Big Data Analysis with No Digital Footprints Available: Evidence from Cyber-Telecom Fraud

Laura Xiaolei Liu[1*]
Yufei Liu[2]
Xinghua Ruan[3]
Yu Zhang[1]

This draft: 12/21/2021.

Abstract

Cyber-telecom fraud is an increasingly severe problem globally. We focus on a special type of cyber-telecom financial fraud, in which criminals induce innocent people to borrow online. Since no digital footprints are available for the fraudsters behind the borrowing cases, identifying the fraud is difficult. Using a proprietary dataset of online consumer financing from a large Fintech company in China, we estimate the extent to which an intervention based on big data and machine learning can identify this type of fraud and prevent customers' financial losses. We find that female borrowers are more likely to become victims of fraud generally, that young and inexperienced users are more likely to become victim of fraud schemes targeting a lack of financial literacy, and that experienced and inexperienced users are equally likely to become victims of fraud schemes targeting overconfidence. Overall, the intervention effectively identifies fraud targeting either financial literacy or behavioral biases. However, it is more difficult to persuade victims of fraud targeting behavioral biases to change their behaviors.

Keywords: Fintech, big data, machine learning, cyber-telecom fraud, Internet finance

---

[1] Guanghua School of Management, Peking University.
[2] China Central Depository & Clearing Co., Ltd.
[3] Du Xiaoman Financial.
* Corresponding author: Laura Xiaolei Liu (laura.xiaolei.liu@gsm.pku.edu.cn). We thank Woo-Young Kang (discussant), David Robinson, Keer Yang (discussant), Bohui Zhang (discussant), the seminar participants at Fudan University Fanhai School of Finance and the conference participants at CFRC, CICF, China Meeting of the Econometric Society, EasternFA, SWFA and 2021 Law and Finance Forum for suggestions that improve this research. We thank Guanghua Thought Leadership for financial support. All errors are our own.

I. Introduction

With the advent of the Internet, cyber-telecom financial fraud has become a fast-growing field of white-collar crime globally, causing severe financial losses for the telecommunications industry and its customers. Cyber-telecom financial fraud, which can be understood as "the abuse of telecommunications products or services with the intention of illegally acquiring money from a communication service provider or its customers," costs the global economy approximately US$32.7 billion in losses each year (Europol, 2019).[1]

Cyber-telecom financial fraud has also become an increasingly important threat in China, the world's second largest financial market. The Ministry of Public Security registered approximately 590K cases in 2015; these reflected a year-on-year increase rate of 32.5% and caused RMB22.2 billion in financial losses (US$3.43 billion). In a survey of 30,000 randomly selected customers conducted by Tencent News, 90% of the respondents reported receiving cyber-telecom financial fraud messages in different forms.[2] Figure 1 shows the number of cyber-telecom-related criminal lawsuits in recent years.[3] From 2016 to 2018, criminal cases related to cyber crime increased by approximately 40% per year. In 2019, the number of cases doubled compared to 2018. This, however, remains an incomplete statistic; given that less than 5% of cyber-telecom financial fraud cases resulted in lawsuits, the real number of cases is likely to be 20 times that reflected in lawsuits.[4]

Fraudulently induced borrowing is one of the most serious forms of cyber-telecom financial fraud. Unlike bank deposit scams, fraudulently induced borrowing creates

[1] See https://www.europol.europa.eu/activities-services/public-awareness-and-prevention-guides/telecommunications-fraud for Europol's report.
[2] More details about the survey can be found at https://news.qq.com/cross/20170309/49rpD72V.
[3] The data are based on the China Justice Big Data Research Institute report on cyber crime (http://www.court.gov.cn/upload/file/2019/11/22/10/53/20191122105337_66635.pdf) and Legal Daily (https://www.chinanews.com/gn/2020/04-08/9150640.shtml).
[4] Difficulties in detecting cyber-telecom fraud can be found at https://m.66law.cn/laws/413213.aspx.

more severe problems, as victims commonly cannot pay back their debts. To make repayments, these individuals sometimes have to borrow from different platforms.[5] As cyber-telecom financial fraud causes severe financial losses to victims, research is needed to better understand this type of fraud and to identify prevention mechanisms. Our study fits this gap.

Traditional methods of intervening in financial scams or fraud are generally ineffective as "cooling off laws provide little protection, nudges cannot help, and it is difficult for preventative educational interventions to be timely enough to be salient" (Fernandes, Lynch, and Netemeyer, 2014, p. 1875). Using a large proprietary dataset on online consumer financing from a large Fintech company in China—Du Xiaoman Financial,[6] a subsidiary of big tech firm Baidu in China—we investigate two questions: (1) what types of individuals in the financial market are more likely to fall victim to cyber-telecom fraudulently induced borrowing; and (2) whether and how Fintech, specifically big data and machine learning, can help to prevent this type of cyber-telecom fraud.

The situation we focus on differs from identifying fraudulent borrowers on online platforms who have no intention to pay back. Recently, there is a burgeoning stream of studies on how big data usage can assist a firm's risk management in lending. For instance, Agarwal, Qian, Ren, Tsai, and Yeung (2020) and Berg, Burg, Gombović, Karolyi, and Puri (2020) show that digital footprints can be used to model borrowers' creditworthiness, and Dai, Han, Shi, and Zhang (2020) show that digital footprints can be used as collateral in debt collection.

In contrast, we show how big data analysis and machine learning can help identify cyber-telecom fraud when fraudsters leave no digital footprints. It is considerably more

[5] Aware of the danger of this particular type of fraud, the Ministry of Public Security launched the "Sword on Cloud 2020-Fighting cyber-telecom fraud on borrowing" campaign in May 2020 (See http://www.gov.cn/xinwen/2020-05/08/content_5509648.htm for more details about the campaign).
[6] Previously called "Baidu Finance."

difficult to identify fraud cases where fraudsters hide in the background, evading the collection of their digital footprints. Such an approach not only involves using big data to analyze borrowers and their digital footprints to determine their creditworthiness, ability, or willingness to pay their debts. It more importantly also involves assessing the motivations of borrowing behaviors and linking borrowers to someone else in the background who is conducting fraud. To make it more difficult, cyber-telecom fraud criminals are usually strangers to the victims, and there are little data available to link victims to criminals.

Our data come from Du Xiaoman Financial, a company primarily engaged in online lending. Since May 2019, the company has recorded occurrences of fraudulently induced borrowing and accumulated data on transactions proven ex post to be related to fraud. Commonly, the victims will report the cases to the police. However, in many cases, the criminals cannot be identified, and the loans are subject to legal dispute. [7] Understanding the legal perspective of these fraudulently induced borrowing is beyond the scope of this study; we focus on the occurrence of the fraud itself. The company has developed machine learning algorithms to help identify fraud-related applications. After the implementation of these algorithms, the company began to provide ex-ante warnings to loan applicants whose applications were identified as fraud-related (hereafter referred to as the "intervention").

