Journal of Financial Stability 73 (2024) 101284

Contents lists available at ScienceDirect

Journal of Financial Stability

journal homepage: www.elsevier.com/locate/jfs

How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm

Leonardo Gambacorta [a,b,*], Yiping Huang [c,d], Han Qiu [a], Jingyi Wang [d,e]

[a] Bank for International Settlements, Centralbahnplatz 2, Basel 4002, Switzerland
[b] Centre for Economic Policy Research (CEPR), London, UK
[c] National School of Development, Peking University, No. 5 Yiheyuan Road Haidian District, Beijing 100871, China
[d] Institute of Digital Finance, Peking University, No. 5 Yiheyuan Road Haidian District, Beijing 100871, China
[e] School of Finance, Central University of Economics and Finance, 39 South College Road Haidian District, Beijing 100081, China

ARTICLE INFO

JEL Classification:
G17
G18
G23
G32

Keywords:
Fintech
Credit scoring
Non-traditional information
Machine learning
Credit risk

ABSTRACT

This paper compares the predictive power of credit scoring models based on machine learning techniques with that of traditional loss and default models. Using proprietary transaction-level data from a leading fintech company in China, we test the performance of different models to predict losses and defaults both in normal times and when the economy is subject to a shock. In particular, we analyse the case of an (exogenous) change in regulation policy on shadow banking in China that caused credit conditions to deteriorate. We find that the model based on machine learning and non-traditional data is better able to predict losses and defaults than traditional models in the presence of a negative shock to the aggregate credit supply. This result reflects a higher capacity of non-traditional data to capture relevant borrower characteristics and of machine learning techniques to better mine the non-linear relationship between variables in a period of stress.

1. Introduction

Financial technology (fintech) is taking on an ever more important role in lending decisions, while lending by fintech companies is gaining a significant share of certain market segments. In the United States, for instance, online lenders accounted for about 8–12 % of new mortgage loan originations, with Quicken Loans being recognised as the country's largest mortgage lender in terms of flow at the end of 2017 (Buchak et al., 2018; Fuster et al., 2019). However, fintech credit is not limited to the US and reached USD 694 billion globally in 2018. China is a country where fintech credit is relatively well-developed, representing around 3 % of total outstanding credit to the non-bank sector at the end of 2019 (Cornelli et al., 2022). In recent years, the pace of growth of alternative form of credit such as big tech credit has been larger than that for bank credit. For instance, during 2020–21, big tech credit in China recorded an average annual growth rate of 37 %, compared to 13 % for bank credit (De Fiore et al., 2023).

New credit scoring models used by fintech lenders differ from traditional models in two key ways. The first is that technology allows financial intermediaries to collect and use a larger quantity of information. Fintech credit platforms may use alternative data sources, including insights gained from social media activity (Jagtiani and Lemieux, 2019; US Department of the Treasury, 2016) and users' digital footprints (Berg et al., 2020). In the case of large technology companies (big techs) with existing platforms, data collection extends to orders, transactions and customer reviews (Frost et al., 2019).[1]

The second difference is the adoption of machine learning techniques. In contrast to traditional linear models such as the logit model, machine learning can mine the non-linear information from variables. For example, Khandani et al. (2010) construct a non-linear, non-parametric forecasting model for consumer credit that is based on machine learning techniques and find that this new model can outperform other models in a range from 6 % to 25 % of total losses. However, the prediction capability of machine learning models has mainly been demonstrated in applications with a stationary external environment. Their performance also needs to be assessed in the case of a structural shock that changes the main relationships between the variables.

This paper contributes to the literature by addressing the following

* Corresponding author at: Bank for International Settlements, Centralbahnplatz 2, Basel 4002, Switzerland
E-mail address: leonardo.gambacorta@bis.org (L. Gambacorta).

[1] For a discussion on the differences in credit scoring models between fintech companies and big techs see Annex A.

https://doi.org/10.1016/j.jfs.2024.101284
Received 18 August 2022; Received in revised form 17 May 2023; Accepted 6 September 2023
Available online 4 June 2024

four questions:

i) Are machine learning-based fintech credit scoring models better able to predict borrowers' losses and defaults than traditional empirical models?
ii) What is the information content of non-traditional sources such as digital applications on mobile phones and e-commerce platform data?
iii) How do the different models perform in the event of an (exogenous) shock?
iv) How do the different models perform for customers with a limited credit history?

The first two questions have also been analysed by other papers, with mixed results. Our contribution is mostly to highlight and explain differences in the results due to different institutional settings and database characteristics. The third and fourth questions are completely new and represent the main contribution of the paper.

To answer these four questions, we use a unique data set from a leading Chinese fintech company at loan-transaction level for the period between May and September 2017. The fintech firm has requested to remain anonymous but has given us access to a very comprehensive data set. Compared to previous studies, this data set allows us to disentangle the effects of traditional bank-type information (credit card information) and non-traditional information (obtained from the use of digital applications on mobile phones and e-commerce platforms). Moreover, we can assess the performance of the credit scores calculated by the fintech company using machine learning methods and such large volumes of data. Papers based on data from Renrendai, a Beijing-based company providing P2P financial services (see, for example, Braggion et al., 2022) cannot use credit card transaction information because Renrendai's borrowers typically did not have a current account with a bank.[2]

Furthermore, unlike other fintech companies, in which borrower information is self-reported by the users themselves (see for example Berg et al., 2020), our fintech company is able to read both credit card transactions and digital app information directly from the system (with the user's permission). The information is therefore collected more comprehensively to include both credit card information and additional non-traditional information.

We analyse personal loans, most of which are repayable in up to one year. We also observe the borrowers' repayment record until October 2018 in order to track the status (viable or defaulted) of each loan after origination. This enables us to evaluate the performance of each loan ex post in terms of losses and defaults.

In order to answer the third question, we analyse the effects of a largely unexpected regulatory change that occurred in China in the period under review. On 17 November 2017, the People's Bank of China (PBoC) – the Chinese central bank – issued specific draft guidelines to tighten regulations on shadow banking. This regulatory change has led many financial intermediaries to increase their lending requirements, causing credit conditions for borrowers to deteriorate. In particular, the aggregated data indicate a significant increase in the default rate and a drop in lending after the shock. A similar pattern can be observed at our fintech company, which enables us to study how the different models performed during this stress period.

The main conclusions of our paper can be summarised as follows:

i) The fintech's machine learning-based credit scoring models outperform traditional empirical models (using both traditional and non-traditional information) in predicting borrowers' losses and defaults.
ii) Non-traditional information enhances the predictive power of the model, resulting in a 2.2 % increase of the AUROC.
iii) While the models perform similarly well in normal times, the model based on machine learning is better able to predict losses and defaults following a negative shock to the aggregate credit supply. One possible reason for this is that machine learning can better mine the non-linear relationship between variables in the event of a shock.
iv) The predictive power of all the models improves when the length of the relationship between bank and customer increases. However, the comparative advantage of the model that uses the fintech credit scoring technique based on machine learning tends to decline when the length of the relationship increases.

2. Literature review

A few studies have started to analyse how credit supplied by fintech firms and their scoring models perform compared with traditional bank lending. Jagtiani and Lemieux (2019) compare loans made by a large fintech lender and similar loans that were originated through traditional banking channels. Specifically, they use account-level data from LendingClub and the Y-14 M data reported by bank holding companies with total assets of $50 billion or more. They find a high correlation between interest rate spreads, LendingClub rating grades and loan performance. Interestingly, the correlations between the rating grades and FICO scores declined from about 80 % (for loans that were originated in 2007) to only about 35 % for later vintages (originated in 2014–2015), indicating that LendingClub has increasingly used non-traditional alternative data.

Using market-wide, loan-level data on US mortgage applications and originations, Fuster et al. (2022) show that fintech lenders process mortgage applications about 20 % faster than other lenders, even when controlling for detailed loan, borrower and geographic observables. It is interesting to note that faster processing does not come at the cost of higher defaults. Furthermore, fintech lenders adjust their supply more elastically than other lenders in response to exogenous mortgage demand shocks, thereby alleviating the capacity constraints associated with traditional mortgage lending. Buchak et al. (2018) compare the pricing of online (fintech) lenders in the US mortgage market with the pricing of banks and (non-fintech) shadow banks; they find that fintech lenders charge a premium of 14–16 basis points over bank mortgages. Jagtiani et al. (2019) find that fintech lenders in the United States tend to supply more mortgages to consumers with weaker credit scores than do banks; they also have greater market shares in areas with lower credit scores and higher mortgage denial rates.

While banks usually incentivise borrowers to pay their loans back by requiring them to pledge tangible assets (eg real estate) as collateral, fintech credit is typically uncollateralised. This makes the use of big data particularly relevant when considering a loan application. Using credit data for China, Gambacorta et al. (2023) suggest that big data can act as a substitute for collateral and that the volume of corporate loans supplied by big techs does not correlate with asset prices, whereas bank loans do.

Frost et al. (2019) use data for Mercado Credito, which provides credit lines to small firms in Argentina on the e-commerce platform Mercado Libre. They find that, when it comes to predicting loss rates, credit scoring techniques based on big data and machine learning have so far outperformed credit bureau ratings. A key question here is whether this outperformance will persist through a full business and financial cycle. Indeed, fintech credit could give rise to new forms of non-prudent risk-taking that needs to be tested in the event of an adverse shock. For example, De Roure et al. (2016) find that online lenders in Germany substitute bank loans for high-risk consumer loans. In US consumer credit markets, Tang (2019) finds that online lending acts as a substitute for bank lending among marginal borrowers, while

[2] The whole fintech sector in China was significantly affected by a regulatory crackdown on P2P lenders at the end of 2017, which is discussed in Section 5. For additional information on the development of fintech and big tech credit in China see Annex B.

complementing bank lending for small loans. It is interesting to observe that the performance of online lenders seems to rely on the quantity and quality of information available to them.

