybank.cn/p/f/fd-j9fi9ern/index.html. The composite annual lending rate in the Wenzhou informal market was 15.66 percent in June 2020 (see http://www.wzpfi.gov.cn/).

[9] Due to business confidentiality, this data set is not available to the public for cross-checking.

Figure 6: Geographic Distribution of Borrowers

a. Distribution by province
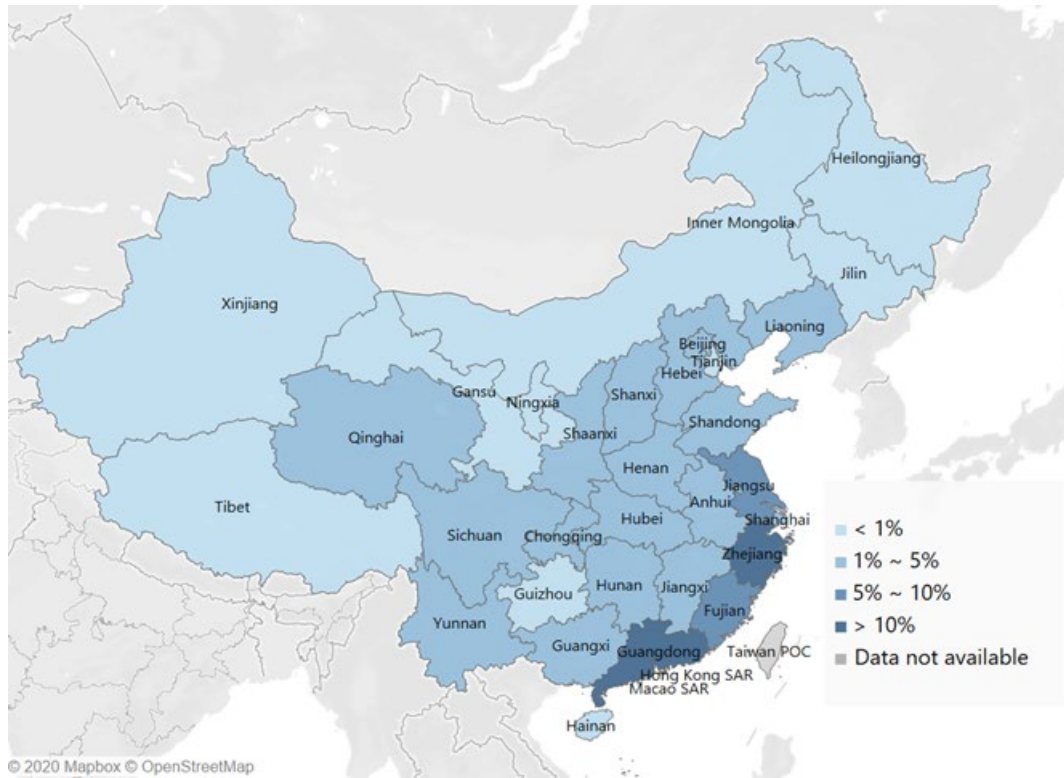
b. Distribution by city tier
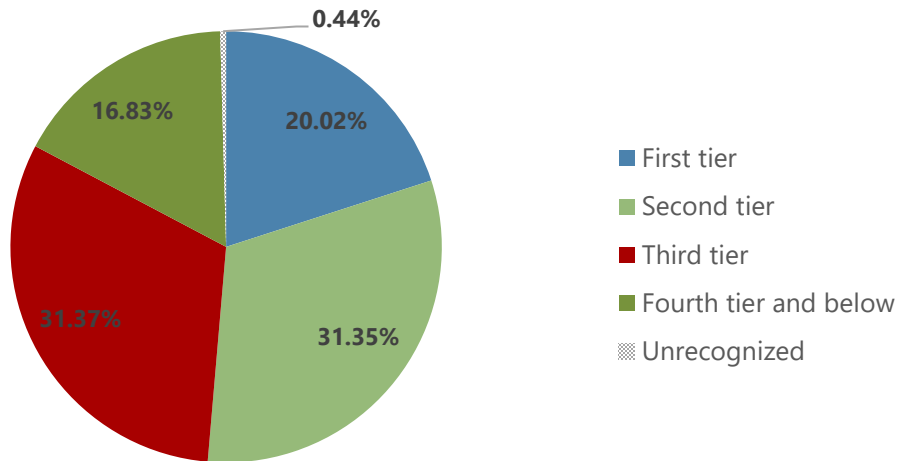
Source: Calculations using data from MYbank.
Note: In panel a, ranges of values represent the share of borrowers in each province (i.e. number of borrowers in each province divided by total borrowers at MYbank). Ranges are not provided for Hong Kong, Macao, and Taiwan due to missing data. In panel b, the percentages are the shares of borrowers in each city tier.

We divide the 76 variables on firm characteristics used in this study into two broad categories: *traditional data* and *proprietary information*, and then further divide them into subcategories. The 32 traditional data variables are classified into four groups: (a) asset-related information, such as housing property; (b) credit history with MYbank; (c) vendor-specific information, such as gender, age, and business; and (d) information on the local (provincial and municipal) economy. The 44 proprietary information variables include transaction volume, network effect score, and other digital footprints. In particular, the network effect score measures a borrower's relative importance in the BigTech network based on their fund flows and social interactions, with 0 representing the lowest impact and 100 the highest impact. Digital footprints also reflect information on borrowers' behavior, such as their online consumption pattern on Alibaba's platform and financial transaction style on Ant's platform. In addition, a firm's real-time customer ratings and reviews are captured. Table 1 provides summary statistics for the key proprietary information variables.

Table 1: Descriptive Statistics of the Key Variables

Variable	No.	Mean	St. dev.	Pctl (25)	Median	Pctl (75)	Type of information
House property (0/1)	1,822,423	0.621	0.485	0	1	1	Traditional
Car property (0/1)	1,822,423	0.662	0.473	0	1	1	Traditional
Number of credit cards	1,825,342	2.968	3.838	0	2	4	Traditional
Owner's age (years)	1,825,342	29	6	25	28	31	Traditional
Gender (male = 1; female = 0)	1,825,342	0.65	0.48	0	1	1	Traditional
Firm's age (years)	1,825,342	4.48	2.24	2.70	4.07	5.99	Traditional
City tiers	1,817,380	2.468	1.027	2	2	3	Traditional
Total inflow of funds in Alipay (yuan)	1,825,342	516,708.00	1,149,718.00	138,534.10	264,210.00	534,783.60	Proprietary
Shop rating	1,825,342	4.19	2.96	2	4	6	Proprietary
Transaction volume (last six months, yuan)	1,825,342	9,196.80	58,973.99	1,964.35	4,754.43	9,954.34	Proprietary
Log-ins (number, last six months)	1,825,342	68.06	42.54	33	59	99	Proprietary
Security funds in Taobao (yuan)	1,825,342	312.78	827.78	0	0	1,000	Proprietary
VIP class	1,825,342	2.94	1.28	2	3	4	Proprietary
Number of good feedbacks from clients	1,825,342	950.94	8,002.71	23	118	460	Proprietary
Network effect score	1,825,342	64.48	27.66	48.85	59.95	74.09	Proprietary
Link to credit card (0/1)	1,825,342	0.56	0.50	0	1	1	Proprietary
Daily average Yu'eBao balance (last three months, yuan)	1,825,342	974.84	6,656.89	0	106	532.60	Proprietary
Daily average Alipay wallet balance (last three months, yuan)	1,825,342	871.30	2,487.46	95.92	362.18	1,117.84	Proprietary
Daily payment activity (index, last year)	1,822,407	1,511.85	3,682.70	647	1,007.00	1,587.00	Proprietary
Payment activity (index, last six months)	1,821,221	204.72	623.59	81	140	236	Proprietary
Total amount of e-commerce purchases (last six months, yuan)	1,821,221	33,514.44	128,465.10	9,132.47	18,080.32	34,788.37	Proprietary
Quarterly consumption of goods (yuan)	1,825,342	838.73	7,754.11	0	57.70	323	Proprietary
Number of fulfilled contracts in daily life	1,822,423	1.07	0.63	1	1	1	Proprietary
Stability of contact information (index, last year)	1,822,423	1.87	1.35	1	2	2	Proprietary
Duration in one location (index)	1,815,137	1,815.63	762.28	1,261.00	1,714.00	2,318.00	Proprietary

Note: 1. We include 32 traditional variables in our analysis. However, only part of the list is shown in this table due to commercial confidentiality issues. 2. *Shop rating* is an evaluation of the borrower's business in Taobao, based on a unique evaluation system. A higher rating indicates a better assessment. 3. *Network effect score* is calculated from variables in several tens of dimensions to capture the relative importance of an individual or firm in the entire social and economic networks. A higher score shows a bigger impact 4. *Daily payment activity last year* and *Payment activity over the last six months* are indexes constructed from payment variables of the borrowers to reflect their transaction activities. 5. *Stability of contact information* and *Duration in one location* are constructed from the duration of using specific contact information and a fixed residential address; a larger index indicates a more stable livelihood.