In this study, our identification strategy is similar to a field experiment. The treated samples are loan applications (including ex-post performance) after the machine-learning-based intervention was applied. The control samples are similar loan applications when no such technology used to help prevent fraud. We first show that both samples are similar in all observable dimensions. Next, we show that the treatment sample records fewer occurrences of fraud ex post and substantially lower financial

[7] There are legal debates on whether and under what circumstances victims are still obligated to pay back these loans.

losses than the control sample. Finally, by comparing the machine learning algorithm adopted in the intervention with traditional logistic regressions, we show and quantify the relatively efficiency of the machine learning algorithm in identifying fraud over traditional methods. This evidence suggests that big data analysis and machine learning techniques are useful in helping to identify cyber-telecom fraud even when the perpetrators leave no data available in the analysis.

Studies aiming to improve the quality of household financial decisions have yet to reach a consensus on the relative efficacy of interventions targeting a lack of financial literacy compared with those targeting behavioral biases (Fernandes, Lynch, and Netemeyer, 2014). In this study, two types of fraudulently induced borrowing cases are distinguished: (1) cases where victims are unaware that they are borrowing (reflecting a lack of financial literacy), and (2) cases where victims know that they are borrowing but are overconfident about the high cash returns promised by fraudsters (reflecting behavioral biases). We find that inexperienced users are particularly likely to fall victim to fraud targeting financial literacy, whereas victims of fraud targeting overconfidence are more universal in representation. Female users are also more likely than male users to be victims of fraud, regardless of the type of fraud. Moreover, the intervention in this study, which takes the form of "just-in-time" education (informing victims that they are making a loan) at the same time as persuasion (encouraging victims to reconsider the risks of "too-good-to-be-true" returns), is found to be effective in preventing financial mistakes in both types of fraud. Albeit, when they fall prey to fraud targeting overconfidence, persuading victims to change their actions is more difficult as it takes more rounds of interventions and takes longer persuasion time.

Our paper makes several contributions to the literature. First, it contributes to emerging research on big data analysis. According to IBM (2013), "big data is a term applied to datasets whose size or type is beyond the ability of traditional relational databases to capture, manage and process the data with low latency. Big data has one

or more of the following characteristics: high volume, high velocity or high variety," and "big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes." In both accounting and finance literature, scholars have tried to extract information from financial reports by making textual analyses of the reports (e.g., Li, 2010; Loughran and McDonald, 2011; Li, Lundholm, and Minnis, 2013; Lang and Lawrence, 2015; Frankel, Jennings, and Lee, 2016; Hoberg and Phillips, 2016). Media content is another source of big data (e.g., Tetlock, 2007; Tetlock, Saar-Tsechansky, and Macskassy, 2008; Tetlock, 2010; Tetlock, 2015). With massive data from the Internet and mobile devices made available recently, studies increasingly use unstructured data from a variety of sources (e.g., Liao, Wang, Xiang, Yan, and Yang, 2020).

Second, our paper contributes to the growing literature applying new machine learning technology to traditional research questions. For example, Gu, Kelly, and Xiu (2020) use machine learning methods to measure asset risk premia, and Giglio, Liao, and Xiu (2019) propose a machine learning based procedure to perform multiple hypothesis testing to limit data snooping. Erel, Stern, Tan, and Weisbach (2018) show that machine learning algorithms can identify better performing corporate directors, and Li, Feng, Shen, and Yan (2020) use machine learning to analyze corporate culture. Easley, Lopez de Prado, O'Hara, and Zhang (2020) apply this method to the microstructure field.

Third, our paper contributes to the new line of studies on Fintech. This line of studies explores the economic impacts of the application of Fintech by both traditional financial institutions and the start-up Fintech firms. For example, Agarwal, Qian, Ren, Tsai, and Yeung (2020) and Berg, Burg, Gombović, Karolyi, and Puri (2020) show that digital footprints can be used to model borrowers' creditworthiness, and Dai, Han, Shi, and Zhang (2020) show that digital footprints are useful for debt collection. Liao,

Martin, Wang, Wang, and Yang (2020) show that informing borrowers that their loan performance will be reported to the public credit registry affects their loan take-up and repayment decisions. Our paper differs from these studies as we explore a unique situation in which the application of technology is required to identify fraud when information on the criminals is not available.

Fourth, our paper contributes to the literature on the roles of financial literacy and behavioral biases in shaping suboptimal financial actions. Numerous studies investigate the impact of financial literacy on different financial behaviors and outcomes, such as retirement planning and wealth accumulation (Ameriks, Caplin, and Leahy, 2003; Stango and Zinman, 2009; van Rooij, Lusardi, and Alessie, 2012), investment behaviors (van Rooij, Lusardi, and Alessie, 2011; Graham, Harvey, and Huang, 2009), and credit behaviors (Stango and Zinman, 2009; Brown, Grigsby, van der Klaauw, Wen, and Zafar, 2016). Understanding the effect of financial literacy and behavioral biases on financial decisions is important as it relates to the usefulness of investor education and debiasing. However, distinguishing financial literacy from behavioral biases is difficult. Our dataset and novel setting allow us to make this distinction, addressing a key knowledge gap in the literature.

Finally, our paper contributes to better understanding fraud. Research in this area mainly emphasizes company fraud, such as misreporting (e.g., Dechow, Ge Larson, and Sloan, 2011; Yu and Yu, 2011; Khanna, Kim, and Lu 2015; Karpoff, Koester, Lee, and Martin, 2017; Amiram, Bozanic, Cox, Dupont, Karpoff, and Sloan 2018). Gurun, Stoffman, and Yonker (2018) studies the Madoff Ponzi scheme and find that residents of communities that were more exposed to the scheme subsequently withdrew assets from investment advisers and increased deposits at banks. Studies on individual financial fraud ("scams" in the Western context) generally focus on older investors (Gamble, Boyle, Yu, and Bennett, 2013; DeLiema, Deevy, Lusardi, and Mitchell, 2020; Lee, Cummings, and Martin, 2019; Kumar, Muckley, Pham, and Ryan, 2018). In

comparison, the victims of cyber-telecom financial scams in this study are distinguishably younger. This finding is expected, given that young people constitute the largest group of Internet users. It is also consistent with the explanation of Modic and Lea (2014) who emphasize operational experience; that is, young people may lack financial experience compared to older individuals. Knüpfer, Rantala and Vokata (2021) find in Finland that victims of Ponzi schemes in Finland are also on average younger than the population. In contrast to Knüpfer, Rantala and Vokata (2021) and Rantala (2019) where victims of Ponzi schemes are more likely male, female users in our dataset are significantly more prone to both types of financial fraud in the online borrowing setting. Thus, our findings suggest that the vulnerability of females to financial fraud may need to be more emphasized.