Some of the literature looks at the informational content of digital soft information and credit performance. Dorfleitner et al. (2016) study the relationship between soft factors in P2P loan applications and financing and default outcomes. Using data on the two leading European P2P lending platforms, Smava and Auxmoney, they find that soft factors influence the funding probability but not the default probability. Jagtiani and Lemieux (2019) find that the rating grades assigned on the basis of alternative data perform well in predicting loan performance over the two years after origination. The use of alternative data has allowed some borrowers who would have been classified as subprime by traditional criteria to be slotted into "better" loan grades, enabling them to benefit from lower priced credit. In addition, for the same risk of default, consumers pay smaller spreads on loans from LendingClub than from credit card borrowing. Berg et al. (2020) show that digital footprints are a good predictor of the default rate. Analysis of simple, easily accessible variables from digital footprints is equal to or better than the information from credit bureau scores. Iyer et al. (2016) analyse the performance of new online lending markets that rely on nonexpert individuals to screen their peers' creditworthiness. They find that these peer lenders predict an individual's likelihood of defaulting on a loan with 45 % greater accuracy than the borrower's exact credit score (unobserved by the lenders, who only see a credit category). The results show how aggregating over the views of peers and leveraging nonstandard information can enhance lending efficiency.

A related stream of the literature looks at how contract and relationship characteristics could influence loans' default. Hertzberg et al. (2018) find that when long maturity is available, fewer borrowers take the short-term loan, and those that do, default less. This is consistent with theories of insurance markets with private information that indicate that maturity choice leads to adverse selection. Additional findings suggest borrowers self-select on private information about their future ability to repay. The findings imply that maturity can be used to screen borrowers on this private information. Puri et al. (2017) evaluate the impact of many aspects of customer–bank relationships on loan default rates. They define relationships in many different ways to capture non-credit relationships, transaction accounts, as well as the depth and intensity of relationships, and find each of these can provide information that helps reduce default. The results show that banks with relationship-specific information act differently compared with banks that do not have this information both in screening and subsequent monitoring borrowers which helps reduce loan defaults. Berg et al., (2020) show that loan officers who face volume-based incentives significantly manipulate ratings even in settings where ratings are computed using hard information only. They concludes that reliance on hard information does not overcome loan officer agency problems, and it is important for banks and regulators to take manipulation of hard information into account when using hard information for risk assessment and regulation.

Another stream of the literature analyses *unfair* price discrimination. In particular, sophisticated machine learning algorithms may not be as neutral as their mathematical nature suggests at first glance. Even though artificial intelligence and machine learning algorithms are neither trained nor fed with protected characteristics such as race, religion, gender or disability, they are able to triangulate such information. Using data on US mortgages, Fuster et al. (2019) find that the introduction of machine learning in credit scoring models disproportionately benefits White Non-Hispanic and Asian borrowers, while Black and Hispanic borrowers are less likely to benefit. This finding suggests that the algorithm may have differential effects across groups and potentially contribute to increased inequality. In particular, the results indicate that approximately 65 % of White Non-Hispanic and Asian borrowers benefit from the introduction of machine learning in credit scoring models, compared to around 50 % of Black and Hispanic borrowers.

3. Data description

We access data on proprietary loan transactions from a leading fintech company in China for the period between May 2017 and September 2017.[3] To obtain credit through the platform, customers need to provide the fintech company with bank credit card information and additional non-traditional information (via platform services). For each customer, the fintech company calculates a credit score that assesses risk on the basis of machine learning technology and information provided by customers (via their credit card transactions, the digital apps on their phones and e-commerce platform data).[4]

In our analysis, we will try to disentangle the information content of the credit score, the credit card information (which is typically what a traditional bank observes) and non-traditional information (accessible via social media and platforms use).

The fintech company decides whether to grant a loan or not on the basis of the fintech score. To this end, we consider the credit score at the date of origination for each loan. At the same time we consider traditional and non-traditional information available at the time the loan is originated.

The fintech company provides personal loans with a maturity of up to 24 months, although the vast majority (more than 80 %) mature after one year. In order to analyse performance, we also access the loan repayment records up to October 2018 to let us evaluate loan defaults and the losses incurred by the fintech firm.

Our database includes all the 343,976 loans financed via the fintech firm in the period May-September 2017. The fintech company provides a portal for individual borrowers to apply for a loan and for individual lenders to view the characteristics of the different products assembled by the fintech firm. There is no one to one matching between a lender and a borrower. The fintech company divides the loans and packages them in different bundles by maturity, risk and return. Investors are unaware of the specific details of the borrowers in each bundle and choose their preferred product by maturity, risk and return. Following Berg et al. (2020), we test the representativeness of our dataset at the geographical level. In particular, we compare the geographical distribution of our database of fintech credit at the province level with those of total bank credit (see Fig. 1). The correlation between the shares of fintech loans and bank loans is around 90 percent.[5]

Table 1 provides descriptive statistics of the data set. Some of the variables seem quite skewed, with rather extreme outliers. For instance, the average frequency of credit card usage over the past year is 6.65, but the max is at 2637. Similarly, there appear to be some people with extremely high numbers of defaults on their credit card (e.g. 31 over the past 3 months) and the max repayment is RMB 1.3 million, while the average is 15. Such skewness in the variables could lead the simple "linear" models to do worse than if the independent variables had more "normal" distribution. Therefore, we will use winsorised variables, at

[3] We chose not to use the volume of credit extended by the fintech company in October 2017, prior to the regulatory shock, for two reasons. First, we aimed to avoid any criticism that the fintech company may have had foreknowledge of the impending regulatory change in November. Second, the fintech company implemented modifications to its credit score and risk assessment system in October 2017, resulting in a potential discontinuity in the analysis. It is important to note that our analysis of credit quality, which relies on loans supplied between May 2017 and September 2017, remains unaffected by the change in credit scoring that occurred in October 2017.

[4] The fintech company used a decision tree approach applied to a database of 300 variables. The company ended up using 20 variables to calculate fintech credit scores. One of the reasons why they did not include all 300 variables was to avoid an overfitting problem.

[5] Similar conclusions are found by comparing the geographical distribution of our database with that of population in each province.

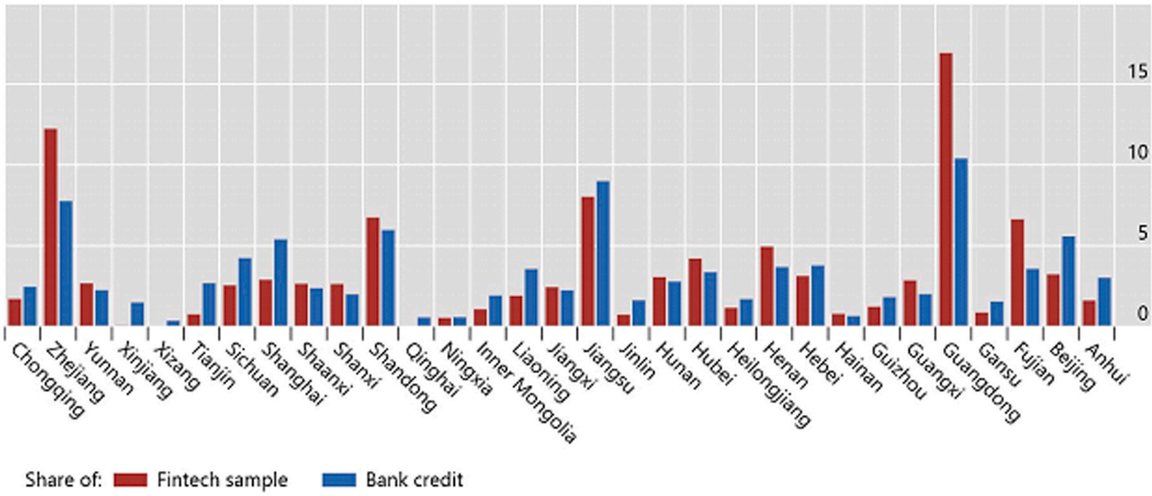

Fig. 1. Geographic distribution of fintech credit and bank credit in our sample In per cent. This figure illustrates the share of loans by province in our sample compared to the total bank loan distribution by province. The sample period is from 1 May to 30 September 2017. Source: Authors' calculations.

the 1 % level in the regressions.

4. A horse race between different credit scoring models

4.1. Empirical strategy

Our first goal is to assess whether fintech credit scoring models (based on machine learning plus big data) are better able to predict borrowers' losses and defaults than linear models (based on traditional and non-traditional data).

We start by estimating different models to predict total losses:

$$L_{i,t} = \alpha CS_{i,t} + \mu_P + \mu_T + \varepsilon_{i,t} \tag{1}$$

$$L_{i,t} = \beta X_{i,t} + \mu_P + \mu_T + \varepsilon_{i,t} \tag{2}$$

$$L_{i,t} = \beta X_{i,t} + \delta Y_{i,t} + \mu_P + \mu_T + \varepsilon_{i,t} \tag{3}$$

where $L_{i,t}$ indicates the loss rate (as a percentage of the origination volume) on a loan. The first information set includes the fintech credit score for borrower i at time t ($CS_{i,t}$). The second information set includes a vector of variables obtained through the credit card ($X_{i,t}$). This set of traditional information is typically also available to a bank. The third set of information also includes a vector of non-traditional variables ($Y_{i,t}$) obtained by the fintech company through customers' mobile phone apps and their activity on the e-commerce platform. All models include province (μ_P) and time (μ_T) fixed effectsand $\varepsilon_{i,t}$ is an error term. Eqs. (1) to (3) are estimated using a tobit model given the censored nature of data (either 0 or positive).

The second set of equations are

$$p(D_{i,t}) = \Phi(\alpha CS_{i,t} + \mu_P + \mu_T + \varepsilon_{i,t}) \tag{1'}$$

$$p(D_{i,t}) = \Phi(\beta X_{i,t} + \mu_P + \mu_T + \varepsilon_{i,t}) \tag{2'}$$

$$p(D_{i,t}) = \Phi(\beta X_{i,t} + \delta Y_{i,t} + \mu_P + \mu_T + \varepsilon_{i,t}) \tag{3'}$$

where $p(D_{i,t})$ indicates the probability for the borrower of a loan to default (and to generate a loss). Eq. (1') to (3') are estimated using a logit model, which is more appropriate than probit models for large sample sizes.