In addition to these variables, we include bank credit history for borrowers who have borrowed from traditional banks in the past. This information allows us to compare the relative importance of traditional data and proprietary information from Ant and bank credit history for credit risk assessment. Since the majority of the MYbank borrowers are not served by traditional banks, the sample size for this exercise using bank credit history information is much smaller, with 145,109 loan transactions, or about 8 percent of the entire sample.

B.   Analytical Strategy and Empirical Results

Our empirical analyses include four steps. First, we conduct a horse race between the fintech approach and the traditional approach for credit risk assessment. This should allow us to quantify the information and model advantages by using the same models and data sets. Second, we assess the role of big data in credit assessment, relative to bank credit history information. In particular, we explore whether big data, especially BigTech's proprietary information, alone are sufficient for reliable credit risk assessment in the absence of bank credit history information. Third, we test whether the relative outperformance of the fintech approach can survive an exogenous shock, such as the adoption of a new regulatory policy. Fourth, we explore the inclusive nature of the fintech approach by comparing its performance with that of the traditional approach for subsamples of borrowers of different sizes and in different cities. The data used in our analysis were provided by MYbank on a confidential basis and are not available to the public given protection of the privacy of the customers.

The basic strategy is to conduct four sets of horse races to evaluate the roles of information and modeling methods in assessing credit risk, which is proxied by default. We compare the contributions of traditional data with that of MYbank's proprietary information by controlling for the same models. And we look at the predictive power of traditional scorecard models and machine learning models, by using the same sets of information. The scorecard model is widely used in the financial industry for calculating credit scores. For the machine learning model, we use the random forest model in our baseline, à la Butaru et al. (2016).[10] We also use other machine learning models (such as the Gradient Boosting Decision Tree) for robustness checks, and the conclusions remain unchanged. In the

---

[10] Appendix B provides a more detailed elaboration of the formulas and algorithms.

following analysis, traditional models refer to standard scorecard models, while machine learning models refer to random forest models. The four models (see Table 2) are model I, or the traditional approach (scorecard models + traditional information) ; model II (scorecard models + all information); model III (machine learning models + traditional information); and model IV, or the fintech approach (machine learning models + all information). The difference in the performance of models I and IV captures the relative advantage of fintech versus the traditional approach, while other pairwise comparisons show the marginal contribution of big data (model IV versus model III and model II versus model I) and the machine learning model (model IV versus model II and model III versus model I).

Table 2. **Model Definitions**

	Scorecard	Random forest
Traditional information	Model I (traditional approach)	Model III
All information (traditional + proprietary data)	Model II	Model IV (fintech approach)

For the out-of-sample tests, we split the sample into two subperiods (March to May 2017 and July to August 2017).[11] The first subsample is used to estimate the models (training set), and the second subsample is used to test the models (testing set). There are 771,596 loan transactions in the training set and 1,053,748, in the testing set.

Baseline

The baseline results of the four models are reported in Figure 7 and Table 3. Figure 7 shows the receiver operating characteristics (ROC) curves. The true positive rate (on the vertical axis) is also known as sensitivity. The false positive rate (on the horizontal axis), known as specificity, is basically the fallout or probability of a false alarm and can be calculated as (1 − specificity). At a given level of specificity (false positive), a superior outcome would have a higher level of sensitivity (true positive). For a given model, it is better if the true positive rate is higher and the false positive rate is lower, which means that the ROC is closer to the upper left-hand corner of the diagram.

The results show that model IV (purple line, random forest + all information) performs best. Model III (blue line, random forest + traditional information) is second in line, followed by model II (green line, scored card + all information). Model I (red line, scored card + traditional information) is the

---

[11] The results are robust to alternative subsample periods.

least efficient among all four. Since we control the size of the sample and the two lists of variables, we can conclude as follows. First, the fintech approach is more reliable than the traditional approach in predicting defaults. Second, if we replace the traditional information in the traditional approach with all information (or adding proprietary information to traditional variables), the performance of the model (model II) improves significantly. This may be characterized as an *information advantage*. Third, if we replace the scorecard model with the machine learning model, the performance of the model (model III) improves compared with model I. This may be characterized as a *model advantage*. In summary, comparison of the ROCs of the four models indicates that the fintech approach exhibits information and model advantages over the traditional approach. In this particular case, the model advantage appears to be greater than the information advantage. However, this finding could be sample dependent.

Figure 7: ROC Curves for Four Models
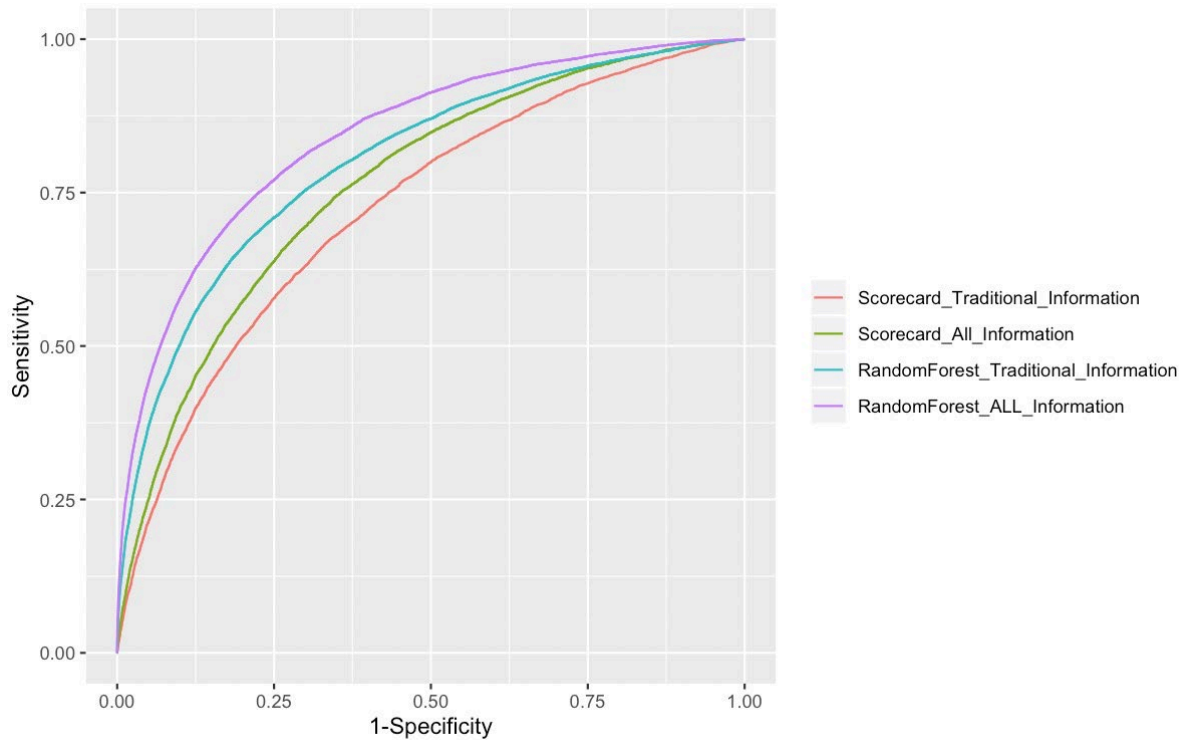
Source: Calculations using data from MYbank.
Note: The graph compares four model specifications by using ROC curves. ROC = receiver operating characteristics.