The rest of this paper is organized as follows. Section II presents background information on the company's consumer lending business and the intervention. Section III describes the data and presents the empirical evidence. Section IV evaluates the efficacy of the intervention on different types of fraud in more detail. Section V presents the paper's conclusions.

II. Institutional Background and AI Technology

a) The Platform

With the development and expansion of China's financial market, emerging credit risks are subjecting traditional financial institutions to increasing operational pressures. As banks venture into new customer pools, the number of borrowers with high-quality hard information in the traditional sense tends to decline, while the cost of customer acquisition increases; this forces growth to slow down. However, Fintech firms have a greater ability to perform big data analysis using machine learning algorithms for risk control in credit lending. This advantage allows Fintech firms to thrive either by providing loan risk management services to traditional banks, by serving as alternative lenders leveraging advanced risk management capabilities in-house, or both.

We investigate the prevention of cyber-telecom financial fraud at the Fintech firm Du Xiaoman Financial. The company was formerly Baidu's financial services business group before it spun off in 2018. Like Alibaba's finance affiliate Ant Group, Du Xiaoman is a multi-business FinTech firm. Baidu obtained a third-party payment license and launched its wealth management platform in 2013, before acquiring a mutual fund sales license in 2014. Baidu's financial services business group was formally established at the end of 2015. In 2017, Baidu and CITIC Bank formed a joint venture, Baixin Bank, and obtained a license to provide online banking and lending services from the China Banking Regulatory Commission (CBRC). In April 2018, Baidu Finance officially split from Baidu and was renamed "Du Xiaoman" to facilitate independent operations.

Du Xiaoman as a Fintech company inherits Baidu's artificial intelligence ability. The suitability of Baidu to set up a Fintech affiliate comes first from the market presence of its flagship product, Baidu Search, and 14 other Internet and mobile applications including Baidu Map that together serve 95% of Chinese Internet users, a presence that enables Baidu to collect data under agreement of the user. Second, Baidu is also one of the first in China to invest in artificial intelligence, with a focus on applying machine learning techniques to its product operations. As such, Du Xiaoman is one of the largest providers to banks and Internet financial institutions (such as itself) of loan risk management solutions, which cover the three stages of loan origination, loan maintenance/servicing, and delinquency management/recovery.

The methods, techniques, and targets of cyber-telecom fraud are always changing motivated by huge profits. Online loans are becoming increasingly popular, because they provide credit to borrowers not served by traditional financial institutions. Accordingly, many fraudsters have shifted their focus to the vast number of users of online lending, who upon application approval receive large amounts of money from

the lending platforms. We focus on this particular type of cyber-telecom financial fraud, in which innocent individuals are tricked into borrowing through legitimate online lending platforms before transferring the funds to fraudsters' accounts.

As one of China's largest Fintech platforms, Du Xiaoman (hereafter referred to as the "Platform") has naturally become a main target for this type of fraud. The Platform has recorded cyber-telecom fraudulently induced borrowing cases since May 2019. In the first few months, approximately 20 cyber-telecom fraud victims were recorded daily, with an average loss of approximately RMB25,000. Some of these victims were tricked into borrowing from several platforms simultaneously, resulting in low payback ability. About half of all victims defaulted on their loans. As such, these cyber-telecom fraudulently induced borrowing cases caused significant financial losses to both users and the Platform.

However, current risk control rules and models cannot be applied to this specific scenario of cyber-telecom fraud. The difficulty comes from the fact that fraudsters do not show up in databases of loan applications, leaving no digital footprints for tracing. Tackling such fraud not only requires assessing borrowers' creditworthiness or ability/willingness to repay their debts via big data analysis of the borrowers' digital footprints as the Fintech literature has traditionally investigated (Agarwal, Qian, Ren, Tsai, and Yeung, 2020; Berg, Burg, Gombović, Karolyi, and Puri, 2020). The non-traditional requirement is that it also involves assessing the motivations of borrowing behaviors, whether the motivation is fraudulently induced, and linking borrowers to the third parties conducting fraud. This aspect of risk control and consumer financial protection is not studied previously. Moreover, because cyber-telecom fraudsters are usually strangers to their victims, in most cases there are no data to link both parties. To address these challenges, the Platform developed an anti-fraud system based on machine learning as an intervention measure, which we describe below.

b) The Gradient Boosting Decision Tree (GBDT) Algorithm

The intervention is based on a gradient boosting decision tree (GBDT) machine learning algorithm. GBDT (Friedman, 2001) is a member of the boosting family of integrated learning, and is an iterative algorithm. In each iterative step, GBDT fits the best "weak learner"—which is a simple nonlinear binary predictor (e.g. a decision tree) using a limited number of dependent variables—by using data from the prediction error of the existing predictive function (the "strong learner") from the last iterative step. GBDT then adds the weak learner to the existing strong learner from the last iterative step and subjects the former to a learning rate.

The weak learner's functional form in the GBDT algorithm is the decision tree model (Lewis, 2000) for binary prediction. This model uses a limited number of "layers", and each layer contains one or more single-variate binary prediction functions. The decision tree model starts from the first layer, where it locates a single variable and a threshold value to partition the dataset into two subgroups, before fitting a simple constant to each observation in the subgroups. The model then proceeds to the next layer, where each subgroup is further partitioned into two smaller subgroups. The structure of the decision tree model resembles its name; the dataset is broken into "tree branches," the subgroups in the highest layer corresponding to "tree leaves." The decision tree model output for each observation in the data a prediction score based on the sum of fitted constants over all the layers.

When fit with a small number of layers, the decision tree model is parsimonious method for accommodating interaction effects in the prediction of binary outcomes. However, decision trees are prone to overfitting when fit with a large number of layers. The GBDT algorithm (Friedman, 2001) overcomes the challenge of extending the decision tree model. By iteratively estimating and summing decision trees with a small number of layers, the GBDT algorithm reduces the likelihood of overfitting while including rich interactions of variables in the data to predict binary outcomes.

The pseudo-code for the GBDT algorithm is as follows:

- Take as input the training set sample $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$ where x_i are a vector of predicting variables and y_i is a binary outcome, the maximum number of iterations T, the loss function L, the learning rate λ, and the depth (number of layers) for weak learner K.