To sum up, we consider three different models with different information sets. Model I only uses the fintech score as the independent variable, while Model II only uses the traditional bank-type information set as independent variables. Model III includes both traditional and non-traditional information as independent variables. We need to stress that for Models II and III we use the same explanatory variables as are used in the machine learning model (13 traditional and 7 non-traditional variables). These explanatory variables were selected from more than 300 series, using a data selection process based on their highest predictive power.[6]

It is worth emphasizing that in the "horse race" between the three models, the comparison is not completely one-for-one. In a sense, the fintech credit score (Model I) is tested "out of sample", while Models II and III are estimated "in sample". So this would in principle produce a bias against Model I. On the other hand, Model I uses more data for training than the data used in Models II and III, so that may be one reason for its better performance. In other words, in Models II and III we use the same set of data selected to be used for the machine learning model, under the assumption that they would be also the best ones for the linear models. We will address some of these points in the robustness check section.

4.2. Results

Table 2 presents the results of Eqs. (1) to (3) that consider the three different information sets. The model in the first column uses only the fintech score as the independent variable (Model I), while the model in the second column provides the result using the traditional credit card information as independent variables (Model II), and the third column provides the result using all variables (Model III). All models are estimated using a Tobit regression model. The fintech score is a highly significant predictor of the loss rate. The credit card/bank and non-traditional information are also useful. However, the pseudo R^2 of Model I (0.0367) is almost double that of Model III (0.0217).

Table 3 has a similar structure to Table 2, but presents the estimates of Eq. (1') to (3'), ie the probability that a customer will default. All the models in Table 3 are estimated using a logistic regression model. Consistent with Table 2, Model I – which only uses the fintech credit score – has the highest pseudo R^2, 0.0399. Model II has a pseudo R^2 of

[6] For instance, the popular lasso belongs to this class of estimators that produce sparse representations of predictive models (see Belloni et al. 2011 for a recent survey and examples of big data applications of these methodologies in economics). By contrast, Giannone et al. (2018) point to the need to use dense-modelling techniques that recognise that all possible explanatory variables might be important for prediction, although their individual impact might be small.

Table 1
Descriptive statistics.

Variables	Obs	Mean	Std dev	Min	Max
Default dummy (0/1)	310,919	0.16	0.37	0	1
Loss rate	310,919	0.13	0.27	0	1
Fintech credit score	310,919	623.84	30.18	576	815
Interest rate (annualised)	310,919	47.59	12.91	29.87	101.68
Number of bank accounts used in the last 3 months	310,919	0.65	0.66	0	13
Number of bank accounts used in the last 12 months	310,919	4.72	2.70	0	18
Frequency of credit card usage in the last 3 months	310,919	1.54	6.81	0	494
Frequency of credit card usage in the last 12 months	310,919	6.74	22.50	0	2637
Number of transactions over 5000 RMB in the past 12 month	310,919	36.77	53.57	0	3155
Credit line (RMB)	310,919	41591.05	40206.2	0	3500,000
Credit card defaults in the past 3 months[1]	310,919	0.05	0.34	0	31
Credit card defaults in the past 12 months[1]	310,919	0.40	1.16	0	59
Repayments (RMB)	310,919	15.75	3077.98	0	1281,800
Credit history (months)	310,919	26.60	17.82	0	126
Salary deposited in current account	310,919	931.03	8385.18	0	1500,000
Gender (0=male)	310,919	0.24	0.43	0	1
Telephone call duration in the last 12 months	310,919	2904.11	1807.58	0	10157
Number of calls with family members in the last 12 months	310,919	306.01	456.11	0	2340
Number of calls in the last 3 months	310,919	1149.41	854.87	0	4329
Average number of calls per day in the past 12 months	310,919	6.50	4.34	0	22.22
Purchase (in RMB) on Taobao platform in the past 12 months	310,919	542.09	1229.31	0	6914.92
Defaults (Taobao)[2]	310,919	0.0003	0.0221	0	5

[1] Number of credit card defaults. [2] Number of times a borrower has not paid/delivered goods on the Taobao e-commerce platform.

Table 2
Loss rate regressions.

	Dependent variable: Loss rate		
Variables	I. Fintech credit score	II. Traditional information only	III. All information
Fintech credit score	-0.00845***		
	(8.30e-05)		
Traditional information			
Number of bank accounts used in the last 12 months		-0.00198**	-0.00195*
		(0.00100)	(0.00100)
Frequency of credit card usage in the last 12 months		-2.14e-05	-0.00012
		(0.00015)	(0.00015)
Frequency of credit card usage in the last 3 months		-0.00076	-0.00067
		(0.00048)	(0.00047)
Number of transactions over 5000 RMB in the past 12 month		-0.00126***	-0.00100***
		(5.87e-05)	(5.80e-05)
Credit line (RMB)		-2.97e-07***	-1.52e-07**
		(6.78e-08)	(6.69e-08)
Credit card defaults in the past 12 months		0.0121***	0.0159***
		(0.00197)	(0.00196)
Credit card defaults in the past 3 months		0.00978	0.0117*
		(0.00674)	(0.00669)
Repayments (RMB)		-3.60e-07	-4.32e-07
		(7.71e-07)	(7.96e-07)
Number of bank accounts used in the last 3 months		0.0140***	0.0155***
		(0.00347)	(0.00351)
Credit history (months)		-0.00624***	-0.00600***
		(0.00015)	(0.00015)
Salary deposited in current account		-4.55e-06***	-4.64e-06***
		(5.13e-07)	(5.21e-07)
Non-traditional information[1]			
Telephone call duration in the last 12 months			-3.55e-10***
			(1.16e-10)
Number of calls with family members in the last 12 months			2.97e-06
			(4.70e-06)
Average number of calls per day in the past 12 months			-0.0191***
			(0.00098)
Purchase (in RMB) on Taobao platform in the past 12 months			-1.69e-05***
			(1.64e-06)
Other variables			YES
Observations	310,919	310,919	310,919
Pseudo R^2	0.0367	0.0169	0.0217

[1] The model with non-traditional information also includes the number of defaults on the Taobao platform, the number of calls in the last three months and gender. All models include monthly and province fixed effects.

Table 3
Default rate regressions.

Variables	Dependent variable: Default rate I. Fintech credit score	II. Traditional information only	III. All information
Fintech credit score	-0.0178***		
	(0.00018)		
Traditional information			
Number of bank accounts used in the last 12 months		0.00111	0.00083
		(0.00209)	(0.00211)
Frequency of credit card usage in the last 12 months		0.00012	-6.65e-05
		(0.00031)	(0.00031)
Frequency of credit card usage in the last 3 months		-0.00206**	-0.00195*
		(0.00101)	(0.00102)
Number of transactions over 5000 RMB in the past 12 month		-0.00283***	-0.00225***
		(0.00013)	(0.00013)
Credit line (RMB)		-7.97e-07***	-5.09e-07***
		(1.47e-07)	(1.46e-07)
Credit card defaults in the past 12 months		0.0236***	0.0309***
		(0.00406)	(0.00405)
Credit card defaults in the past 3 months		0.0209	0.0252*
		(0.0138)	(0.0138)
Repayments (RMB)		-6.20e-07	-7.20e-07
		(1.62e-06)	(1.69e-06)
Number of bank accounts used in the last 3 months		0.0371***	0.0409***
		(0.00722)	(0.00734)
Credit history (months)		-0.0123***	-0.0118***
		(0.00031)	(0.00032)
Salary deposited in current account		-1.06e-05***	-1.12e-05***
		(1.25e-06)	(1.29e-06)
Non-traditional information[1]			
Telephone call duration in the last 12 months			-7.76e-10***
			(2.88e-10)
Number of calls with family members in the last 12 months			1.07e-05
			(1.01e-05)
Average number of calls per day in the past 12 months			-0.0397***
			(0.00212)
Purchase (in RMB) on Taobao platform in the past 12 months			-3.67e-05***
			(3.70e-06)
Observations	310,910	310,910	310,910
Pseudo R^2	0.0399	0.0180	0.0231

[1] The model with non-traditional information also includes the number of defaults on the Taobao platform, the number of calls in the last three months and gender. All models include monthly and province fixed effects.

0.0180, while model III has a pseudo R^2 of 0.0231.[7]

Fig. 2 and the first panel of Table 4 present a comparison between the three models with different information sets. Fig. 2 shows the receiver operating characteristics (ROC) curve of each model. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The TPR is also known as sensitivity. The FPR is also known as the fall-out or probability of false alarm and can be calculated as (1 − specificity). The first panel of Table 4 reports the area under the ROC curve (AUROC) for every model. The AUROC is a widely used metric for judging the discriminatory power of credit scores. The AUROC ranges from 50 % (purely random prediction) to 100 % (perfect prediction). The formal test on the difference in performance across the models can be done comparing the 95 % confidence interval reported in the last column of the first panel of Table 4.

The left-hand panel of Fig. 2 reports the results for the three different models. The results show that Model I performs better than the other two models. Model III is the second-best. Model II performs worst. This means that the model based on the fintech credit score (Model I) is better than the traditional model that use bank-type information (Model II) in predicting default rates for this sample of borrowers. But Model I, which uses machine learning techniques, is also superior to logit regressions that use also non-traditional information (Model III). The better performance of the fintech company in predicting defaults could depend on both: (i) specific selection of the variables that better fit Model I than Model II or III; and (ii) the use of machine learning techniques that are able to capture relevant non-linearities among the variables. The three models are statistically different at the 5 % level, as verified by the test at the bottom of Table 4. In terms of contribution of non-traditional data and machine learning to predictive power, non-traditional data contribute an additional 2.2 % increase in the AUROC (=(0.607–0.594)/0.594), while applying machine learning techniques provides a 5.3 % increase in the AUROC (=(0.639–0.607)/0.607).

We conducted three additional tests with a view to shedding further light on this result. First, we considered the distribution for the expected default rate for the three different models over both the whole original sample and the winsorised sample. The results reported in Figure C2 in Annex C indicate that Model I has a greater discriminatory power than Model II and III (i.e. the expected default rates encompass a larger set of plausible data).

Second, as many of the continuous variables used in the database are positively skewed we have re run the model by log transforming the winsorised variables. The second panel of Table 4 indicates very similar results.