Table 3 reports the area under the ROC curve (AUC), which is widely used to measure the discriminatory power of credit scores (Berg et al. 2019). The AUC ranges from 50 percent (purely random prediction) to 100 percent (perfect prediction). The general rule is that the model is reasonably reliable if the AUC is above 60 percent, and it performs strongly if the AUC is above 70 percent (Berg et al. 2019). Table 3 reveals some interesting results. First, with relatively few

variables, the machine learning models (random forest) show no advantage (illustrated by smaller AUCs for machine learning models than the scorecard models in the first two columns in Table 3). This finding implies that machine learning models are more powerful for large data sets. Second, while relying on Ant's proprietary information ((e) in Table 3) should be sufficient for conducting credit risk assessment (with AUC of 0.76), its advantage over model I (traditional variables + scorecard model, with AUC of 0.72) is limited. Third, other things being equal, adding more variables improves the effectiveness of credit risk assessment. Fourth, the combination of a large data set and a machine learning model greatly improves credit risk assessment. Adding Ant's proprietary information to the traditional approach increases the AUC by 5.6 percent, while applying machine learning techniques adds an additional 11.1 percent to the AUC. Table 4 shows the 95% confidence intervals of the AUC for the four models (models I to IV) in Figure 7. The differences in AUCs among the four models are significant.

Table 3: Comparison of AUCs for the Scorecard and Random Forest Models

Model	(a)(b)	(c)	(a)(b)(c)	(a)(b)(c)(d)	(e)	(a)(b)(c)(d)(e)
Scorecard	0.70	0.58	0.72	0.72	0.69	0.76
Random forest	0.67	0.54	0.76	0.80	0.76	0.84

Source: Calculations using data from Mybank.
Note: 1. The table shows the discriminatory power of different model specifications by providing the AUC. 2. (a) refers to asset and financial information; (b) refers to credit history (only from MYbank); (c) refers to vendor-specific information; (d) refers to local economy information; and (e) refers to Ant proprietary information. AUC = area under the receiver operating characteristics curve.

Table 4: 95% Confidence Intervals of the AUCs for Four Models

Model	AUC	95% conf. interval	
Model I (scored card + traditional information)	0.7246	0.7222	0.7269
Model II (scored card + all information)	0.7636	0.7615	0.7658
Model III (random forest + traditional information)	0.8022	0.8001	0.8043
Model IV (random forest + all information)	0.8414	0.8396	0.8433

Source: Calculations using data from Mybank.
Note: AUC = area under the receiver operating characteristics curve.

Table 5 reports the reduction in the NPL ratio when switching from the traditional approach to the fintech approach, using alternative thresholds of the expected default rate. Loans with expected default rates below the threshold will be accepted. We then calculate the default rates (NPL ratios) of loans accepted by the fintech and traditional approaches. The right column shows the reduction in NPL ratios from using the fintech approach, compared with the traditional approach, at given threshold levels. For example, if the threshold is an expected default probability of 10 percent, the fintech approach reduces the NPL ratio by 1.05 percentage points. The improvement becomes smaller as we relax the threshold values from 10 to 30 percent.

Table 5: Reduction in NPL Ratio from Traditional Approach to Fintech Approach

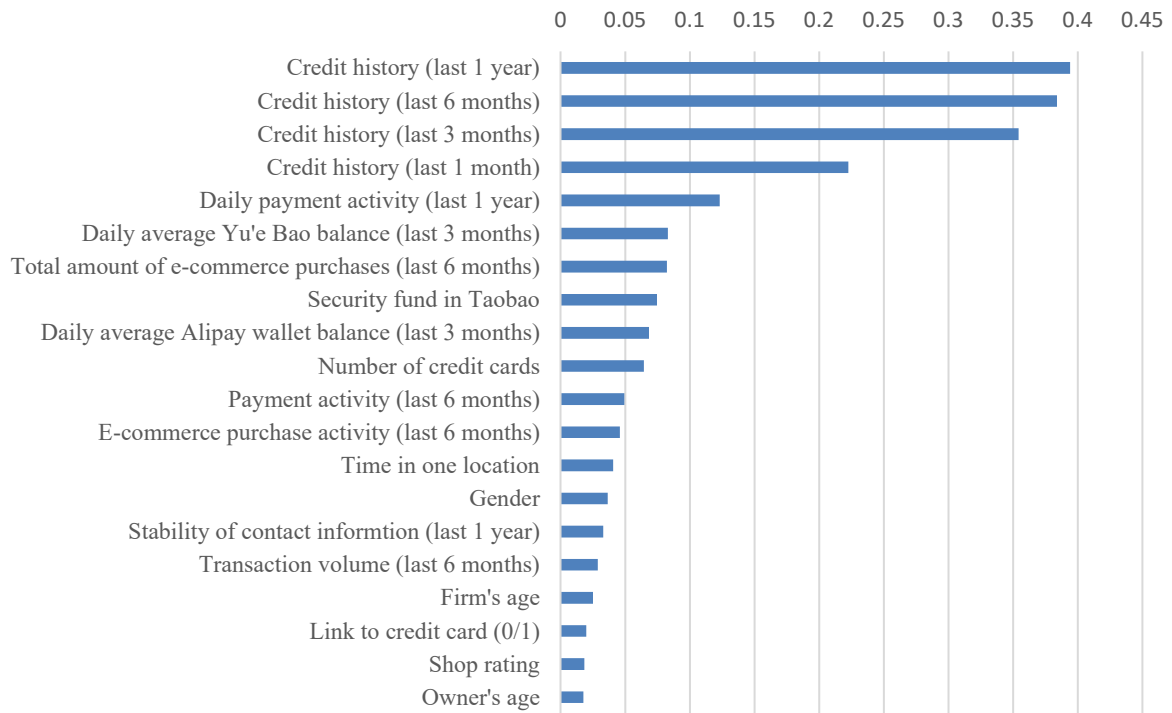| Threshold (%)	Reduction in NPL ratio (percentage points)
10	1.05
15	0.76
20	0.48
25	0.27
30	0.15

Source: Calculations using data from MYbank.
Note: NPL = nonperforming loans.

When comparing the performance of different modeling methods, we control for the same information set. However, the same variables may contribute to the results differently using different models. To illustrate this point, we report the top 20 contributing variables for the scorecard and random forest models. Since the two models use different criteria in evaluating the importance of the variables, the two sets of parameters are not comparable numerically. The scorecard model results are primarily driven by five variables, four on credit history and one on daily payment activity (Figure 8, panel a). Among the top 20 variables, most are on transactions and payments. In contrast, in the machine learning model results, information values are more evenly distributed across a wide range of variables (Figure 8, panel b), and transaction and payment data play a more important role than in the scorecard model. Critically, the key difference in the two models is the role of Ant's proprietary information, such as customers' ratings of vendors and network effect scores in the machine learning models. In a way, these indicators represent the borrowers' "digital collateral" pledged in the Alibaba ecosystem, as their reputation and business operations would be immediately affected in the case of default. These variables do not appear in the list of the top 20 variables in the scorecard model, probably because the relationships between these proprietary information variables and predicted defaults are not linear. Rather, these variables affect the results through interactions with other variables. For example, a high network effect score alone cannot guarantee loan repayment; it is useful only when combined with healthy cash flows.

Figure 8: Top 20 Information Variables in the Scorecard and Random Forest Models

a. Ranking of information value: Scorecard model

Variable	Value
Credit history (last 1 year)	~0.39
Credit history (last 6 months)	~0.38
Credit history (last 3 months)	~0.35
Credit history (last 1 month)	~0.22
Daily payment activity (last 1 year)	~0.12
Daily average Yu'e Bao balance (last 3 months)	~0.08
Total amount of e-commerce purchases (last 6 months)	~0.08
Security fund in Taobao	~0.07
Daily average Alipay wallet balance (last 3 months)	~0.065
Number of credit cards	~0.06
Payment activity (last 6 months)	~0.05
E-commerce purchase activity (last 6 months)	~0.045
Time in one location	~0.04
Gender	~0.035
Stability of contact informtion (last 1 year)	~0.03
Transaction volume (last 6 months)	~0.03
Firm's age	~0.025
Link to credit card (0/1)	~0.02
Shop rating	~0.02
Owner's age	~0.02

b. Ranking of Gini impurity: Random forest model

Firm's age
Daily payment activity (last 1 year)
Time in one location
Network effect score
Total inflow of funds in Alipay
Transaction volume
Total amount of e-commerce purchases (last 6 months)
Daily average Alipay wallet balance (last 3 months)
E-commerce purchase activity (last 6 months)
Number of good feedbacks from clients
Payment activity (last 6 months)
Owner's age
Credit history (last 1 month)
Daily average Yu'e Bao balance (last 3 months)
Credit history (last 6 months)
Number of credit cards
Quarterly consumption of commodities
Credit history (last 3 months)
Ratio of sham transactions (last 6 months)
Shop rating

Source: Calculations using data from MYbank.
Note: In panel a, the x-axis shows the information value of each variable in the scorecard model. In panel b, the x-axis shows the Gini impurity of each variable in the random forest model. Both indicators measure the predictive power of the variables, but the values are not directly comparable. Appendix B provides a more detailed elaboration of information value and Gini impurity.