- The algorithm will yield an output $f(x)$, which is a strong learning of the binary outcome variable given the predictor variables, given the following iterative procedure:

1) Initialize the strong learner $f_0(x)$ in the 0-th iteration in the simplest possible way, where $f_0(x)$ is an estimated constant c that minimizes the sum of the loss function $L(y_i, c)$:

$$f_0(x) = \arg\min_c \sum_{i=1}^{m} L(y_i, c)$$

2) In each iteration $t = 1,2, \ldots, T$:

   a) For the sample $i = 1,2, \ldots, m$, calculate the negative gradient r_{ti}, which represents the prediction residual of the strong learner $f_{t-1}(x)$ in the last iteration:

$$r_{ti} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial (f(x_i))}\right]_{f(x)=f_{t-1}(x)}$$

   b) Use $(x_i, r_{ti})\ (i = 1, 2, \ldots, m)$ to fit the t-th decision tree, and its corresponding tree representation (leaf areas) is $R_{tj}, j = 1,2, \ldots, J$, where J is the number of leaf areas of regression tree t with depth K

   c) For each leaf area $j = 1, 2, \ldots, J$ calculate the best fitting constant value c_{tj}:

$$c_{tj} = \arg\min_c \sum_{x_i \in R_{tj}} L(y_i, f_{t-1}(x_i) + c)$$

   d) Update the strong learner $f_{t-1}(x_i)$ by adding the weak learner from the step above, shrunken by the learning rate $\lambda \in (0,1)$:

$$f_t(x) = f_{t-1}(x) + \lambda \sum_{j=1}^{J} c_{tj} I(x \in R_{tj})$$

3) After the T-th iteration, the algorithm obtains the predictive function $f(x)$ that assigns a value from zero to one given the predicting variables x:

$$f(x) = f_T(x) = f_0(x) + \sum_{t=1}^{T}\sum_{j=1}^{J} c_{tj} I(x \in R_{tj})$$

The Platform uses a training sample that includes demographic, behavioral and credit registry variables corresponding to a borrower, and an ex-post labeled binary outcome variable of the user being fraudulently induced to borrow. We describe these variables in detail in the next section. The GBDT algorithm specifies the following parameters (as determined from experiments with experimented values in parentheses): a learning rate of 0.1 (0.05, 0.1, 0.5), a maximum number of iterations of 100 (100, 200), a maximum depth of the decision tree in each iteration of 5 (3, 4, 5, 6), and a logistic loss function.

III. Data and Empirical Results

In this section, we describe the data and identification strategy used in the analysis. Our identification strategy is similar to that of a field experiment. The treatment sample consists of loan usage applications (including ex-post performance) after the deployment of the machine learning-based intervention, and the control sample includes comparable loan applications with no such technology used to help prevent cyber-telecom financial fraud.

[[INSERT Figure 2 around Here]]

Figure 2 illustrates the intervention received by the treatment group in the anti-fraud experiment. An applicant makes a loan usage application if they decide to take out a loan after observing the loan terms. In the control group, no actions are undertaken to prevent fraudulently induced borrowing from the Platform at the loan

usage stage.[8] In the treatment group, loan usage requests undergo anti-fraud screening. The anti-fraud system assigns a score for each loan application, with a higher score indicating a higher probability of cyber-telecom fraudulently induced borrowing. Loan usage requests are then approved without intervention for applicants with low scores. In contrast, applicants with high scores receive alerts and requests for feedback from the Platform via phone calls. Most applicants who are subject to cyber-telecom fraudulently induced borrowing recognize that a fraud has been committed and withdraw their loan usage requests. Applicants who are alerted but insist that they are not undertaking fraudulently induced borrowing also have their loan usage requests approved. The full sample contains all loan usage applications from the control and treatment groups.

The data is at the application level, and includes ex-post outcomes. We define a loan usage application as $Fraud = 1$ when it involves a cyber-telecom fraudulently induced borrowing case. This is identified based on (a) applicants' feedback to the Platform's warning phone calls, and (b) post-borrowing feedback from borrowers. Applicants and borrowers have little incentive not to report fraud to the Platform, because their responses remain private and not shared with their social network, and because recognition by the Platform is used in filing the fraud case with legal enforcement. We also define a loan usage application as $Use = 1$ when the borrower subsequently experiences financial losses due to cyber-telecom fraudulently induced borrowing. This means that a defrauded applicant successfully takes out the loan and transfers the money to a fraudster's accounts. This happens when the defrauded borrowing victim was not alerted, or alerted but the intervention was unsuccessful in convincing the victim. We then define a monetary variable $Loss$ that refers to the amount of loss caused by cyber-telecom fraud; it takes a value of 0 when $Use = 0$.

[8] In fact, the Platform reviews loan usage applications and intercepts abnormal loan usage applications as a routine procedure before and after the application of the anti-fraud system. However, this routine review procedure has nothing to do with cyber-telecom fraudulently induced borrowing. So we do not show this step in Figure 3 to make the flow diagram more straightforward.

The characteristics of the loan usage applicants include demographics (gender, age, income), behavioral (apply amount) and credit registry information (total amount of loans in the previous 12 months, number of loan accounts in the previous 12 months, number of days after the last credit inquiry, total credit card utilization, etc.) Machine learning in principle is able to utilize other dimensions of data, but in experimentations these characteristics provide almost all power to predicting whether an applicant is defrauded, and the marginal contribution of other data is negligible. The main purpose of our empirical analysis is to explore how these characteristics relate to the probability of *Fraud* and to evaluate the treatment effect of the intervention on the probability of *Use* and the amount of *Loss*.

a) Sample Balance of the Treatment and Control Groups

To examine the quality of randomization, we plot the applicant characteristics for the control and treatment groups in Figure 3. The distribution of gender, age, education, income, loan amount, and deal approval rates is almost identical for the control and treatment groups.

[[INSERT Figure 3 about Here]]

The same summary statistics are also used to characterize the loan applicants. These statistics show that most of the applicants are young men. The applicant base is dominated by individuals under 35. Male applicants account for 66% and 68% of the treatment and control groups, respectively. As the disclosure of education level is performed on a voluntary basis, most of the applicants do not report such information. In terms of income, most of the applicants earn RMB4,000–8,000 per month. More than half of the applicants apply for a loan of less than RMB5,000. Finally, less than 10% of the loan applications are rejected in the control and treatment groups.

To identify differences in the probability of cyber-telecom fraudulently induced borrowing across different age and gender groups, Table 1 shows the occurrences of fraud-induced loan usage applications (*Fraud*) and eventual credit use following a fraud-induced loan usage application (*Use*) in subsamples of the data across gender and age. Panel A shows this information separately for male and female applicants in the control and treatment groups. Panels B and C show fraud-induced applications and credit use across different age groups in the control and treatment groups, respectively.

[[INSERT Table 1 about Here]]

As shown in Panel A, the ratio of fraud-induced loan applications (*Fraud*) to all loan applications for female loan applicants is more than 12 (7) times that of male loan applicants in the control (treatment) group (0.446%/0.036%≈12 and 0.421%/0.057% ≈ 7, respectively), partly because female users account for more fraud induced applications, and partly because male users account for more normal loan applications. In addition, while the probability of fraud perpetration (*Use*) conditional on *Fraud* is quite similar between male and female applicants, the average *Loss* for each victim is 16% (225%) higher for female than for male applicants in the control (treatment) group conditional on *Use*.