Third, we performed similar tests using the information content of the interest rates. Interest rate is assigned in buckets and it is not very granular. In particular, as the interest rate is highly correlated with the

[7] It is worth noting that the fintech credit score remains significant also in a specification model that includes also traditional and non-traditional information. In this case the Pseudo R-square passes from 0.0399 to 0.0495.

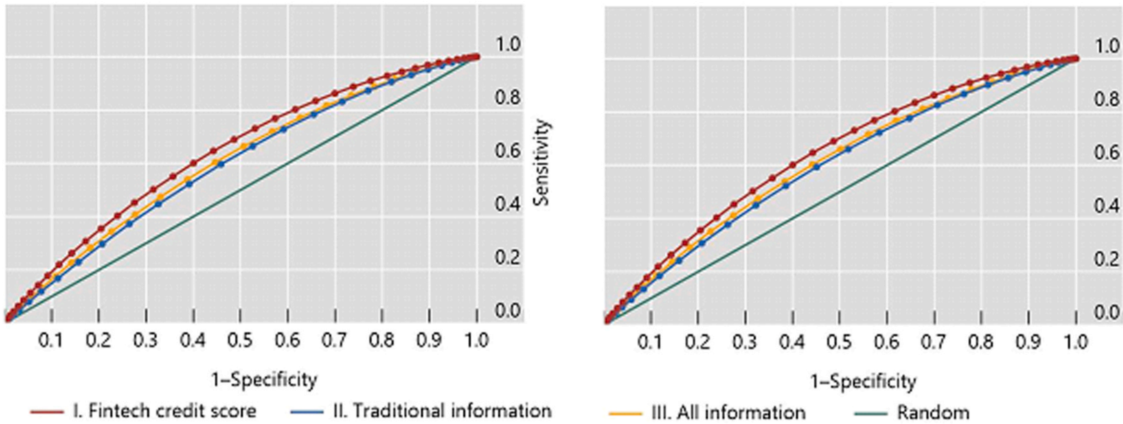

Fig. 2. ROC curves for different models, Source: Authors' calculations.

Table 4
A horse race between the three different models.

I. Baseline models				
	AUROC	Std err	95 % conf. interval	
I. Fintech credit score	0.6391	0.0012	0.63686	0.64143
II. Traditional information	0.5939	0.0012	0.59149	0.59621
III. All information	0.6070	0.0012	0.60462	0.60932
II. Models that use log transformed variables				
	AUROC	Std err	95 % conf. interval	
I. Fintech credit score	0.6397	0.0016	0.6367	0.6427
II. Traditional information	0.5669	0.0015	0.5639	0.5699
III. All information	0.6133	0.0016	0.6101	0.6164
III. Models that also include interest rate information				
	AUROC	Std err	95 % conf. interval	
I. Fintech credit score	0.6391	0.0012	0.63686	0.64144
II. Traditional information	0.5971	0.0012	0.59477	0.59951
III. All information	0.6095	0.0012	0.60712	0.61183
IV: Fintech credit score trained with 30 % of the sample in May-Sept 2017. Only borrowers which received credit in September.				
	AUROC	Std err	95 % conf. interval	
I. Fintech credit score	0.6986	0.0015	0.6887	0.7048
II. Traditional information	0.5911	0.0015	0.5811	0.6011
III. All information	0.6262	0.0015	0.6153	0.6370
V: Subsample of borrowers with roll over of credit rejected after the regulatory shock				
	AUROC	Std err	95 % conf. interval	
I. Fintech credit score	0.6186	0.0024	0.61385	0.62327
II. Traditional information	0.5744	0.0025	0.56962	0.57927
III. All information	0.5886	0.0025	0.58375	0.58375

fintech score, we have included the residual of a regression of the interest rate on the credit score in Models II and III. The test aims to control for the fact that the interest rate could take into account additional information not included in the list of explanatory variables but that the fintech company can use in its assessment. The other reason is that high interest rates could also affect repayment behaviour even with same risk. As can be seen from the right-hand panel of Fig. 2 and the third panel of Table 4, the results are qualitatively very similar.

The AUROC results of Model I can be compared with other existing studies that examine the information content of traditional and non-traditional data and performance of machine learning approaches. Figure C1 in the Annex C reports on the vertical axis the AUROC values for a number of retail datasets used in the literature. For example, Berg et al. (2020) find a higher AUROC than in our paper (0.736 vs 0.639), and a larger improvement when incorporating digital footprint data (0.053 vs 0.046).

A first reason for the lower performance of Model I in our paper is the high default rates (reported on the horizontal axis) registered in China after the regulatory shock (18.6 %), much larger than that experienced by the German ecommerce company analysed by Berg et al. (3.0 %). Looking at other studies, there is a clear negative relationship between model performance and average credit quality. For example, Iyer et al. (2016) that analyse the case of loans from Prosper in the US (default rate of 10.2 %), find a lower AUROC of 0.714. Figure C1 in the Annex C shows a clear negative correlation between annualized defaults rates (x-axis) and AUROC values (y-axis) for the different retail databases analysed by the literature.

A second reason for the lower performance of our Model I with respect to other papers in the literature is that we include only the fintech company credit score, under the hypothesis that the credit offer is based on this score. The other models reported in Figure A1 consider more complete specifications in which the credit score is included together with all other traditional and non-traditional information. If we include in Model I also (traditional and non traditional) information the AUROC increases to 0.6790.

A third reason is that the fintech credit score is typically trained on past data and could discount some information lags. Including all contemporaneous data in the model could boost the model performance. For this reason we have repeated the analysis with a random forest model trained on 30 % of the loans originated in the period May-September 2017 and including all available information (more than 100 variables). The results reported in fourth panel of Table 4 show that the AUROC curve for Model I increases to 0.698. The AUROC increases for Model II and Model III are only marginal compared to the baseline exercise reported in the first panel of Table 4.

5. Performance of the models in the event of an (exogenous) change in regulation

In finance, particularly in the context of credit, there is a key distinction compared to fields such as medicine and diagnostic procedures like Computed Tomography (CT) scans. While medical procedures are based on objective measurements and outcomes, financial relationships involve human actors who adapt their behaviour and expectations in response to changes in the economic environment. Therefore, the assumption that past repayment behaviour will consistently predict future repayment performance can be unreliable particularly when structural changes occur in the economy.

In this section, we want to test model performance in the event of an exogenous change in regulation. The current debate highlights one possible problem for machine learning models.[8] In particular, some of the literature stresses that the machine learning technology could only be useful in situations where the relationship between inputs and outputs remains the same – but this is often not the case in financial applications. We want to note that all the information in the model, including credit score, is taken before the negative shock, so the model outcomes and loan decisions are exogenous to the shock.

So far, machine learning, especially supervised learning, has been applied successfully in applications where there are stationary patterns. For example, when a CT scan is performed, we know that well trained doctors will make the same diagnosis every time they see a certain pattern in the scan. Application to financial data does not respond to this situation of stable correspondence. In credit scoring, for example, the relationship between the characteristics of the borrowers and whether they defaulted or are delinquent might not be stable at all. So the performance of machine learning models in stress situations remains to be fully explored.

In this section, we analyse the impact of a regulatory change in China on the performance of credit scoring models. On 17 November 2017, the PBoC issued draft guidelines to tighten regulations on financial institutions' asset management activities, a key component of the country's growing shadow banking sector. The main aim of the new rules, which affected $15 trillion of asset management products, was to unify regulatory practices across the financial industry. These changes were largely unexpected and caused a significant impact on fintech firms' business models. In particular, starting 17 November 2017, financial institutions have not been allowed to use asset management products to invest in commercial banks' credit assets or provide "funding services" for other institutions (such as fintech companies) to bypass regulations. The new rule had a huge impact on fintech companies' funding sources. The PBoC also set a limit on the interest rates charged by P2P lending companies. All annualised interest rates, which include the upfront fees charged for loans, were capped at 36 %. The effects of these new rules were reinforced on 1 December 2017 when China's Internet Financial Risk Special Rectification Work Leadership Team Office rolled out strict measures concerning online micro-lending.

The impact of the regulatory changes was to reduce loan supply, especially to riskier borrowers. Fig. 3 shows that the pace of growth in total credit in the Chinese economy fell by 4 percentage points in less than one year after these regulatory changes. Moreover, the sudden freeze on rolling over credit lines to risky borrowers caused many sole proprietorships to default. The histograms in Fig. 4 show that the default rate of the loans supplied by the fintech company analysed in this study (the number of the loans which defaults in each month divided by the number of total existing loans) increased by 3 percentage points at the end of December 2017, and then decreased smoothly to pre-shock levels after one year. Some of the borrowers might also have defaulted strategically, especially those with interest rates in excess of 36 % whose credit could not be rolled over at the same conditions. However, the lines in Fig. 4 indicate that the default rate of borrowers with an interest rate in excess of 36 % – despite being structurally higher because of the higher associated risk – followed a similar pattern to the default rate of the other borrowers.

We use the loan repayment records to evaluate the discriminatory power of the three models (Model I: fintech score; Model II: traditional information; Model III: all information) on a monthly basis. The discriminatory power is measured by the area under the ROC curve (AUROC), as depicted in Fig. 5. The vertical line represents the date of the regulatory shock. Our findings reveals that prior to the regulatory change, both Model I and Model III outperform Model II, and the difference between Model I and Model III is not statistically significant. After the regulatory shock, the discriminatory power of all three models decline. However, in relative terms, Model I exhibits superior performance compared to Model II and Model III.[9]

Fig. 6 shows the gap between the discriminatory powers of Model I (based on the credit scoring obtained using machine learning with big data) and Model II (traditional bank model). We decompose the gap into two parts. The first part (light blue) is the value added provided by non-traditional information (the gap between Model II and Model III). The second part (dark blue) is the gain obtained from machine learning technology (the gap between Model I and Model III).[10] Based on this graph, non-traditional information represents the main reason why Model I performs better than Model II prior to the shock. The contribution of machine learning technology is particularly relevant after the shock. This result could be due to the fact that machine learning technology can mine richer information from the variables during a period of stress. This may be due to the non-linearity of the model, which better captures dynamic relationships after the regulatory shock.