Role of Bank Credit History

Next, we repeat the same exercises for a subsample of borrowers with bank credit history, which accounts for 7.5 percent of the total. We compare the importance of Ant information ((a)(b)(c)(d)(e) in Table 3) relative to bank credit history. Again, a comparison of the ROC curves shows that the approach with machine learning models and all information is the most efficient, while the approach with scorecard models and bank credit history information is the least efficient (Figure 9). Adding Ant's information set to the traditional approach or replacing the scorecard models with machine learning models significantly improves credit risk assessment.

Figure 9: ROC Curves for Different Models for Traditional Bank Borrowers
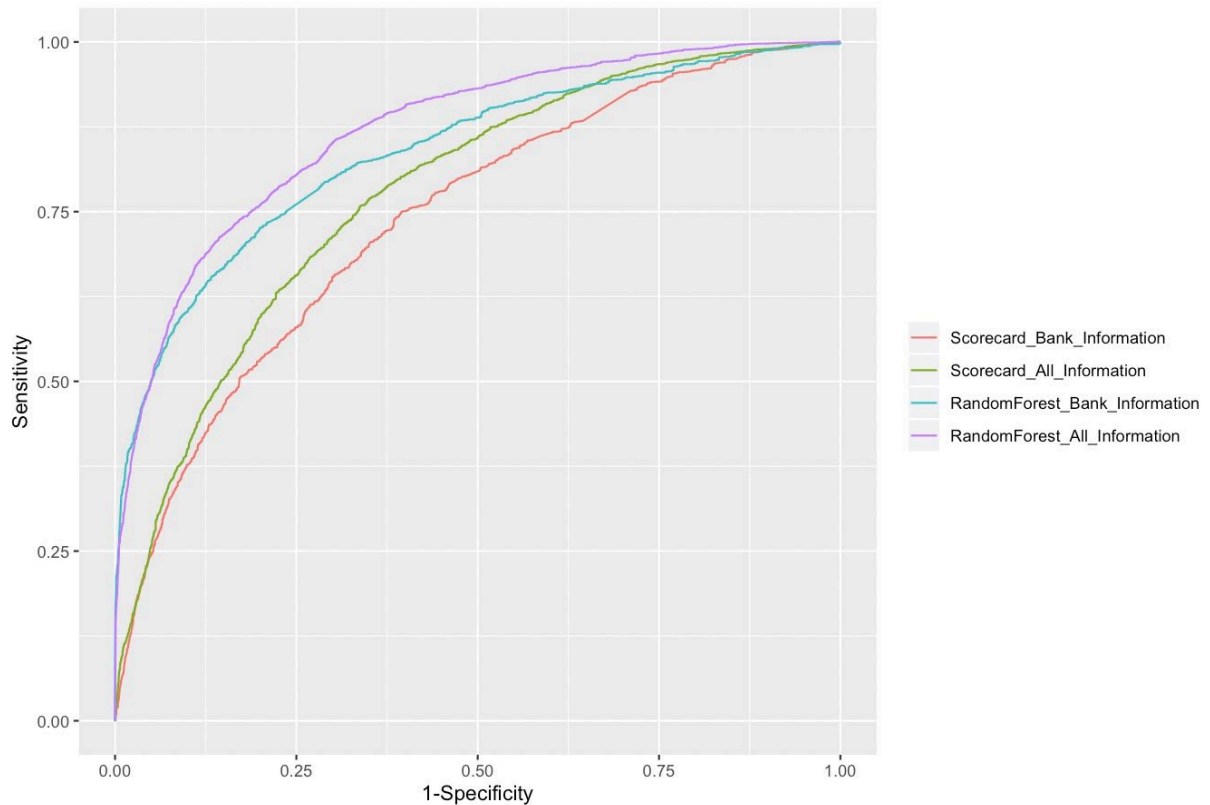
Source: Calculations using data from MYbank.
Note: The graph shows the discriminatory power of four model specifications by providing the ROC curves. The sample only includes borrowers who have a credit history in traditional banking. ROC = receiver operating characteristics.

Table 6 compares the AUCs of different combinations of models and information sets. First, bank credit history information is quite effective for predicting defaults, with an AUC of 0.74 in the traditional approach using the scorecard models. Second, again, applying the machine learning models significantly increases the accuracy of predicting defaults, raising the AUC to 0.84. Third, Ant information alone generates similar results as bank credit history information, using the scorecard and machine learning models. This finding is especially significant because many SMEs are not registered in the credit registration system. This result suggests that fintech lenders could cover the massive number of small borrowers that have never been serviced by banks. And finally, the combination of the complete set of information and machine learning models delivers the best outcome for predicting defaults. Table 7 shows the 95% confidence intervals of the AUCs for the four models in Figure 9. The differences in the AUCs among the four models are significant.

Table 6: Comparison of AUCs for Different Models for Traditional Bank Borrowers

Model	Bank credit history information	Ant information	Bank credit history information + Ant information
Scorecard model	0.74	0.72	0.78
Random forest	0.84	0.83	0.87

Source: Calculations using data from MYbank.
Note: The table shows the discriminatory power of different model specifications by providing the AUC. Ant's information includes all the information (traditional information and Ant's proprietary information) we used in the baseline model. The sample only includes borrowers who have a credit history in traditional banking. AUC = area under the receiver operating characteristics curve.

Table 7: 95% Confidence Intervals of the AUCs for Four Models

Model	AUC	95% conf. interval	
Random forest + bank credit history information + Ant information	0.8681	0.8636	0.8725
Random forest + bank credit history information	0.8372	0.8318	0.8427
Scorecard model + bank credit history information + Ant information	0.7767	0.7713	0.7820
Scorecard model + bank credit history information	0.7397	0.7337	0.7456

Source: Calculations using data from MYbank.
Note: AUC = area under the receiver operating characteristics curve.

Impact of a Policy Shock

So far, our results are broadly in line with the findings of Berg et al. (2019) and others on the value of digital footprints for credit risk assessment, and Gambacorta et al. (2019) on the advantages of machine learning models over scorecard models. However, policy makers and academics frequently caution that the robustness of the fintech approach for credit risk assessment still needs to be tested, especially if the financial cycle changes direction.

To address this concern, we analyze the impact of an exogenous regulatory shock on the performance of the credit risk models. On November 17, 2017, the People's Bank of China issued draft guidelines to tighten regulations on the asset management activities of financial institutions, to curb the growing risks in China's shadow banking sector. The new regulations is a significant move to ensure China's long-term financial stability, but also led to a tightening of financial conditions in the short-run, with credit growth dropping by 4 percentage points the following year. SMEs were among the hardest hit, given their heavy reliance on shadow banking. The NPL ratio of the banking system rose from 1.74 percent before the announcement to 1.87 percent one year later. MYbank's NPL ratio edged up from 1.23 to 1.3 percent.

We use loan repayment records to calculate the discriminatory power of the four models in two periods, before and after the shock (October 2017–November 2017 and December 2017–January 2018, respectively). Again, discriminatory power is measured by the AUC. Table 8 shows the

differences in the AUCs of the four models. The difference in AUCs between model IV (random forest + all information) and model I (scorecard model + traditional information) is even greater after the shock than before the shock. The outperformance of the fintech approach over the traditional approach, measured by the AUC, increased from 0.11 before the shock to 0.14 after the shock. Comparisons of the other models also reveal information and model advantages in the face of the policy shock.

Table 8: AUC Differences before and after a Policy Shock

AUC difference	October–November 2017 (before shock)	December 2017–January 2018 (after shock)
Model IV – Model I (comprehensive evaluation)	0.11	0.14
Model IV – Model II (model advantage)	0.07	0.09
Model II – Model I (information advantage)	0.04	0.05

Source: Calculations using data from MYbank.
Note: The table shows the difference between the areas under the AUC curves of four models before and after a shock. AUC = area under the receiver operating characteristics curve.

These results suggest that the fintech approach is not only more effective, but also more robust than the traditional approach. For the short-term loans investigated in this paper, the fintech approach still outperforms the traditional approach when the financial cycle turns. Similar to the case in normal times, this resilience in abnormal times can be explained by the approach's information and model advantages. In addition to traditional data, the fintech approach relies on real-time data and behavior information, which are not only more up-to-date, but also more stable. And the machine learning model captures dynamic and interactive relationships. Comparatively speaking, the scorecard models are linear, and some of the traditional data can easily become outdated.