As shown in Panels B and C, the probability that loan applicants are deceived by fraudsters and make loan applications (*Fraud*) decreases with their age. However, the average loss unconditionally and conditional on *Use* are both higher among mature applicants. The treatment group shows the similar pattern for *Fraud* and average loss unconditionally across age, but a more muted pattern for loss conditional on *Use*, possibly owing to the low number of fraud-induced credit use following the intervention. The three panels show a substantial decline in the probability of fraud-induced credit use (*Use*) in the treatment group. For instance, the likelihood of *Use* conditional on *Fraud* for applicants aged 26 to 30 is much lower for the treatment group (2.88%) than

for the control group (95%).

Table 2 presents the characteristics of fraud-induced loan usage requests; they are separated into subsamples based on (1) group type (i.e., control versus treatment), (2) whether the machine learning algorithm successfully detects the influence of fraudsters on applicants in the treatment group, and (3) whether loan applicants proceed with credit use after being alerted to fraud. Panel A shows the number of fraud-induced loan applications in the control group, which corresponds to a 95.6% probability of Use. Panels B and C show that the machine learning algorithm identifies many fraud-induced loan applications (346 identified applications against 39 non-identified applications); this corresponds to a detection rate of 89.8% (346/(346+39)).

Panel D shows the number of cases where individuals making fraud-induced applications are alerted by the anti-fraud system but fail to recognize that a fraud has been committed. These applicants take out their loans and transfer the funds to fraudsters (4%, 14/346). As shown in Panel D, fraud-induced loan applicants who ignore the Platform's alert are often young women with a relatively high value of *Loan Amount*.

[[INSERT Table 2 about Here]]

b) What Types of People are Likely to Fall Victim to Cyber-Telecom Financial Fraud?

In this section, we investigate the factors determining individual vulnerability to fraudulently induced borrowing. Unlike previous studies on financial fraud or scams in the Western context focusing on older populations (Gamble et al., 2013; DeLiema et al., 2018; Lee et al., 2018; Kumar et al. 2018), victims of cyber-telecom fraudulently induced borrowing appear to be distinctly younger.

We use logit and probit regressions at the loan level to make these results

interpretable (Table 3). The dependent variable, *Fraud*, is a dummy variable equal to 1 when a loan usage request is made under cyber-telecom fraudulently induced borrowing. The explanatory variables include *Age*, gender (*Female*), *Income*, *Loan Amount*, *Total Credit in the Past 12 Months* (excluding mortgages), *Number of Loan Accounts in the Past 12 Months*, *Historical Consumer Loan Amounts* (including settled and outstanding loans), *Days After the Last Credit Report Inquiry*, and *Credit Card Utilization Rate*.

[[INSERT Table 3 about Here]]

The results in Table 3 illustrate that young women are more likely to become potential victims of cyber-telecom fraudulently induced borrowing than young men, confirming the patterns identified in Tables 1 and 2. The results in all columns show that female applicants are much more likely to make loan usage requests induced by fraudsters than male applicants in the same age groups; this is evident from the coefficient of gender and the interaction terms between gender and age. The pattern that female applicants have a greater chance to be influenced by fraudsters may be because of gender differences in personality. For example, Cadsby, Maynes and Trivedi (2006) find that female experiment subjects are more compliant to monetary tax contributions than male subjects, and Buchan, Croson and Solnick (2008) find in monetary investment games that while male are more trusting when being the initializing party of the game, female are more trusting in reciprocal and communal situation. It may be that tactics of the fraudsters induced the reciprocal aspect of trusting behaviors of females more than males.

Young female and male applicants also have a higher probability of falling victim to fraud compared to older applicants. Applicants applying for larger loan amounts are also more likely to be manipulated by fraudsters, with the Loan Amount $>$ 10,000 group having the highest logit and probit coefficients. Applicants with little or no credit

experience are also at higher risk of being manipulated by fraudsters, as indicated by the coefficients of the external credit record variables. Our finding that younger applicants to be more gullible to financial frauds is partly because our online lending setting and sample involves mostly younger and middle age individuals. Agarwal, Driscoll, Gabaix and Laibson (2009) report that younger and older adults make more financial mistakes than middle-aged adults. Calvet, Campbell and Sodini (2009) find that younger people are financially less sophisticated, possibly because of lower wealth and financial experience.

Finally, applicants who attempt to borrow from multiple institutions simultaneously (triggering report inquiry records on the day of the loan usage application) are at higher risk of fraudulently-induced borrowing (as evidenced by the coefficient for Days After the Last Inquiry = 0). The prediction power of such abnormal application behavior on the probability of being defrauded is also as expected.

c) The Effect of the Intervention

We evaluate the effect of the Platform's machine learning based intervention on the probability of successful fraud-induced credit use (*Use*) and customer losses (*Loss*) by comparing the treatment and control groups (Table 4).

Ex-ante Similarity between Treatment and Control Groups

Table 4 shows that 385 and 315 loan usage requests are induced by fraudsters in the treatment and control groups, respectively. In the treatment and control groups, the probability of fraud-induced borrowing is 0.18% and 0.17%, and the average loss (conditional on ex-post successful credit use) is RMB23.3K and RMB29.0K, respectively. Using t-tests, we find that the two groups do not differ significantly in terms of their average probability of being deceived by fraudsters and their average loss conditional on credit use. This confirms that the treatment and control groups are largely similar, with the exception of the intervention.

[[INSERT Table 4 about Here]]

Reduction in Financial Fraud Perpetration

Despite their similarities, the number of loans with ex-post credit use under the influence of fraudsters is 301 in the control group but only 35 in the treatment group. The probability of eventual credit use (*Use*) conditional on being deceived by fraudsters (Fraud = 1) in the control group is 95.56% (= 301/315), much higher than the probability of 9.09% (= 35/385) for the treatment group. The average loss conditional on being deceived by fraudsters also shows a similar pattern (control group = RMB27.7K; treatment group = RMB2.1K). The machine learning algorithm uses a cut off of 0.3 on the predicted probability of *Fraud* and identifies 13,298 applications to be at high risk of fraud and sends alerts to the applicants. From this pool, 346 fraud cases are correctly identified, achieving a model accuracy of 2.60% (= 346/13,298) and a detection rate of 89.9% (= 346/385). Overall, these differences between the treatment and control groups correspond to economically and statistically significant improvements in outcomes for people targeted by financial scams.

Efficacy Compared to Logit and Linear Probability Models

The machine learning algorithm significantly improves the efficiency with which calls are used to intervene in potential cases of financial fraud. This high efficacy is clearly illustrated by a comparison of the effect of random warning calls. Given that the probability of fraudulently induced borrowing cases in the treatment group is 0.18%, an intervention that randomly chooses from 13,298 applicants (the number of positives in the machine learning prediction model) to make warning calls would detect only 24 (= 13,298*0.18%) fraud cases on average, with a detection rate of only 6.23%.