One concern could be that the above results only hold for the period of estimation (May 2017 – September 2017) and that different results could be obtained by estimating the logit models for a different period. As a robustness test, we therefore estimate the coefficient of the logit models for the period from January 2015 to December 2016 using a random sample of 10,000 customers of the fintech company. We then apply these coefficients to the explanatory variables of the borrowers in the period from May 2017 to September 2017 to verify any possible changes. The results reported in Fig. 7 indicate that, in relative terms, Model I performs better than Model III even in "normal" times, but the difference between the two models widens significantly after the regulatory shock.

Another possible concern regards the role of the gender variable in evaluating the differences between Model II and Model III. In the results presented so far, we have considered the gender variable in the non-traditional information. Banks may not include this variable in the set of traditional information out of concern for discrimination issues.[11] However, a borrower's gender is easily detectable and could be highly informative. We have therefore re-run the model by including the gender variable in traditional information. The results reported in Fig. 8 indicate that, also in this case, the model based on machine learning is better able to predict losses and defaults after the regulatory shock. However, the difference between Model II and Model III becomes less evident.

We also carried out some checks for loans with differing maturities. In particular, 80.3 % of the loans have a maturity up to 12 months, 5.4 % have a maturity between 12 and 18 months, and 14.3 % have a maturity between 18 and 24 months. As we evaluate loans originated in the period between May 2017 and September 2017, until October 2018 we are only able to observe the whole life cycle of credit maturing in up to one year. For this reason, we conducted the same structural break

[8] https://www.risk.net/risk-management/4120236/academics-warn-against-overuse-of-machine-learning#cxrecs_s

[9] We needed to consider how to include the time-fixed effects in the analysis of the explanatory power of the three different models in response to the exogenous regulatory shock. The results reported from Fig. 5 onwards include the average effects for the time dummies. Another possibility would have been to estimate the models up to month t and then make a prediction for month t+1 under the assumption that the month fixed effect in month $t+1$ is identical to the one in month t. The results obtained using this second assumption are very similar and not reported for the sake of brevity.

[10] The selection of the parameters to be used in the machine learning algorithm requires not only a knowledge of technical aspects but also experience in the selection of the appropriate weights and variables. In doing so, the technology officers may use also their own experience (soft information) in the evaluation. This means that this gap captures not only technological aspects but also experience (which is not easily replicable).

[11] Similarly the US Fair Housing Act (FHA) and Equal Credit Opportunity Act (ECOA) prohibit discrimination based on race, national origin, sex or religion.

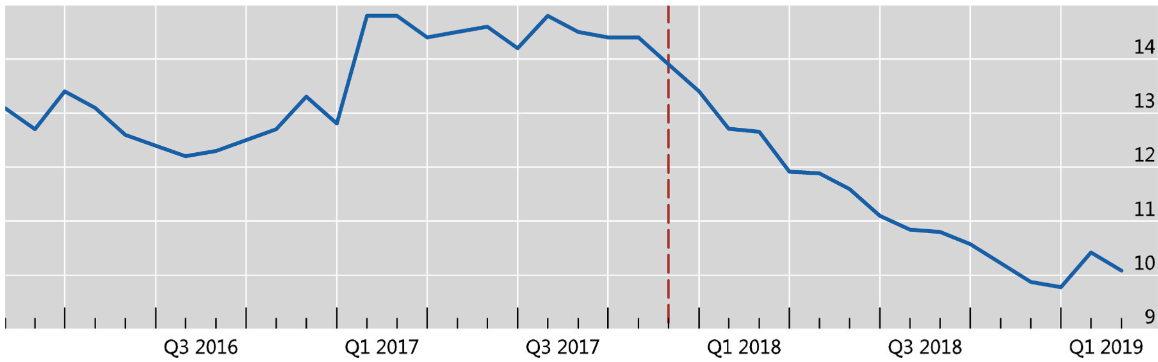

Fig. 3. Total credit to the Chinese economy (yearly credit growth). In per cent, The vertical dashed line indicates 17 November 2017, when the PBoC issued specific draft guidelines to tighten regulations on shadow banking. Source: The People's Bank of China.

Fig. 4. Default rate for the fintech company, In per cent, The vertical dashed line indicates 17 November 2017, when the PBoC issued specific draft guidelines to tighten regulations on shadow banking. Among these new rules, the PBoC set also a limit on the interest rates charged by P2P lending companies. All annualised interest rates, which include the upfront fees charged for loans, were capped at 36 %. This figure wants to analyse if those borrowers with credit contracts with an interest rate greater than 36 % reacted strategically and defaulted by more with respect to the others. Source: Authors' calculations based on an anonymous Chinese fintech company.

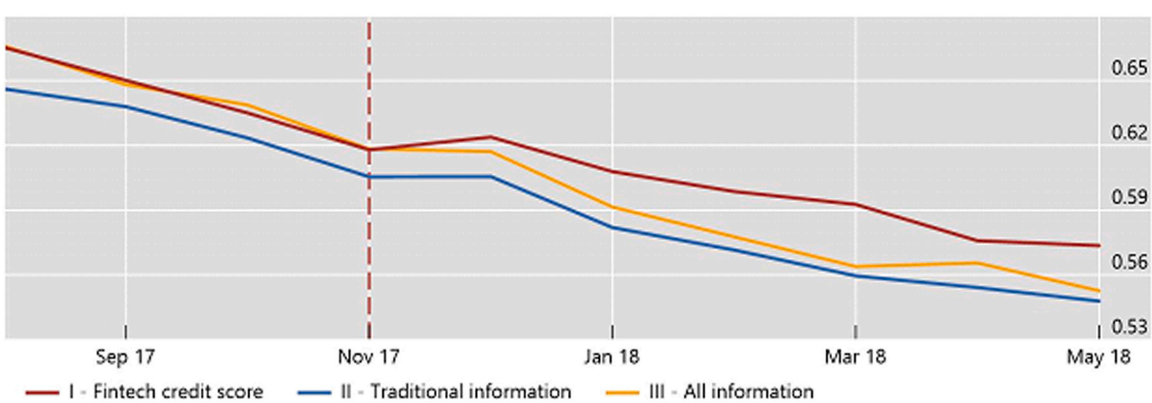

Fig. 5. Discriminatory power of the models before and after the regulatory shock, The vertical axis reports the area under the ROC curve (AUROC) for every model. The AUROC is a widely used metric for judging the discriminatory power of credit scores. The AUROC ranges from 50 % (purely random prediction) to 100 % (perfect prediction). The vertical dashed line indicates the date of the shock. In particular, it refers to a largely unexpected regulatory change that occurred in China on 17 November 2017, when the PBoC issued specific draft guidelines to tighten regulations on shadow banking. This regulatory policy has led many financial intermediaries to increase their lending requirements, resulting in deteriorating credit conditions for borrowers. Source: Authors' calculations.

analysis only for those loans with a maturity of up to one year. These loans are observed over their whole life. (The last credit line extended in September 2017 expires in October 2018.) The results of this test are reported in Fig. 9 and are qualitatively very similar. This is also in line with the statistical observation that most of the defaults take place in the first months of a loan contract. Table C1 in Annex C reports the performance of the three different models for loans with different contractual length (up to one year and more than one year). The results indicate a lower performance (reflecting higher default) for contracts with length more than one year confirming the finding in Hertzberg et al. (2018).

Based on the above checks, the result that the model based on machine learning is better able to predict losses and defaults after the regulatory shock seems quite robust. We look at this aspect by

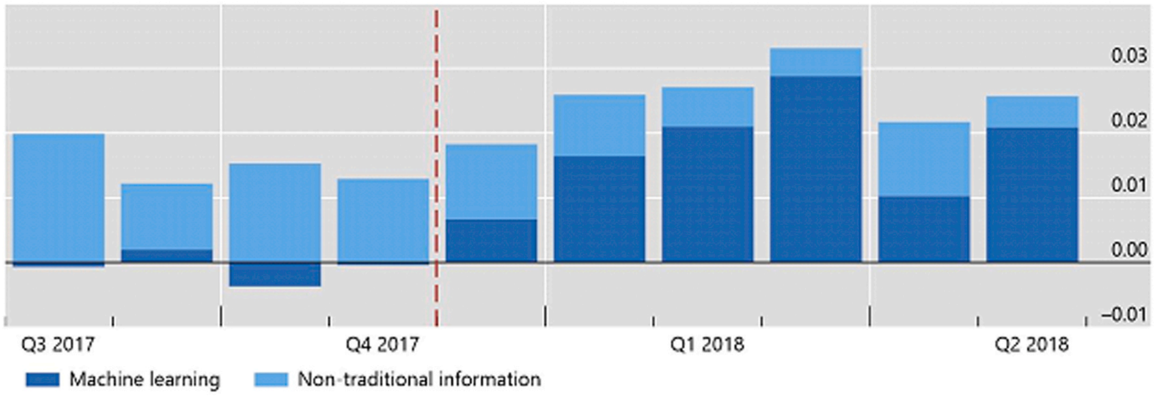

Fig. 6. The contribution of machine learning and non-traditional information, In per cent, The vertical dashed line indicates 17 November 2017, when the PBoC issued specific draft guidelines to tighten regulations on shadow banking. Source: Authors' calculations.