Inclusiveness of Fintech Lending

The stylized facts show that the firms served by MYbank are much smaller than China's average SMEs. Does the advantage of fintech credit assessment differ for different SMEs in this sample? The answer to this question has important implications for market structure: if Ant's fintech lending favors larger firms in the sample, this could indicate that there is a greater chance of competition with bank lending. However, if the fintech approach favors smaller firms, then there should be greater complementarity between fintech and bank lending. Even the "larger" firms in our sample are still SMEs.

We apply the cumulative distribution function of increase in estimated default probability in the fintech and scorecard models in figure 10. Each line in the figure represents a subgroup. Borrowers for whom this difference is negative (to the left of 0 on the horizontal axis) are "winners" in the fintech approach (in the sense of having a lower estimated default probability), and those with a positive difference (to the right of 0 on the horizontal axis) are "losers" (Fuster et al. 2020).

Figure 10, panel a, compares the differences in predicted default probabilities between the fintech approach and the traditional approach for different subgroups of SMEs (measured by annual sales turnover). Among the smallest borrowers, the share for whom the estimated default probabilities drop using the fintech model is around 57 percent, while it is around 27 percent for the largest borrowers. These findings mean that smaller businesses have lower estimated default probabilities and, therefore, are more likely to obtain loans when the fintech approach is applied. Thus, compared with the traditional approach, the fintech approach benefits smaller firms proportionately more than larger firms.

Figure 10, panel b, compares the predicted default probabilities between the fintech approach and the traditional approach for SMEs in different cities. For instance, for borrowers in Tier-4 cities,[12] about 62 percent have lower predicted default probabilities using the fintech model, compared with 52 percent for borrowers in Tier-1 and Tier-2 cities. This finding implies that SMEs in lower-tier cities benefit more from the fintech approach than those in higher-tier cities, likely because SMEs in smaller cities lack traditional data and rely more on proprietary information to secure small fintech loans.

---

[12] There is no official government guideline on city classification. The tiering system adopted here is widely used in the media, reflecting a confluence of factors, such as economic development, population size, and administrative hierarchy. Tier-1 cities represent the most densely populated and developed urban areas in China, while Tier-4 cities are small and less developed.

Figure 10: Comparison of Predicted Default Probabilities across Models

a. Comparing SMEs of different sizes
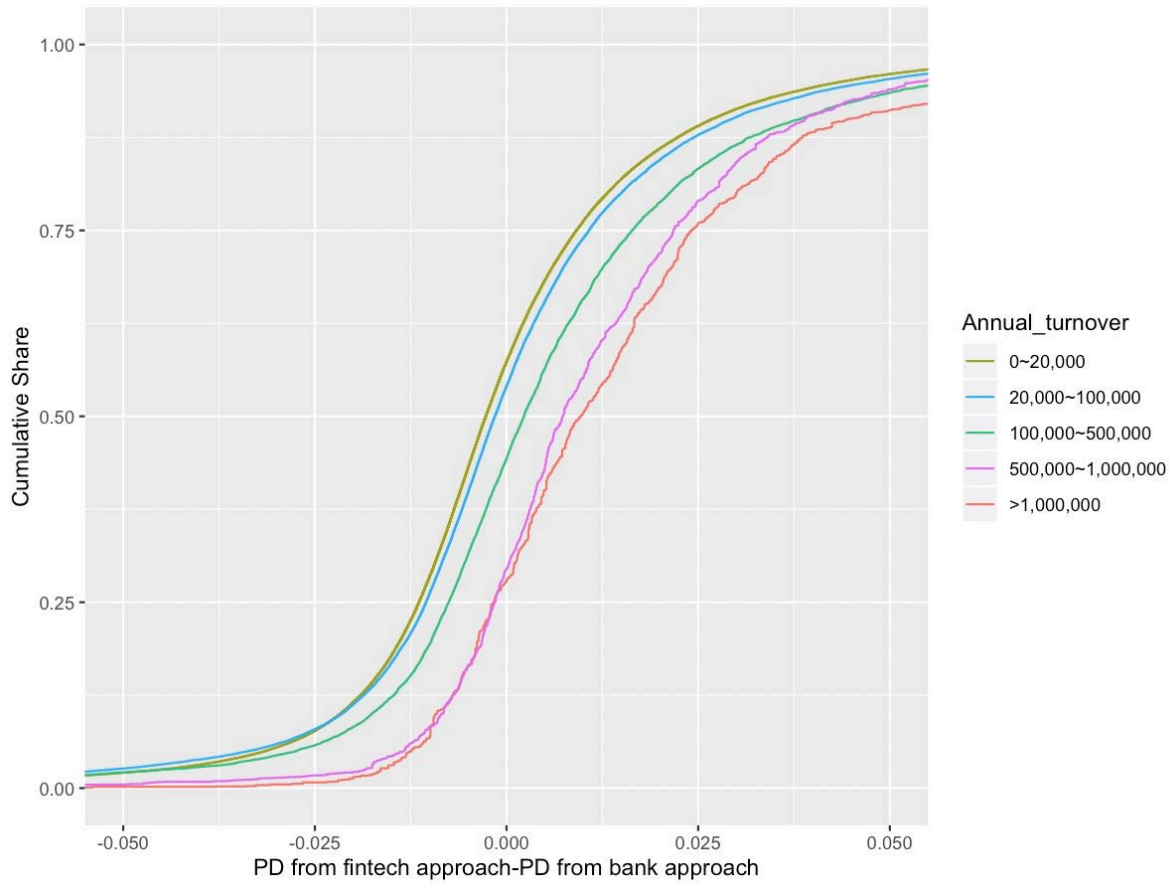
b. Comparing SMEs in different city tiers
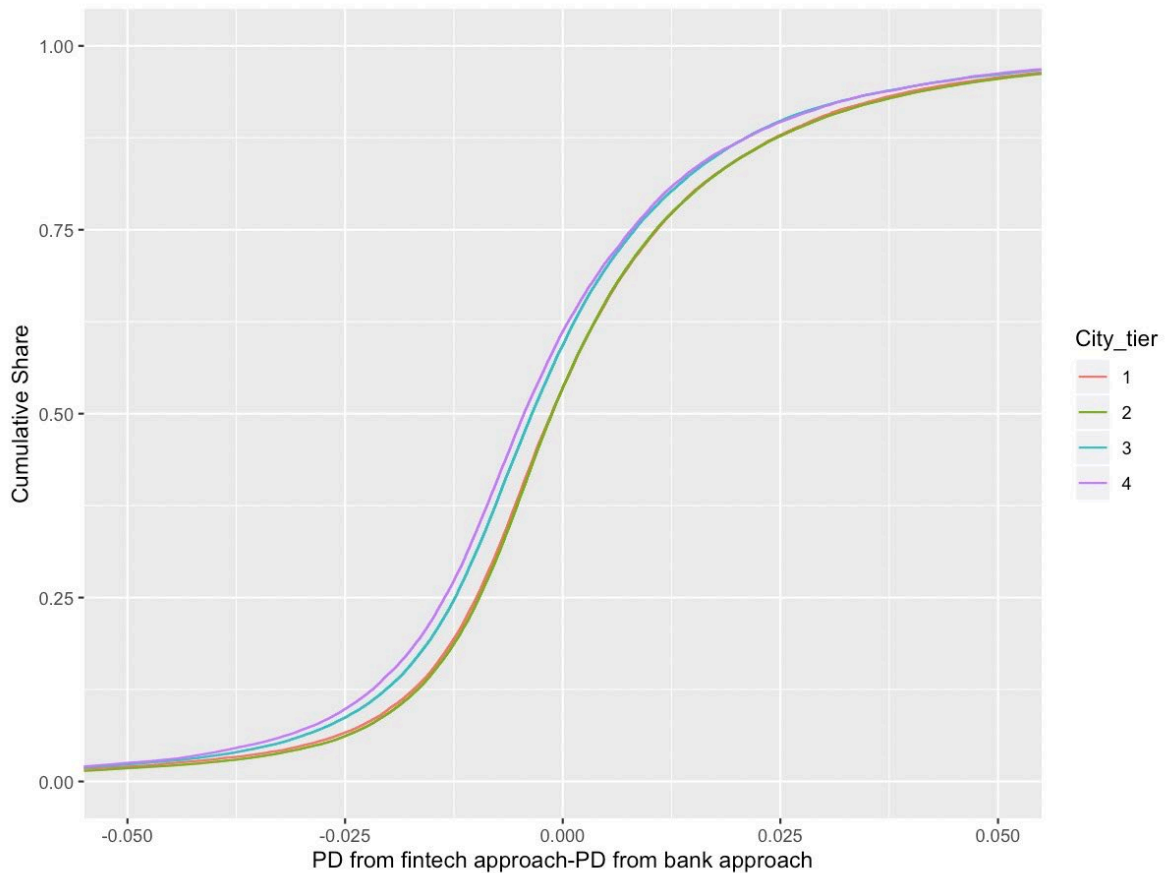
Source: Calculations using data from MYbank.
Note: The graphs plot the cumulative distribution function of differences in the predicted probabilities of default between the fintech model and the traditional model by differentiating SME sizes (panel a) and city tiers (panel b). PD = predicted default; SMEs = small and medium-size enterprises.