We also compare the predictive power of conventional logit and linear regressions for binary outcomes to that of the Platform's machine learning prediction. We compare the logit model (2) in Table 3, a conventional regression model for binary outcomes and

that has the highest pseudo R-squared value among the logit and probit models, with the Platform's machine learning algorithm based on GBDT. Note that as described in the beginning of this section, the logit model (2) includes all the main variables that contribute to the machine learning algorithm's prediction power.

This comparison is shown in Figure 5, which plots the receiver operating characteristic curve (ROC curve) for the logit model, a linear probability model with the same explanatory variables, and the machine learning prediction model. The ROC curve shows the performance of a binary prediction model at all detection thresholds by plotting sensitivity (true positive) against 1-sensitivity (false positive). One important consideration when evaluating binary prediction algorithms is the trade-off between false positives and false negatives. For example, while a high threshold for the logit model can be set to reduce false positives, this simultaneously increases false negatives. The closer the ROC curve is to the 45-degree diagonal, the more accurate the prediction model is, because the model can achieve both lower false negatives and lower false positives compared to another model with a ROC curve farther away from the 45-degree diagonal.

As shown in Figure 4, the ROC curve for the machine learning prediction model is closer to the upper left corner than the ROC curve for the logit model as well as the linear probability model, showing that the machine learning model is more accurate. The area under the curve (AUC) of the machine learning prediction model is 0.98, and is statistically significantly higher than the AUC of the logit model (0.95) and the linear probability model (0.94) with a p-value lower than 0.1%. Finally, if the Platform follows the logit model or the linear probability model and carry out the same number (13,298) of alert calls, the detection rate would only be 68.8% for the former and 65.7% for the latter. To achieve the same 89.9% detection rate of the machine learning model, the logit model and the linear probability model would require 3.17 times and 3.40 times more alert calls, respectively. Considering alert calls on average cost RMB6-9 per user

($1.0-1.5 USD) for the Platform, the machine learning prediction model produce a lower bound cost saving of RMB253K and RMB1.3 million ($54K and $213K USD) on calls over the cost of alert calls based on the logit prediction model and universal alert calls, respectively, for protecting borrowers from financial frauds.

[[INSERT Figure 4 about Here]]

Reduction of Borrower Financial Losses

Next, we estimate the economic benefits created by machine learning based system for loan applicants using the numbers in Table 4. The actual loss in the treatment group owing to cyber-telecom fraud is RMB818.5K. If no intervention via the anti-fraud system were undertaken, the probability of *Use* conditional on *Fraud* (~95%) and the average loss per capital (taking the average of the treatment and control groups; ~RMB25K) would be similar for the treatment and control groups. There would be 366 (= 385*95%) incidences of *Use* in the treatment group, and the loss for individual victims of cyber-telecom fraud would reach RMB9.15 million (= 366*25K).

Based on the back-of-the-envelope calculations above, the intervention saved more than RMB8 million for applicants during the 3-week experiment, reducing borrower financial losses by over 90%. This economic benefit is substantial, given the short duration of the experiment. In addition, the magnitude of the loss prevented at the individual level is comparable to the annual disposable income of the average person in China, suggesting a sizable micro-level impact. The economic effect of the intervention is also likely to be understated because a subset of fraud-induced loan applicants are likely to apply for loans on several different platforms.

We then show that the significant decrease in the probability of loss conditional on cyber-telecom fraud and the reduction in customer loss are not driven by fewer customers being targeted by fraudsters. This is evidenced by a regression showing that

the probability of being targeted by fraudsters in the treatment and control groups is similar. Table 5 compares the probability of fraud for the control and treatment groups using a logit model and a probit model, respectively, at the level of loan usage requests. The dependent variable is *Fraud*, as defined above. The groups are distinguished, with *Treatment* = 1 corresponding to the treatment group and *Treatment* = 0 for the control group. The coefficients of *Treatment* indicate the differences in Fraud probability between the treatment and control groups. In the four specifications, the coefficients of *Treatment* are not significant, suggesting that the likelihood of being manipulated by fraudsters ex ante (before the anti-fraud alert) is similar for the treatment and control groups.

As we mention earlier, given that communication with the Platform is entirely private, applicants and borrowers have little incentive not to report fraud. The similarity in the sample probability of *Fraud* is also consistent with this institutional detail, for there is no evidence that reported *Fraud* is any different between the treatment group and the control group.

[[INSERT Table 5 about Here]]

Last, we use a regression to demonstrate the effect of the intervention on the probability of fraud-induced credit use influenced by cyber-telecom fraud (*Use*) and the financial value of customer losses (*Loss*), and report the results in Table 6. We use a logit model and a probit model for fraud-induced credit use and ordinary least squares (OLS) for the financial value of customer losses at the loan usage level. The coefficients of *Treatment* indicate differences in the probability of fraud-induced credit use and customer losses between the treatment and control groups. In the four specifications, the coefficients of *Treatment* are significantly negative at $p < 0.01$, suggesting that the probability of fraud-induced credit use and customer losses are statistically lower for the treatment group than for the control group, confirming the results in Table 4.

The corresponding average marginal effect of *Treatment* for the logit model in column (1) is -0.17%; that is, the probability of fraud-induced credit use is reduced by 0.17% after the intervention. Given that the probability of fraud-induced credit applications is 0.18% (= 385/213,584) in the treatment sample, the intervention successfully identifies and intervenes in a large majority of fraud cases, resulting in an economically significant effect.

The OLS results presented in columns (3)–(4) indicate that all else being equal, the average loss per customer is RMB67.75 lower and the average loss conditional on being targeted by a fraudster is RMB20,534.22 lower for the treatment group. This suggests that the intervention involving the machine learning algorithm prevents considerable financial loss for individuals using the online borrowing platform.

[[INSERT Table 6 about Here]]

Back Test

We also present the results of a back test, which applies the machine learning algorithm on data for the control group and evaluates its performance in fraud identification (Figure 5). Assuming that the lower probability of *Use* and *Loss* in the treatment group is indeed due to the effect of the intervention, the algorithm would succeed in detecting fraud cases in other samples. Figure 5 shows that the distribution of predicted probability of *Fraud* for normal loan usage applications is concentrated at the low end of probabilities, and the distribution of predicted probability of *Fraud* for applications in fact under the influence of financial fraudster being more concentrated at the higher end. Using the same cut-off point of 0.3, the anti-fraud system correctly identifies 281 of the 315 actual fraud cases. The detection rate is very similar to that of the treatment group (Back Test = 89.2% vs. Treatment group = 89.9%), which suggests that the prediction accuracy of the machine learning algorithm displays stability across samples.

[[INSERT Figure 5 about Here]]

False Positives: Does Overprotection Deter Normal Credit Use?

A remaining question on the cost-benefit analysis of the intervention is that while the intervention benefits the borrowers by preventing monetary losses and potentially also benefits the Platform by reducing non-performing loans, the intervention can cost the Platform if alert calls to applicants impact the borrower experience in a way that disturbed applicants ("false positives") borrow less. Financial fraud is a low probability event -- predicting it inevitably give rise to false positives, especially given there is no digital footprints on the fraudsters. If this cost exists, then it may be cost-ineffective for financial institutions to carry out such intervention that prevents financial fraud.