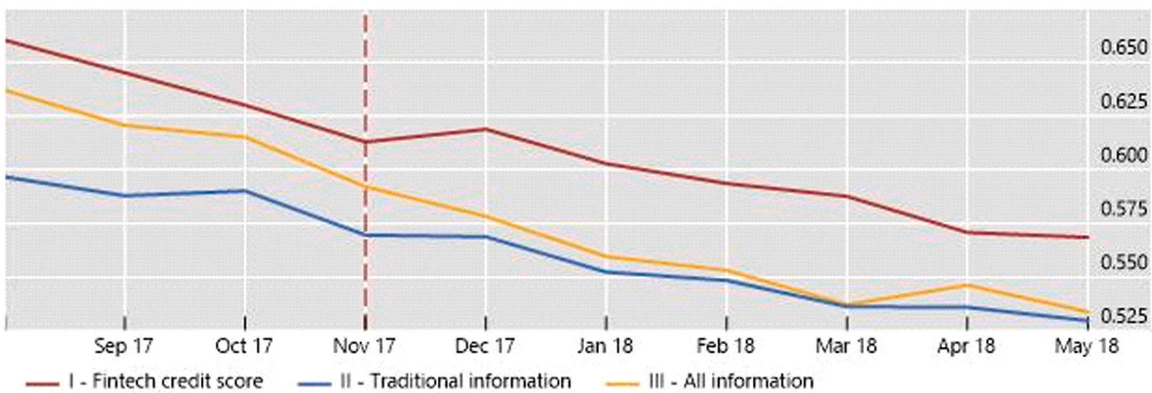

Fig. 7. Robustness check using a different estimation period for the logit models, The vertical axis reports the area under the ROC curve (AUROC) for every model. The AUROC is a widely used metric for judging the discriminatory power of credit scores. The AUROC ranges from 50 % (purely random prediction) to 100 % (perfect prediction). The vertical dashed line indicates the date of the shock. In particular, it refers to a largely unexpected regulatory change that occurred in China on 17 November 2017, when the PBoC issued specific draft guidelines to tighten regulations on shadow banking. This regulatory policy has led many financial intermediaries to increase their lending requirements, resulting in deteriorating credit conditions for borrowers.Source: Authors' calculations.

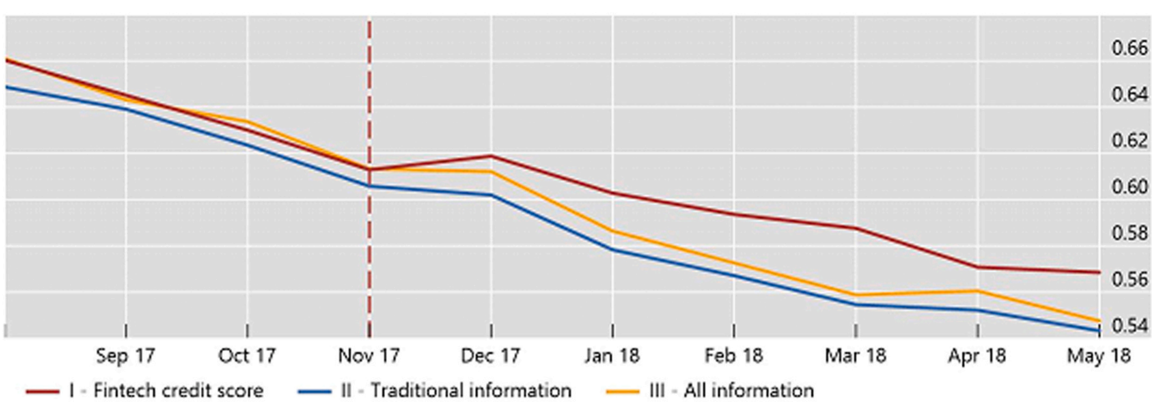

Fig. 8. Robustness check including gender among traditional information, The vertical axis reports the area under the ROC curve (AUROC) for every model. The AUROC is a widely used metric for judging the discriminatory power of credit scores. The AUROC ranges from 50 % (purely random prediction) to 100 % (perfect prediction). The vertical dashed line indicates the date of the shock. In particular, it refers to a largely unexpected regulatory change that occurred in China on 17 November 2017, when the PBoC issued specific draft guidelines to tighten regulations on shadow banking. This regulatory policy has led many financial intermediaries to increase their lending requirements, resulting in deteriorating credit conditions for borrowers. Source: Authors' calculations based on an anonymous fintech company data.

considering the loans that defaulted in the two-month window before and after the shock for the three different models. In particular, Fig. 10 plots the distribution of the three models' expected default rate before and after the regulatory shock. The sample prior to the shock includes the months of October and November 2017, while that after the shock covers December 2017 and January 2018. If the shock does not affect the model's predictive power, the distributions of the two samples should not be significantly different. Fig. 10 shows that the distributions

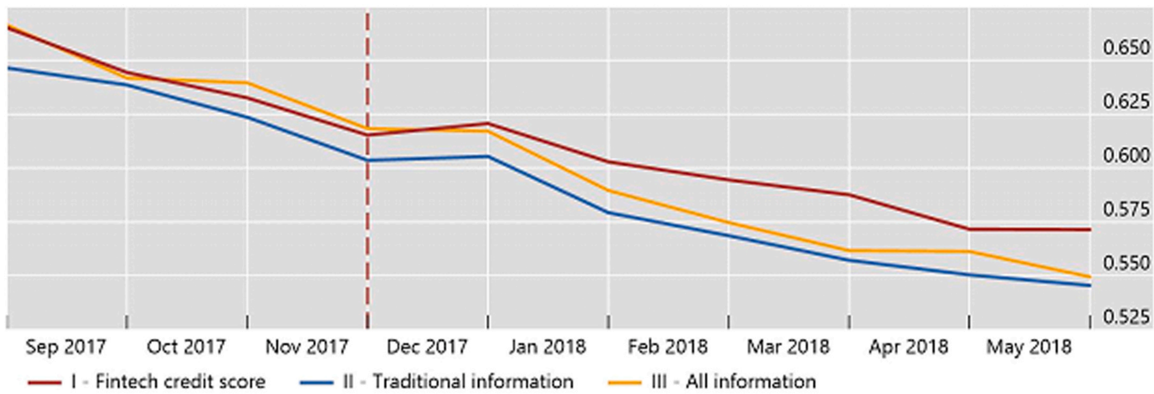

Fig. 9. Robustness check: loans up to one year, The vertical axis reports the area under the ROC curve (AUROC) for every model. The AUROC is a widely used metric for judging the discriminatory power of credit scores. The AUROC ranges from 50 % (purely random prediction) to 100 % (perfect prediction). The vertical dashed line indicates the date of the shock. In particular, it refers to a largely unexpected regulatory change that occurred in China on 17 November 2017, when the PBoC issued specific draft guidelines to tighten regulations on shadow banking. This regulatory policy has led many financial intermediaries to increase their lending requirements, resulting in deteriorating credit conditions for borrowers. Source: Authors' calculations based on an anonymous fintech company data.

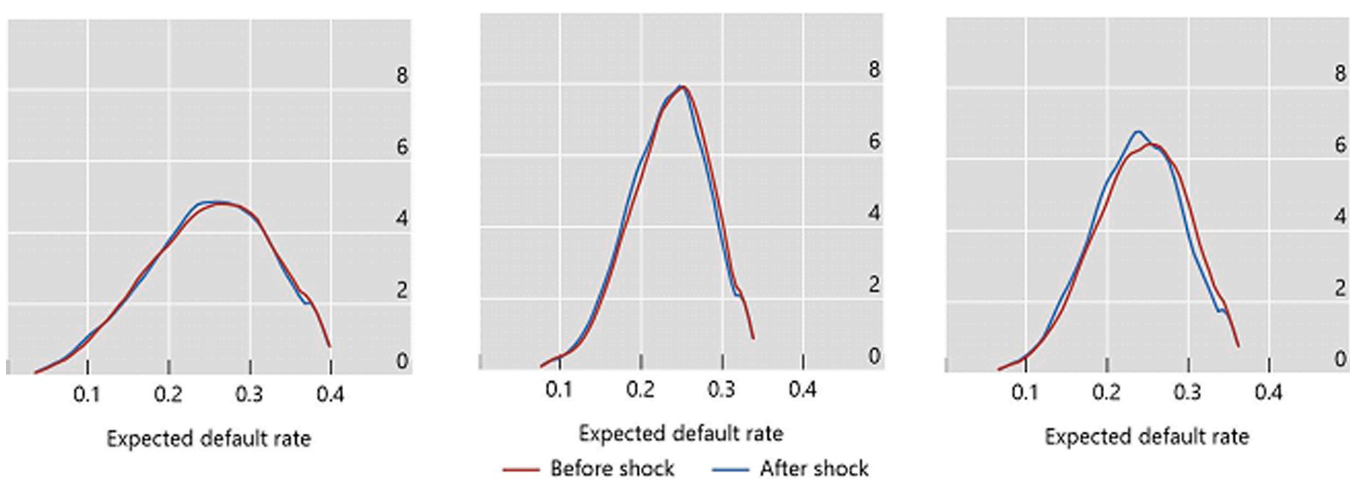

Fig. 10. The distribution of expected default rate before and after shock, Note: These graphs compare two-month windows of defaults before and after the shock for the three different models. The comparison makes it possible to see whether the distribution of expected default rates changes after the shock. If the distribution moves to the left, this means that the power of the model has reduced. Models tend to underestimate the probability of default after the shock. Source: Authors' calculations.

Table 5
Quantile regression before and after shock.

	Fintech credit score			Traditional information			All information		
Variables	q25	q50	q75	q25	q50	q75	q25	q50	q75
After shock	7.50e-05	-0.00172	-0.00145	-0.004***	-0.004***	-0.004***	-0.005***	-0.006***	-0.007***
	(0.00160)	(0.00114)	(0.00118)	(0.00086)	(0.00076)	(0.00072)	(0.00134)	(0.00086)	(0.00098)
Constant	0.197***	0.255***	0.307***	0.203***	0.240***	0.272***	0.202***	0.245***	0.285***
	(0.00117)	(0.00083)	(0.00098)	(0.00074)	(0.00075)	(0.00076)	(0.00134)	(0.00086)	(0.00102)
Observations	30,216	30,216	30,216	30,216	30,216	30,216	30,216	30,216	30,216

based on the fintech credit model are qualitatively very similar, while those based on Model II and Model III shift to the left after the shock. This means that, prior to the shock, these models were too optimistic regarding customers' capacity to repay their loans. A more precise evaluation is reported in Table 5 using the results of a quantile regression. The dependent variable is the expected default rate over the four months from October 2017 to January 2018. The right-hand side includes a dummy variable that takes the value of one in the post-shock period. For Models II and III, the dummy has a negative value for all quantiles, indicating that people who defaulted after the shock have a lower expected default rate based on traditional models. In other words, those people who have a higher evaluation based on the traditional model defaulted. This effect is not significant in the fintech credit model, indicating more stability.