To illustrate the results for subgroups of SMEs, Figure 11 plots the share of SMEs that obtain a lower predicted default rate using the fintech approach, by size and city location. The fintech approach favors smaller SMEs, especially those with annual turnovers of less than RMB 100,000. There is no clear advantage of applying the fintech approach to larger SMEs, as banks may already have sufficient information on these companies. When grouped by city tier, all the SMEs benefit from the fintech approach, although those in lower-tier cities benefit even more.

Figure 11: Share of SMEs with Lower Predicted Default Rates Using the Fintech Approach,

by Firm Size and City Location
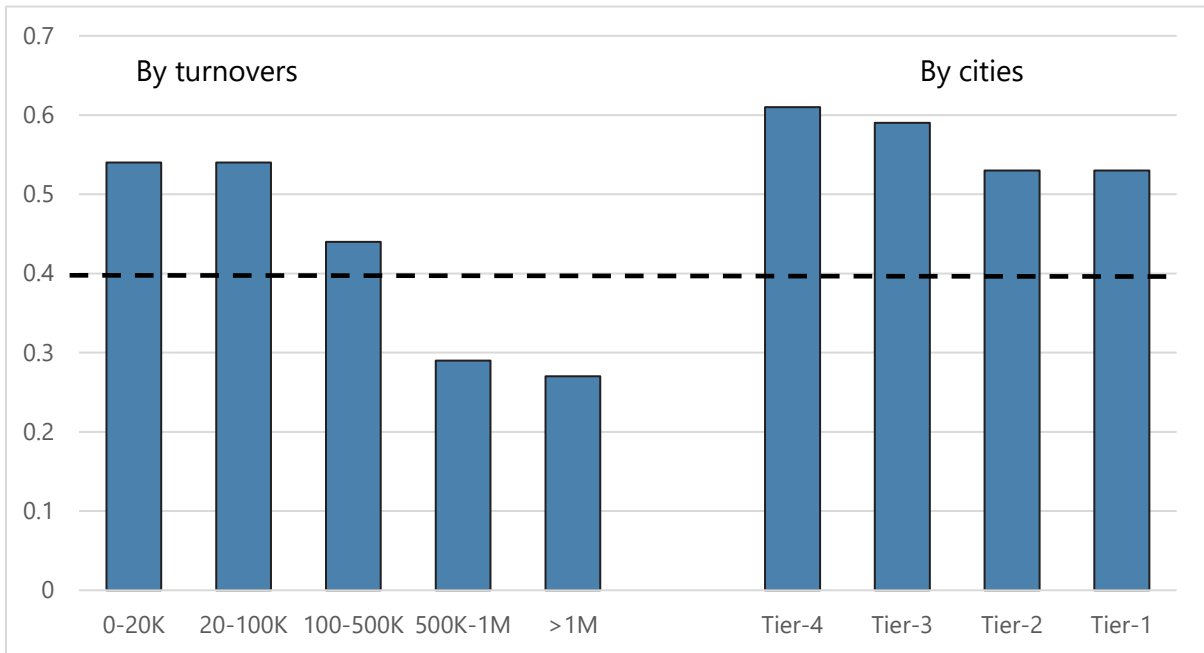
Source: Calculations using data from MYbank.

These findings confirm that, compared with the traditional approach, the fintech approach benefits smaller SMEs and SMEs in smaller cities. A follow-up question is whether such lower predicted default probabilities are as reliable as the other predictions. Here again, we use the AUCs to measure the reliability of the models for different borrower groups (Table 9). The first row in Table 9 shows the AUCs of the traditional approach for borrowers of different sizes. Apparently, the model works better for relatively larger firms, that is, firms with annual turnover of at least RMB 0.5 million. The AUCs shown in the second row for the fintech approach offer a similar pattern. However, by comparing the first and second rows, we find that the differences are greater for smaller firms than for larger firms (third row).

Table 9: AUCs for Different Borrower Groups under Different Models

a. Firm size (RMB, thousands)

Borrowers, by annual turnover	< 20	20-100	100-500	500-1,000	> 1,000
Traditional approach	0.72	0.72	0.71	0.8	0.76
Fintech approach	0.84	0.84	0.82	0.89	0.85
Fintech – traditional	0.12	0.12	0.11	0.09	0.09

b. Firm location

Borrowers, by city tier	Tier-4	Tier-3	Tier-2	Tier-1
Traditional approach	0.71	0.73	0.73	0.73
Fintech approach	0.84	0.85	0.84	0.84
Fintech – traditional	0.13	0.12	0.11	0.11

Source: Calculations using data from MYbank.
Note: The table shows the discriminatory power of the traditional and fintech model specifications across groups by providing the AUCs. The fintech approach refers to a combination of random forest models and all information. The traditional approach refers to a combination of scorecard models and traditional information. AUC = area under the receiver operating characteristics curve.

V. CONCLUSIONS AND POLICY IMPLICATIONS

This study conducted analyses of the fintech approach for credit risk assessment, using big data and machine learning models, relative to the traditional approach, featuring standard financial information and scorecard models. Using a unique data set from China's MYbank, the study finds that the fintech approach has a significant advantage in strengthening credit risk management and promoting financial inclusion for smaller SMEs.

First, compared with the traditional approach, the fintech approach more accurately predicts loan defaults, reflecting the combined advantages of the information (e.g., proprietary information from BigTech companies) and model (e.g., machine learning models). The information advantage is likely associated with more real-time financial data and behavior information. Compared with lagging financial histories, real-time data do a better job in assessing borrowers' capacity to repay loans. In addition, behavior information better captures borrowers' willingness to repay.[13] The model advantage shows that machine learning models help capture the nonlinear relationships between the response and feature variables and the interactive effects among the feature variables. Some of the BigTech-specific variables, such as customer ratings and network effect scores, are quite important in the fintech approach but not in the traditional approach.

Second, the study analyzed a small subsample of MYbank borrowers who have borrowed from traditional banks. We find that bank credit history information alone predicts loan defaults reasonably well using scorecard models. However, adding traditional data and Ant's proprietary information to the analysis and replacing the scorecard models with machine learning models significantly improves

---

[13] One caveat of our analysis is that traditional banks may have other soft information about borrowers and expert judgment that is not reflected in the data. In addition, our study focuses on uncollateralized loans, while traditional banks are more experienced in assessing collateral quality and issuing collateralized loans.

forecast accuracy (with the AUC increasing from 0.74 to 0.87). Clearly, additional data and machine learning models contribute to the efficiency of credit risk assessment. More importantly, machine learning models based on all the information without bank credit history information generate an AUC of 0.83. This means that proprietary information not only complements bank credit history information, but also replaces it, when needed. Such finding means that providing fintech lending to the vast number of unbanked customers, who account for more than 90 percent of the sample used in this study, has become possible.

The study tested the robustness of the findings in the face of an exogenous policy shock. As part of its deleveraging campaign, the People's Bank of China issued new regulations on asset management at the end of 2017, which greatly improved long-term financial stability, while also led to a tightening of financial conditions in the short-run, in particular for small firms. We find that the advantage of the fintech approach is more pronounced during the after-shock period compared with the pre-shock period, as evidenced by the larger gain in the AUC, reflecting the contributions of the information and model advantages. This finding suggests that the fintech approach is robust when facing regulatory policy shocks. That said, given the short history of fintech development, its performance during other types of financial shocks remains to be tested.

Finally, the study compared the performances of the fintech and traditional approaches in credit assessment for SMEs of different sizes and in different cities. We find that the improvement in loan default prediction with fintech is largest for the smallest firms. Similarly, SMEs in lower-tier cities, such as Tier-3 and Tier-4 cities, benefit more than those in Tier-1 and Tier-2 cities. These results demonstrate the inclusive feature of fintech lending.

These findings have important implications for policy makers to strengthen credit risk assessment and promote financial inclusion.