We explore this issue regarding the potential effect that a false positive fraud alert call may have on ex-post credit use. The estimation sample is the treatment sample, but with applications actually fraudulently induced excluded from the sample in order to focus on false positives. The dependent variable *CreditUse* is a dummy variable defined to be 1 if the applicant eventually used the credit service from the Platform, and 0 if the applicant did not use the credit service. The main explanatory of interest is a dummy variable *FalsePositive*, which is defined to be 1 if the applicant received an alert call from the Platform despite being not induced by financial fraudsters. The control variables are the explanatory variables in the previous regressions. Table 7 report the estimation result, which show that holding all observable characteristics constant, receiving a *FalsePositive* call does not associate with lower probability of credit use with the Platform. If anything, *FalsePositive* has a borderline significant positive association with the chance of borrowing relationship. We therefore conclude that we find no evidence that the intervention against fraudulently induced borrowing lead to side costs to the financial institution in terms of impacted credit use on the false positives.

[[INSERT Table 7 about Here]]

IV. Fraud that Targets Financial Literacy or Behavioral Biases

In this section, we investigate the impact of financial literacy and behavioral biases on the likelihood of victimization and the effectiveness of the intervention. The potential victims of cyber-telecom fraudulently induced borrowing in our sample can be classified into two groups: those who do not know that they are applying for a loan and those who know that they are going through a loan application process.

The first group of potential victims fall into traps such as refund scams and false account cancellations. They follow the instructions of the fraudsters and have no idea that they are applying for a loan, even when they are filling out a loan application form. In general, this group of potential victims lacks financial literacy. In contrast, the second group of potential victims fall into traps such as promises of investment opportunities and unlicensed online gambling, often believing that they can make a fortune. This group of potential victims exhibits behavioral biases and more specifically, overconfidence.

The Platform keeps a record of all fraudulently induced borrowing cases, including tricks used by fraudsters, the number of warning calls made, and the length of each communication. This makes it possible to distinguish the two types of potential victims and to measure the difficulty of persuading a potential victim of fraud. In our sample, potential victims who lack financial literacy represent about four fifths of all fraud cases, and potential victims who fall prey to schemes targeting overconfidence represent the remaining one fifth of the fraud cases.

We investigate the characteristics associated with the two types of potential victims. Table 8 shows the results of the logit models and the differences in coefficients between columns (1) and (2). The outcome variable in column (1), *Fraud_FL*, is equal to 1 in fraud cases related to a lack of financial literacy. Similarly, the outcome variable in

column (2), *Fraud_ OC*, is equal to 1 in fraud cases related to overconfidence. In keeping with Table 3, the results of column (1) indicate that young and inexperienced women are more likely to be victims of fraud targeting their lack of financial literacy.

Fraud cases targeting overconfidence are different from those targeting a lack of financial literacy. Victims of fraud targeting overconfidence are more universally distributed across young and old, as the coefficients of Age 18–25 in column (2) are significantly smaller than those in column (1). Historical credit experience does not affect the likelihood of fraud targeting overconfidence, as the coefficients of *Total Loan in Previous 12 Months*, *Number of Loan Accounts in Previous 12 Months*, and *No Previous Credit Inquiry*[9] are not significant. In addition, compared with potential victims who lack financial literacy, overconfident individuals apply for relatively larger amounts; indeed, the coefficient of *Loan Amount* $>$ *10K* in column (2) is significantly higher than that in column (1).

In summary, young and inexperienced users are more likely to be victims of fraud targeting a lack of financial literacy, whereas fraud targeting behavioral biases attract young and older victims and experienced and inexperienced users. Female users are more likely to be victims of both types of fraud.

[[INSERT Table 8 about Here]]

A key contentious question in the literature on improving the quality of individual financial decisions is whether interventions targeting financial literacy (e.g., financial education) are more effective than those targeting behavioral biases (e.g., nudging). Our study distinguishes victimization due to a lack of financial literacy from that due to overconfidence. Specifically, the intervention takes the form of a call (often a robocall), during which users are first asked whether they are aware that they are

[9] As the central bank keeps loan inquiry records for 2 years, No Previous Credit Inquiry = 1 means that an applicant has no credit experience in the past 2 years.

applying for a loan (and educated if they are not). They are then asked whether they have been promised returns that are too good to be true (and persuaded not to move forward with their loans if this is indeed the case). Hence, the intervention combines "just-in-time" financial education with persuasion.

We analyze the behavioral responses of cyber-telecom financial fraud victims following the intervention. Specifically, we investigate whether these victims can be persuaded not to proceed with their loans (thereby preventing suboptimal action), and the level of difficulty in persuading those who are not receptive to warnings from the Platform (reflecting the resources required for effectively correcting the suboptimal action). We specifically investigate whether the intervention is effective for victims of fraud targeting a lack of financial literacy, for fraud targeting overconfidence, or both, and the time and resources required to successfully intervene in the cases.

We report the findings comparing the effectiveness of the intervention for behavioral biases and a lack of financial literacy in Figure 6. We also present the difficulty of deterring two types of potential victims from engaging in fraudulently induced borrowing, measured by the number of warning calls made and the total length of these warning calls. As shown in Panel A, a large proportion of potential victims lacking financial literacy can be persuaded with only one warning call, whereas a large proportion of overconfident potential victims require more than three warning calls to see through the deception. As shown in Panel B, more than half of the overconfident potential victims belong to the group that takes the longest time to be persuaded, almost double the proportion of potential victims lacking financial literacy and difficult to persuade. The results of Figure 6 indicate that potential victims lacking financial literacy are more willing to listen to the Platform's warnings than overconfident potential victims. It is more difficult to persuade potential victims with behavioral biases, as these cases involve more warning calls and longer communication times.

In summary, we find evidence that it takes much longer to persuade users who are victims of fraud due to overconfidence than users who are victims of fraud due to a lack of financial literacy. Nevertheless, both types of victims are eventually able to be persuaded.

V. Conclusion

Cyber-telecom fraud has become a serious problem globally. In this study, we focus on a particular type of cyber-telecom fraud involving criminals who trick innocent people into borrowing from online lending platforms. As the fraudsters responsible leave no digital footprints, identifying them is difficult. Using a proprietary dataset of online consumer financing from a large Internet company in China, we find that big data analysis and machine learning techniques can help identify this type of fraud and reduce customers' financial losses.