6. Credit scoring and relationship lending

The provision of financial services (lending, insurance, wealth management etc) traditionally relies on trust and human interaction – they are relationship-based. By contrast, fintech lending is transaction-

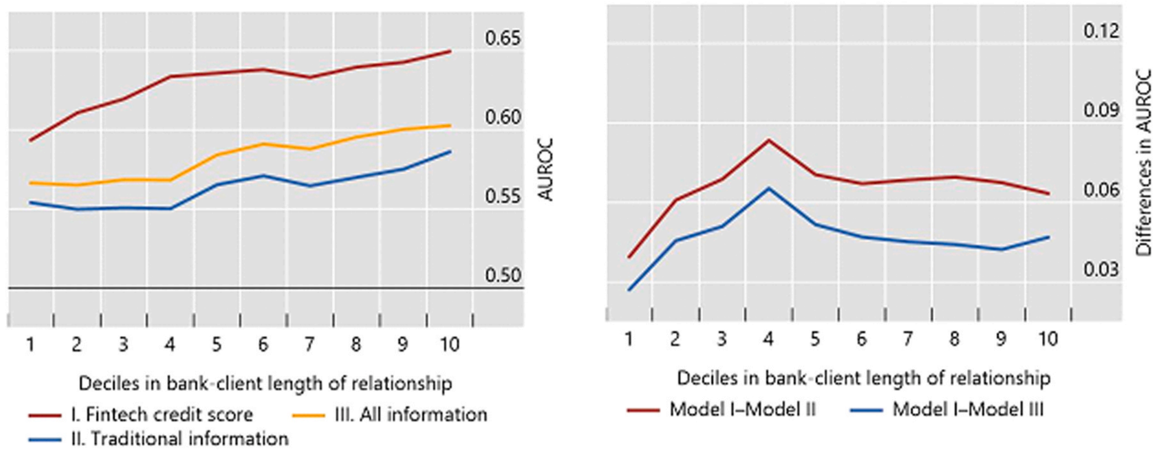

Fig. 11. Predictive power of the models and length of the bank-customer relationship, Source: Authors' calculations.

based and does not involve human intervention or a long-term relationship with the customer.

The loans offered by fintech lenders are strictly transactional, typically short-term credit lines that can be automatically cut if a customer's condition deteriorates. It is therefore interesting to study how the model's performance evolves for customers with different credit histories and compare with the existing literature (Degryse et al., 2009; Banerjee et al., 2021).

Fig. 11 highlights how the comparative advantage of the model that uses the credit scoring technique based on machine learning and big data could be modified by the length of the relationship between bank and customer. Please note that we use the length of the relationship between borrower and bank to calculate the borrower's credit history. This is because borrowers typically enter into a credit relationship with a bank first. In particular, we use the number of months from the opening of the bank account as a proxy of the bank-customer relationship. The average number of months is 27 with a standard deviation of 17. We have divided the sample into ten deciles according to the length of the relationship and calculate the predictive powers (level and gap) of the three models for the ten different buckets. We find that the performance of the three models – measured by the AUROC – improves with the length of the relationship (see Fig. 11, left-hand panel). On the right-hand side of Fig. 11, we compare the predictive power of the model based on the fintech score (Model I) with the model that considers only traditional information (Model II) and the model that includes all information (Model III). Interestingly, the comparative advantage of Model I over Models II and III tends to increase for low levels of the bank-customer relationship. However, when the relationship becomes stronger, the differences between Model I and the other two models decrease. This tallies with the idea that a longer relationship between bank and customer tends to attenuate asymmetric information problems. This is also reflected in the relationship between borrower and fintech company.

A final test was to insulate demand and supply side factors related to the regulatory shock. After the shock the Fintech firm faced deteriorating funding condition and had to reduce the supply of lending. Borrowers with a low fintech credit score have probably had harder time rolling over their debt. In other words, funding issues may have led the Fintech firm to ration credit by more to those borrowers that were identified as risky based on their credit scores. This would bias the performance tests in favour of the credit score model. To insulate this effect we have selected a subsample of loans for customers in the sample who wanted to roll over the credit but whose applications were refused by the fintech company. The results reported in the fifth panel of Table 4 indicate that the three models are statistically different at the 5 % level. In terms of contribution of non-traditional data and machine learning to predictive power, non-traditional data contribute an additional 2.4 % increase in the AUROC (=(0.5886–0.744)/0.5744), while applying machine learning techniques provides an additional 5.1 % increase in the AUROC (=(0.6186–0.5886)/0.5886).

7. Conclusion

The main goal of this paper is to compare the predictive power of credit scoring models based on machine learning techniques and big data with that of traditional loss and default models. Using a unique data set at the transaction level from a leading fintech company in China, we test the performance of different models to predict losses and defaults both in normal times and when the economy is hit by a shock. In particular, we analyse the case of an (exogenous) change in regulatory policy on shadow banking in China that caused lending to contract and credit conditions to deteriorate.

We find that the model based on machine learning and big data is better able to predict losses and defaults than traditional models in the event of a negative shock to the aggregate credit supply. This result reflects a higher capacity of non-traditional data to capture relevant borrower characteristics and of machine learning techniques to better mine the non-linear relationship between variables in a period of stress. By analysing different types of data, we find that non-traditional information, obtained from mobile phone applications and e-commerce platforms, has high predictive value. Finally, the comparative advantage of the model that uses the fintech credit scoring technique based on machine learning and big data tends to decline for those borrowers with a longer credit history.

Acknowledgements

We would like to thank the Editor Fariborz Moshirian, Sebastian Doerr, John V. Duca, Jon Frost, Xiang Li, Julapa Jagtiani, Weiyi Zhao and, in particular, two anonymous referees for comments and suggestions. We would also like to thank seminar participants at the 35th Australasian Finance and Banking Conference, the University of Basel, the Bank for International Settlements, Bocconi University and the Irving Fisher Committee – Central Bank of Malaysia conference for useful comments. We thank Giulio Cornelli for excellent research assistance. Jingyi Wang acknowledges the financial support by the National Natural Science Foundation of China (72003214). The views in this paper are those of the authors only and do not necessarily reflect those of the Bank for International Settlements. The authors wish to highlight that the data and analysis reported in this paper may contain errors and are not suited for the purpose of company valuation or to deduce conclusions about the business success and/or commercial strategy of the anonymous Chinese fintech firm. All statements made reflect the private opinions of the authors and do not express any official position of the anonymous fintech firm or its management. The analysis was undertaken in strict compliance with Chinese privacy law. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The anonymous fintech firm did not exercise any influence on the content of this paper but has ensured the confidentiality of the (raw) data.

Annex A. Differences in Credit Scoring Models between fintech companies and big techs

Fintech companies employ varying credit scoring models tailored to their specific needs and data sources. For example, the fintech company mentioned in this paper utilizes a combination of users' credit card-related information and non-traditional data, such as social networking activity, to assess creditworthiness. By analysing patterns in credit card usage and incorporating social network data, they gain a more comprehensive understanding of individuals' financial behaviours and social connections, which can be valuable for risk assessment.

While some fintech companies still require collateral as a traditional risk mitigation strategy, they complement it with non-traditional information to further enhance their risk assessment capabilities. This approach allows them to evaluate a borrower's creditworthiness not solely based on assets but also by considering their transactional history, social influence, and other alternative data points. By combining traditional and non-traditional information, these fintech companies aim to create more inclusive credit models that provide opportunities for individuals with limited credit history.

Big tech companies, operating online platforms, can leverage their vast user bases and access to diverse datasets encompassing various aspects of users' lives, including financial transactions, e-commerce activity, and social interactions. This wealth of data enables big tech companies to develop sophisticated credit scoring models based on a data-network-activity or "DNA" feedback loop, which takes into account multiple dimensions of an individual or small and medium enterprise's digital footprint. By analysing the extensive data available within their ecosystem, big tech companies can make more accurate and personalized lending decisions.

Overall, fintech risk assessment models primarily incorporate non-traditional information and employ novel approaches to credit evaluation. The availability of a rich dataset and the utilization of complex models can enhance the accuracy and effectiveness of risk assessment to a certain extent. However, it is important to note that these models heavily rely on historical data, and their reliability may be affected when sudden external environmental changes occur, as the models may not account for unprecedented events or shifts in economic conditions. Continuous monitoring, adaptation, and incorporating real-time data become crucial to mitigate potential risks associated with reliance on historical data alone.

Annex B. The development of fintech and big tech credit in China

The development of fintech and big tech credit in China can be traced back to the early 2000s when the country experienced a rapid economic growth and embraced technological advancements. China's large population, rising middle class, and limited traditional banking infrastructure created a fertile ground for the emergence of innovative financial technologies.

One of the earliest pioneers in Chinese fintech was Alipay, launched in 2004 by Alibaba Group. Initially designed as an online payment platform for Alibaba's e-commerce marketplaces, Alipay quickly expanded its services and introduced a digital wallet system. Alipay's success paved the way for other players to enter the market and fostered the growth of mobile payment solutions.

In 2013, Tencent, another major Chinese tech company, introduced WeChat Pay, integrated into its popular messaging app WeChat. WeChat Pay's seamless integration with daily life activities, such as sending messages, ordering food, and hailing rides, contributed to its widespread adoption and solidified China's position as a leading mobile payment market.

As fintech companies gained traction, they started collecting vast amounts of data on consumer behaviour, transaction history, and social connections. This data, combined with advancements in artificial intelligence and machine learning, allowed companies like Ant Group (an affiliate of Alibaba) to develop robust credit scoring systems. Ant Group's Sesame Credit, launched in 2015, utilized alternative data points to assess creditworthiness, enabling individuals with limited credit history to access financial services.

The Chinese government, recognizing the potential of fintech and big tech credit, embraced regulatory support and innovation. This supportive environment allowed companies like Ant Group and Tencent to expand their financial services beyond payments and credit scoring into wealth management, insurance, and lending.

Also peer-to-peer (P2P) platforms developed rapidly. The number of platforms reached 1931 in 2017.[*] Subsequently, the industry faced stricter regulations, and the number of platforms declined rapidly. In contrast to the P2P industry, the financial businesses of China's big tech companies continued to develop. The scale of big tech credit reached 516 billion USD while fintech credit reached 111 billion as of 2019 (Cornelli et al., 2022). In recent years, the pace of growth of big tech credit has been larger than that for bank credit. For instance, during 2020–21, big tech credit in China recorded an average annual growth rate of 37 %, compared to 13 % for bank credit (De Fiore et al., 2023).