Promoting fintech lending while strengthening the regulatory framework. Given the significant advantage of fintech lenders in reaching unbanked SMEs, with unique data access and enhanced risk modeling, the government could actively encourage fintech lending to promote financial inclusion. Nonetheless, the supporting policies should be complemented with strict licensing and regulatory checkups. Although the fintech approach can achieve better credit risk assessment, it requires two critical elements: access to big data and capacity for machine learning analysis. The failing peer-to-peer lending industry in China offers some useful lessons, including that a pure online lending platform without big data analysis could build up significant financial risks. Given the short history of fintech development, regulators need to stay vigilant on new risks that might emerge from its business model, particularly in the downturn of financial cycles. Overall, regulators must strike a delicate balance between promoting financial innovation and maintaining financial stability. Given the wide range of data and analytical models, and rapid changes in technology, regulators could use a sandbox to incubate new business models and practices.

Encouraging collaboration between fintech lenders and traditional banks. On the one hand, fintech lenders have the advantages of information and analytical skills, but their capacity to lend remains small due to capital and funding constraints. On the other hand, traditional banks have substantial financial resources and other soft information on existing borrowers, yet limited capacity in big data and risk modeling. Policy makers could encourage more collaboration by combining their strengths. In China, for instance, traditional banks have started to develop their own contactless and data-driven SME financing by accumulating data and analytical skills, often in collaboration with fintech firms. As of the end of 2019, RMB 2 trillion in business loans (equivalent to 2 percent of gross domestic product) were jointly issued by banks and fintech firms, where fintech lenders provided the initial screening of borrowers and banks provided most of the funding. Therefore, providing joint loans between fintech firms and traditional banks can help promote financial inclusion, and the detailed modality of cooperation can be designed on a case-by-case basis. Such a business model could also be a useful reference for other countries where there are significant shortages of lending to SMEs. Going forward, the overall financial landscape is likely to change significantly. For instance, the fintech approach for credit risk assessment may become a common tool for all lenders.

Enhancing the policy framework on data protection. Given the critical role of data in fintech risk assessment, governments should set clear data policy standards to maximize the benefits of big data analysis while protecting individuals' privacy. Currently, the data used by fintech lenders come from multiple sources, including e-commerce platforms, industry and commerce registration, and business operations. Greater emphasis needs to be placed on the data's authenticity, legality, and security. In China, the government is in the process of enacting a personal information protection law and a data security law. It might also be useful to foster development and competition in fintech lending, by allowing individuals to carry their data to other platforms.

Ensuring data access. Our analysis shows that the proprietary data of fintech firms is highly valuable for credit risk assessment. Since such information often comes from parent companies in the platform industry, such as e-commerce or social media, it might add a new dimension to the discussion of competition policy in an area characterized by a high degree of concentration that often reflects network effects and economies of scale. A lack of sharing of proprietary data within the financial sector would further add to this problem, while ensuring such data can be used widely by others will alleviate it. In the meantime, ensuring data access needs to be balanced with privacy protection and proper data usage.

China's experience and success in fintech lending could provide a useful lesson for other countries where SME lending remains a challenge. However, it is important to recognize that the development of fintech requires a strong digital infrastructure, an established online ecosystem with a wide user reach, as well as big data analytical skills. These are critical elements for an internet company to become a successful fintech lender. Hence, government priority could be sequenced in investing in digital infrastructure, supporting e-commerce and broader digitalization, and later branching out to fintech lending.

APPENDIX A. MYBANK

China has been at the forefront of fintech development and is the largest fintech market in the world, with virtual banking being one of its wide-reaching features. Enabled by digital technology and big data, China's big four tech players—Alibaba, Baidu, Tencent, and JD—have made incursions into financial services. During 2014–16, China's banking regulator issued 11 new privately-owned banking licenses, including to MYbank. By aiming to bring private money into the Chinese financial services sector, three banks (MYbank, WeBank, and XW Bank) have set up a wider reach to millions of small and medium-size enterprises (SMEs).

Headquartered in Hangzhou, Zhejiang province (in southeast China), MYbank was founded in 2015 by Alibaba's affiliate firm, Ant Group, through a 30 percent stake in a joint venture comprising a group of private firms. MYbank uses big data, machine learning, and the associated flexible risk management approach to offer credit to SMEs and manage risks. MYbank has about 20 million SME borrowers, about 80 percent of whom have never borrowed from banks in the past, and keeps its nonperforming loan (NPL) ratio at about 1 percent.

MYbank provides financing for SMEs and individuals in urban and rural areas. Its data/cloud-based model with no physical branches makes its operational costs much lower than the traditional brick-and-mortar banking model. By harnessing its credit-profiling techniques driven by big data analytics (e.g., e-commerce and cash flow), MYbank manages to approve small loans for individuals and businesses around the clock 24 hours a day, seven days a week, providing capital and liquidity for the masses in China.

In 2019, MYbank's total assets were RMB 139.6 billion ($20 billion), the average loan size was RMB 31,000 ($4,500), the accumulated SMEs served reached 20.9 million, the capital adequacy ratio was 16.4 percent, and the NPL ratio was 1.3 percent.

APPENDIX B. ALGORITHMS

Random Forest

Random forest is one of the most popular machine learning algorithms. As its name implies, it builds a large collection of uncorrelated *decision trees*, and then averages them. A typical decision tree is shown in Figure B1. The independent variables (attributes) are income and age. A decision tree is a flowchart-like tree structure, where each internal node (non-leaf node), denoted by rectangles, represents a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label (default or not).
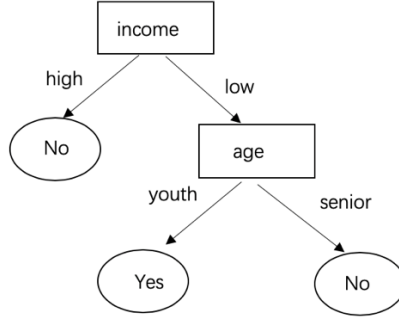
Figure B1. Decision Tree for Classifying Default

How to build a decision tree? The first step is to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a splitting attribute (variables) and a split-point (or splitting subset). The optimal criterion is determined by maximizing the effective change in node impurity after applying variable splitting and split-point pairs. We use the Gini index to measure node impurity. Assume there are C classes in the data set. In a node m, representing a region D_m with N_m observations, let

$$p_{mc} = \frac{1}{N_m} \sum_{i \in R_m} I(y_i = c)$$

the proportion of class c observations in node m. The Gini index is calculated as

$$Gini(D_m) = \sum_{c \in C} p_{mc}$$

In our analysis, there are two classes for each observation, default or not. So the Gini index is

$$Gini(D_m) = 2p_m(1 - p_m)$$

where p_m is the proportion of default observations in node m.

We compute a weighted sum of the impurity of each subset after splitting the N_m observations in node m based on attribute A. A binary split on A partitions D_m into $D_{m,1}$ and $D_{m,2}$, and the Gini index of D_m given that partitioning is

$$Gini_A(D_m) = \frac{N_{m,1}}{N_m} Gini(D_{m,1}) + \frac{N_{m,2}}{N_m} Gini(D_{m,2})$$

The reduction in node impurity (Gini index) incurred by a split on attribute A is

$$\Delta Gini(A) = Gini(D_m) - Gini_A(D_m)$$

The attribute and split-point that maximize the reduction in impurity (Gini index) are selected as the splitting criterion.

Having chosen the optimal splitting criterion, we can build a decision tree following steps 1 to 3:

1. The tree starts as a single node, representing the training data set in D. In the process of building the tree, if all observations are of the same class, node m becomes a leaf and is labeled with that class. If there are no remaining attributes on which the observations may be further partitioned, then node m becomes a leaf and is labeled with the most common class in D (majority voting).
2. Otherwise, the splitting criterion automatically separates or partitions the tuples in D into individual classes. Then we get the resulting partitions D_m.
3. The algorithm recursively repeats steps 1 and 2 to form a decision tree for the observations at each resulting partition D_m, until the minimum node size n_{min} is reached.

Random forest builds an ensemble of trees by bootstrapping M subsets from training data and trains M de-correlated trees on these subsets. Each decision tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction.

We calculate feature importance to measure the predictive power of each variable. Feature importance, or the importance of each independent variable, is measured by the improvement in the split criterion attributed to the splitting variable at each split in each tree and accumulated over all the trees in the forest.

In our analysis, the data set consists of 76 feature variables and a response variable indicating whether to default or not. We build 500 decision trees (M=500) in the forest and use open-source package RandomForestClassifier in Python to train random forest models.