[*] The number is based on the statistics from www.wdzj.com (网贷之家).

Annex C. Additional robustness tests

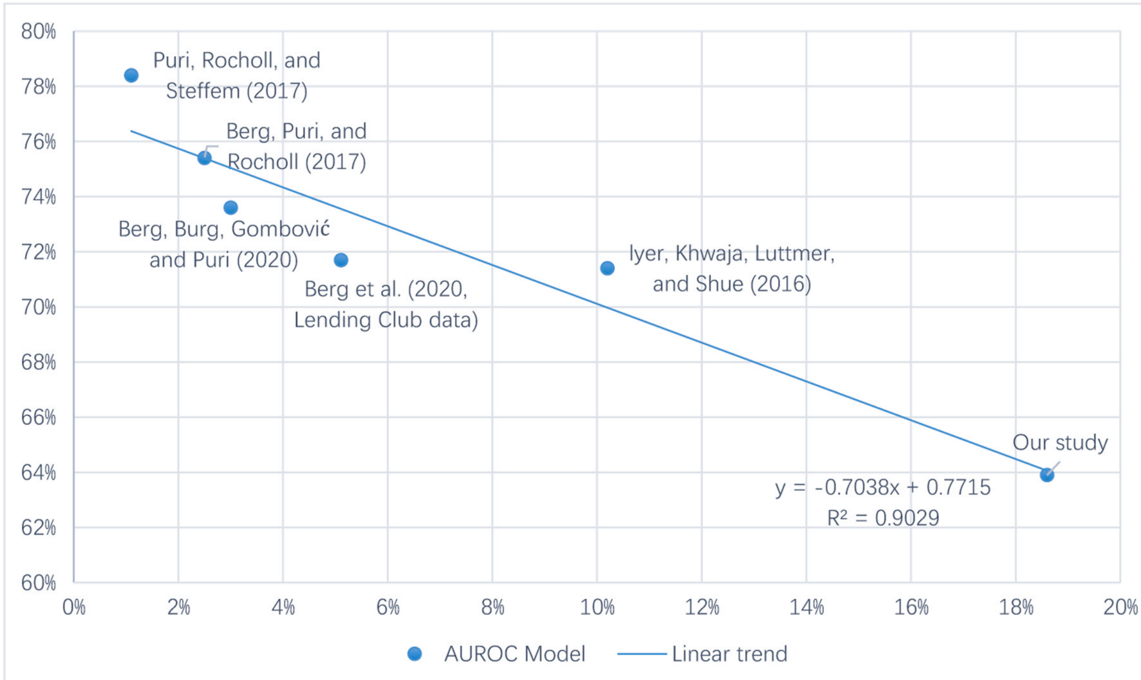

Fig. C1. Comparability of area under the ROC curve to other retail data sets, Note: The AUROC values reported on the vertical axis are taken from Table A2 in Berg, Burg, Gombović and Puri (2020). The results are not in the original papers but were provided by the authors using the same data set from the paper. The horizontal axis reports default rates. Source: Berg, Burg, Gombović and Puri (2020) and authors' calculations.

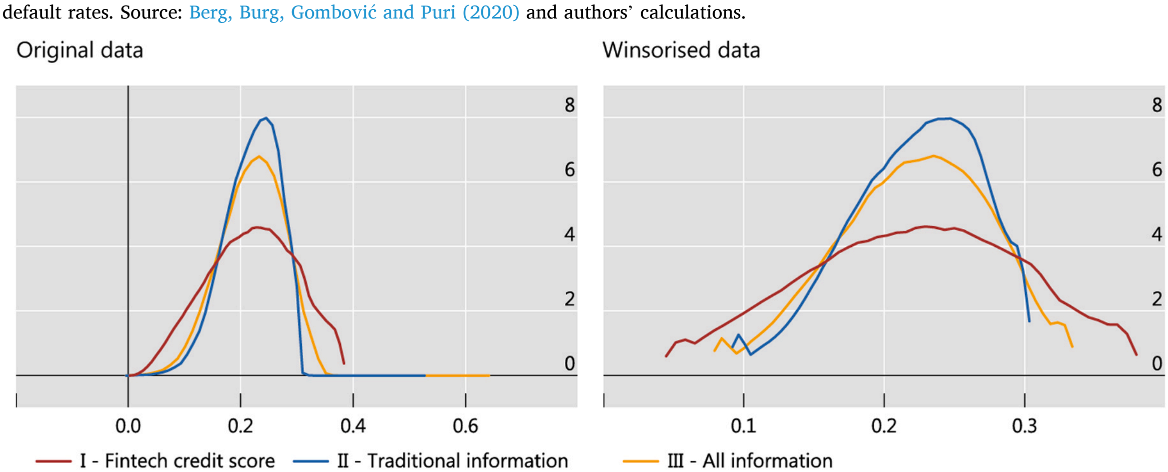

Fig. C2. Distribution of the expected default rate for the three different models, Note: The graphs show that the expected default rate for the whole sample is more dispersed for the fintech credit score model. This implies that Model I better captures the heterogeneity among borrowers. Source: Authors' calculations.

Table C1
A comparison between loans contracts with different contractual terms

I. Loan contracts with length up to one year				
	AUROC	Std err	95 % conf. interval	
I. Fintech credit score	0.6453	0.0013	0.64266	0.64789
II. Traditional information	0.604	0.0014	0.60134	0.60674
III. All information	0.6163	0.0014	0.61358	0.61895
II. Loan contracts with length of more than one year				
	AUROC	Std err	95 % conf. interval	
I. Fintech credit score	0.5997	0.0025	0.59482	0.60453
II. Traditional information	0.5779	0.0025	0.57308	0.58282
III. All information	0.585	0.0025	0.58015	0.58987

References

Banerjee, R.N., Gambacorta, L., Sette, E., 2021. "The real effects of relationship lending". J. Financ. Inter. 48 (October).

Belloni, A., V. Chernozhukov and C. Hansen (2011): "Inference for high-dimensional sparse econometric models," in Advances in Economics and Econometrics, 10th World Congress of the Econometric Society.

Berg, T., Burg, V., Gombović, A., Puri, M., 2020. "On the rise of fintechs – credit scoring using digital footprints". Rev. Financ. Stud. 33, 2845–2897.

Berg, T., Puri, M., Rocholl, J., 2020. "Loan officer incentives, internal rating models, and default rates". Rev. Financ. 24, 529–578.

Braggion, F., Manconi, A., Haikun, Z., 2022. "Household credit and regulatory arbitrage: Evidence from online marketplace lending". Manag. Sci. 69 (10), 6271–6292.

Buchak, G., Matvos, G., Piskorski, T., Seru, A., 2018. "Fintech, regulatory arbitrage, and the rise of shadow banks". J. Financ. Econ. 130 (3), 453–483.

Cornelli, G., Frost, J., Gambacorta, L., Rau, P.R., Wardrop, R., Ziegler, T., 2022. "Fintech and big tech credit: Drivers of the growth of digital lending". J. Bank. Financ. 148.

De Fiore, F., Manea, C., Gambacorta, L., 2023. "Big techs and the credit channel of monetary policy". BIS Work. Pap. 1088.

De Roure, C., Pelizzon, L., Tasca, P., 2016. "How does P2P lending fit into the consumer credit market?". Dtsch. Bundesbank Discuss. Pap. 30.

Degryse, H., Kim, M., Ongena, S., 2009. Microeconometrics of Banking. Oxford University Press.

Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., Kammler, J., 2016. "Description-text related soft information in peer-to-peer lending – Evidence from two leading European platforms". J. Bank. Financ. 64 (March), 169–187.

Frost, J., Gambacorta, L., Huang, Y., Shin, H.S., Zbinden, P., 2019. "BigTech and the changing structure of financial intermediation". Econ. Policy 34 (100), 761–799.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., Walther, A., 2022. "Predictably unequal? The Effects of Machine Learning on Credit Markets". J. Financ. 77 (1), 5–47.

Fuster, A., Plosser, M., Schnabel, P., Vickery, J., 2019. "The role of technology in mortgage lending". Rev. Financ. Stud. 32 (5), 1854–1899.

Gambacorta, L., Huang, Y., Li, Z., Qiu, H., Chen, S., 2023. "Data vs collateral". Rev. Financ. 27 (2), 369–398.

Giannone, D., Lenza, M., Primiceri, G., 2018. "Economic predictions with big data: The illusion of sparsity". Fed. Reserve Bank N. Y. Staff Rep. 847 (May).

Hertzberg, A., Liberman, A., Paravisini, D., 2018. "Screening on Loan Terms: Evidence from Maturity Choice in Consumer Credit". Rev. Financ. Stud. 31 (9), 3532–3567.

Iyer, R., Khwaya, A.J., Luttmer, E.F.P., Shue, K., 2016. "Screening Peers Softly: Inferring the Quality of Small Borrowers". Manag. Sci. 62 (6), 1554–1577.

Jagtiani, J., Lambie-Hanson, L., Lambie-Hanson, T., 2019. "Fintech lending and mortgage credit access". Fed. Reserve Bank Phila. Work. Pap. 19–47 (November).

Jagtiani, J., Lemieux, C., 2019. "The roles of alternative data and machine learning in fintech lending: evidence from the LendingClub consumer platform". *Financ. Manag.*, Winter 48 (4), 1009–1029.

Khandani, A., Kim, A., Lo, A., 2010. "Consumer credit-risk models via machine-learning algorithms". J. Bank. Financ. 4 (11), 2767–2787.

Puri, M., Rocholl, J., Steffen, S., 2017. "What do a million observations have to say about loan defaults? Opening the black box of relationship". J. Financ. Inter. 31, 1–15.

Tang, H., 2019. "Peer-to-peer lenders versus banks: substitutes or complements?". Rev. Financ. Stud. 32 (5), 1900–1938.

US Department of the Treasury (2016): "Opportunities and challenges in online marketplace lending", May.