Credit Scorecard

The credit scorecard model is widely used in the financial industry to evaluate borrowers' credit worthiness. The algorithm contains three steps.

The first step is to transform all the independent variables using the weight of evidence (WoE) method. Based on the proportion of good applicants to bad applicants at each group level of each independent variable, this method measures the "strength" of grouping for differentiating good and bad risk and attempts to find a monotonic relationship between the independent variables and the target variable, which takes 0 if there is no default, and 1 in the case of defaults. We first split the data into several bins of each independent variable, such as age and income group. Then we calculate the WoE. $WoE_{i,j}$ for bin i of independent variable j is defined as follows. B_i is the number of bad borrowers in bin i and B_T is the number of total bad borrowers. G_i and G_T are the numbers of good borrowers in bin i and the whole sample, respectively. Next, we replace the raw data with the calculated WoE values. Such a transformation helps to build a strict linear relationship with log-odds, which is used later in logistic regression and can also handle missing values and outliers.

$$WoE_{i,j} = \ln \left(\frac{B_i/B_T}{G_i/G_T} \right)$$

The second step is to use information value to select informative independent variables, or features. Information value (IV) comes from information theory. It measures the predictive power of independent variables and is used in feature selection and analysis of feature importance. In general, variables with IV less than 0.02 can be dropped. Of our 76 variables, 32 were selected and entered into the logit model later.

$$IV_i = \ln \left(\frac{\frac{B_i}{B_T}}{\frac{G_i}{G_T}} \right) * \left(\frac{B_i}{B_T} - \frac{G_i}{G_T} \right)$$

The third step is to fit a logistic regression model using the newly transformed WoE data. Assume we have N dependent variables (N=32 in our paper).

$$y = \frac{1}{1 - e^{-\beta_0 + \sum_{j=1}^{N} \beta_j * WoE_j}}$$

where y takes 0 if there is no default, and 1 in the case of defaults, WoE_j is the transformed WoE for independent variable X_j.

Then we can use the logistic regression coefficients from model fitting and WoE values to scale the model into a scorecard. For each independent variable X_j, its corresponding score is

$$score_j = \left(\beta_j * WoE_j + \frac{\alpha}{n} \right) * factor + \frac{offset}{n}$$

where β_i is the logistic regression coefficient for the variable, α is the logistic regression intercept, WoE is the weight of evidence value for variable X_i, n is the number of independent variables, and factor and offset are scaling parameters. The total score is the sum of all the scores.

$$Total\ score = \sum_{i=1}^{n} score_i$$

We do not calculate the credit score in the analysis but use the logit model results to calculate expected default probability and plot the area under the receiver operating characteristics curve.

We use open-source package *scorecard* in R to train the scorecard models.

REFERENCES

Abdulsaleh, A. M., and A. C. Worthington. 2013. Small and Medium-Sized Enterprises Financing: A Review of Literature. *International Journal of Business and Management* 8:36–54.

Agarwal, S., S. Alok, P. Ghosh, and S. Gupta. 2019. Financial Inclusion and Alternate Credit Scoring for the Millennials: Role of Big Data and Machine Learning in Fintech. SSRN Scholarly Paper, Social Science Research Network, Rochester, NY.

Agarwal, S., and R. Hauswald. 2010. Distance and Private Information in Lending. *Review of Financial Studies* 23:2757–88.

Ahmed, M. S. I., and P. R. Rajaleximi. 2019. An Empirical Study on Credit Scoring and Credit Scorecard for Financial Institutions. *International Journal of Advanced Research in Computer Engineering & Technology* 8:275–79.

Bazarbash, M. 2019. Fintech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk. IMF Working Papers 19, International Monetary Fund, Washington, DC.

Berg, T., V. Burg, A. Gombović, and M. Puri. 2019. On the Rise of FinTechs: Credit Scoring Using Digital Footprints. *Review of Financial Studies*.

Berger, A. N., W. S. Frame, and V. Ioannidou. 2016. Reexamining the Empirical Relation between Loan Risk and Collateral: The Roles of Collateral Liquidity and Types. *Journal of Financial Intermediation* 26:28–46.

Berger, A. N., and G. F. Udell. 2002. Small Business Credit Availability and Relationship Lending: The Importance of Bank Organisational Structure. *Economic Journal* 112:32–53.

———. 2006. A More Complete Conceptual Framework for SME Finance. *Journal of Banking & Finance* 30:2945–66.

Besanko, D., and A. V. Thakor. 1987a. Collateral and Rationing: Sorting Equilibria in Monopolistic and Competitive Credit Markets. *International Economic Review* 28:671–89.

———. 1987b. Competitive Equilibrium in the Credit Market under Asymmetric Information. *Journal of Economic Theory* 42:167–82.

Butaru, F., Q. Chen, B. Clark, S. Das, A. W. Lo, and A. Siddique. 2016. Risk and Risk Management in the Credit Card Industry. *Journal of Banking & Finance* 72:218–39.

Cerqueiro, G., S. Ongena, and K. Roszbach. 2016. Collateralization, Bank Loan Rates, and Monitoring. *Journal of Finance* 71:1295–1322.

Cornée, S. 2019. The Relevance of Soft Information for Predicting Small Business Credit Default: Evidence from a Social Bank. *Journal of Small Business Management* 57:699–719.

Demirgüç-Kunt, A., and L. Klapper. 2013. Measuring Financial Inclusion: Explaining Variation in Use of Financial Services across and within Countries. *Brookings Papers on Economic Activity* 2013:279–340.

Demirgüç -Kunt, A., L. Klapper, D. Singer, S. Ansar, and J. Hess. 2018. *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. Washington, DC: World Bank.

Freel, M., S. Carter, S. Tagg, and C. Mason. 2012. The Latent Demand for Bank Debt: Characterizing "Discouraged Borrowers." *Small Business Economics* 38:399–418.

Frost, J., L. Gambacorta, Y. Huang, H. S. Shin, and P. Zbinden. 2019. BigTech and the Changing Structure of Financial Intermediation. BIS Working Papers 779, Bank for International Settlements, Basel, Switzerland.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. 2020. Predictably Unequal? The Effects of Machine Learning on Credit Markets. SSRN Scholarly Paper, Social Science Research Network, Rochester, NY.

Gambacorta, L., Y. Huang, H. Qiu, and J. Wang. 2019. How Do Machine Learning and Non-Traditional Data Affect Credit Scoring? New Evidence from a Chinese Fintech Firm. BIS Working Papers 834:24, Bank for International Settlements, Basel, Switzerland.

Hand, D. J., and M. J. Crowder. 2005. Measuring Customer Quality in Retail Banking. *Statistical Modelling* 5:145–58.

Holmström, B. 2018. Keynote speech at Toulouse School of Economics, October 10, Toulouse, France.

Hopper, M. A., and E. M. Lewis. 2002. Behavioural Scoring and Adaptive Control Systems. In L. C. Thomas, J. N. Crook, and D. B. Edelman (eds.), *Credit Scoring and Credit Control*. Oxford University Press.

Jagtiani, J., and C. Lemieux. 2019. The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform. *Financial Management* 48:1009–29.

Khandani, A. E., A. J. Kim, and A. W. Lo. 2010. Consumer Credit-Risk Models via Machine-Learning Algorithms. *Journal of Banking & Finance* 34:2767–87.

Kithinji, A. M. 2010. Credit Risk Management and Profitability of Commercial Banks in Kenya. School of Business, University of Nairobi, Kenya.

Miller, M., and D. Rojas. 2004. Improving Access to Credit for SMEs: An Empirical Analysis of the Viability of Pooled Data SME Credit Scoring Models in Brazil, Colombia & Mexico. Working Paper 22, World Bank, Washington, DC.

Siddiqi, N. 2012. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, volume 3. John Wiley & Sons.

———. 2017. *Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards*. John Wiley & Sons.

Sirignano, J., A. Sadhwani, and K. Giesecke. 2018. Deep Learning for Mortgage Risk. arXiv:1607.02470.

Stiglitz, J. E., and A. Weiss. 1981. Credit Rationing in Markets with Imperfect Information. *American Economic Review* 71:393–410.

Thomas, L. C., R. W. Oliver, and D. J. Hand. 2005. A Survey of the Issues in Consumer Credit Modelling Research. *Journal of the Operational Research Society* 56:1006–15.

Vos, E., A. J.-Y. Yeh, S. Carter, and S. Tagg. 2007. The Happy Story of Small Business FInancing. *Journal of Banking & Finance* 31:2648–72.