The effects observed are both economically and statistically significant. Our results suggest that the intervention based on machine learning prevents losses of estimated to be at the order of millions of RMB to customers and the Platform each year. That is, the intervention can protect thousands of applicants from cyber-telecom fraud each year, saving tens of thousands of RMB for each applicant. Our results also show that young women with little or no credit experience are more likely to be victims of cyber-telecom fraudulently induced borrowing. Potential victims lacking financial literacy are more willing to listen to the Platform's warnings, indicating that consumer education can play an important role in preventing cyber-telecom fraudulently induced borrowing. Collectively, the findings of this paper are useful for individuals, Fintech companies, and government departments aiming to prevent cyber-telecom financial fraud.

References

360 Group. 2020. Report on Trend of Cyber-Telecom Fraud. https://www.360.cn/n/11619.html

Agarwal, Sumit, Driscoll, John C., Gabaix, Xavier, and Laibson, David. 2009. The age of reason: Financial decisions over the life cycle and implications for regulation. Brookings Papers on Economic Activity, 2009(2): 51-117.

Agarwal, Sumit, Wenlan Qian, Yuan Ren, Hsin-Tien Tsai, and Bernard Yin Yeung. 2020. The Real Impact of FinTech: Evidence from Mobile Payment Technology. Available at SSRN: https://ssrn.com/abstract=3556340.

Ameriks, John, Andrew Caplin, and John Leahy. 2003. Wealth Accumulation and the Propensity to Plan. Quarterly Journal of Economics, 118(3): 1007-1047.

Amiram, Dan, Zahn Bozanic, James D. Cox, Quentin Dupont, Jonathan Karpoff, and Richard Sloan. 2018. Financial Reporting Fraud and Other Forms of Misconduct: A Multidisciplinary Review of the Literature. Review of Accounting Studies, 23(2): 732-783.

Berg, Tobias, Valentin Burg, Ana Gombović, Andrew Karolyi, and Manju Puri. 2020. On the Rise of FinTechs: Credit Scoring Using Digital Footprints. The Review of Financial Studies, 33: 2845-2897.

Brown, Meta, John Grigsby, Wilbert van der Klaauw, Jaya Wen, and Basit Zafar. 2016. Financial Education and the Debt Behavior of the Young. The Review of Financial Studies, 29(9): 2490-2522.

Cadsby, C. Bram, Elizabeth Maynes, and Viswanath Umashanker Trivedi. 2006. Tax compliance and obedience to authority at home and in the lab: A new experimental approach. Experimental economics 9(4): 343-359.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini. 2009. Measuring the financial sophistication of households. American Economic Review 99(2): 393-98.

China Justice Big Data Research Institute. 2019. China Justice Big Data Report—Cyber-Telecom Criminal Cases 2019. http://www.court.gov.cn/upload/file/2019/11/22/10/53/20191122105337_66635.pdf

Dai, Lili, Jianlei Han, Jing Shi, and Bohui Zhang. 2020. Digital Footprints as Collateral for Debt Collection. Working Paper.

Dechow, Patricia M., Weili Ge, Chad R. Larson, and Richard G. Sloan. 2011. Predicting

Material Accounting Misstatements. Contemporary Accounting Research, 28: 17-82.

DeLiema, Marguerite, Martha Deevy, Annamaria Lusardi, and Olivia S. Mitchell. 2020. Financial Fraud Among Older Americans: Evidence and Implications. The Journals of Gerontology: Series B, 75(4): 861-868.

Easley, David, Marcos Lopez de Prado, Maureen O'Hara, and Zhibai Zhang. 2021. Microstructure in the Machine Age. The Review of Financial Studies, 34: 3316-3363.

Erel, Isil, Lea H. Stern, Chenhao Tan, and Michael S. Weisbach. 2021. Selecting Directors Using Machine Learning. The Review of Financial Studies, 34: 3226–3264.

Europol. 2019. Cyber-Telecom Crime Report 2019. https://www.europol.europa.eu/activities-services/public-awareness-and-prevention-guides/telecommunications-fraud

Fernandes, Daniel, John G. Lynch, and Richard G. Netemeyer. 2014. Financial Literacy, Financial Education, and Downstream Financial Behaviors. Management Science, 60(8): 1861-1883.

Frankel, Richard, Jared Jennings, and Joshua Lee. 2016. Using Unstructured and Qualitative Disclosures to Explain Accruals. Journal of Accounting and Economics, 62: 209-227.

Gamble, Keith Jacks, Patricia Boyle, Lei Yu, and David Bennett. 2013. Aging, Financial Literacy, and Fraud. Netspar Discussion Paper No. 11/2013-066, Available at SSRN: https://ssrn.com/abstract=2361151 or http://dx.doi.org/10.2139/ssrn.2361151

Giglio, Stefano, Yuan Liao, and Dacheng Xiu. 2021. Thousands of Alpha Tests. The Review of Financial Studies, 34: 3456-3496.

Graham, John R., Campbell R. Harvey, and Hai Huang. 2009. Investor Competence, Trading Frequency, and Home Bias. Management Science, 55(7): 1094-1106.

Gu, Shihao, Kelly Bryan, and Dacheng Xiu. 2020. Empirical Asset Pricing via Machine Learning. The Review of Financial Studies, 33: 2223-2273.

Gurun, Umit G., Noah Stoffman, and Scott E. Yonker. 2018. Trust Busting: The Effect of Fraud on Investor Behavior. The Review of Financial Studies 31: 1341-1376.

Hoberg, Gerard, and Gordon Phillips. 2016. Text-based Network Industries and Endogenous Product Differentiation. Journal of Political Economy, 124: 1423-1465.

IBM. 2013. What is Big Data? Bringing Big Data to the Enterprise. ibm.com.

Karpoff, Jonathan M., Allison Koester, D. Scott Lee, and Gerald S. Martin. 2017. Proxies and Databases in Financial Misconduct Research. The Accounting Review, 92(6): 129-163.

Khanna, Vikramaditya, E. Kim, and Yao Lu. 2015. CEO Connectedness and Corporate Fraud. Journal of Finance, 70(3): 1203-1252.

Knüpfer, Samuli, Ville Rantala, and Petra Vokata. 2021. Scammed and Scarred: Effects of Investment Fraud on Its Victims. Fisher College of Business Working Paper.

Kumar, Gaurav, Cal B. Muckley, Linh Pham, and Darragh Ryan. 2018. Can Alert Models for Fraud Protect the Elderly Clients of a Financial Institution? Michael J. Brennan Irish Finance Working Paper Series Research Paper No. 18-16. Available at SSRN: https://ssrn.com/abstract=3230188 or http://dx.doi.org/10.2139/ssrn.3230188

Lang, Mark, and Lorien Stice-Lawrence. 2015. Textual Analysis and International Financial Reporting: Large Sample Evidence. Journal of Accounting and Economics, 60: 110-135.

Lee, Steven, Benjamin F. Cummings, and Jason Martin. 2019. Victim Characteristics of Investment Fraud. Academic Research Colloquium for Financial Planning and Related Disciplines. Available at SSRN: https://ssrn.com/abstract=3258084 or http://dx.doi.org/10.2139/ssrn.3258084.

Li, Feng. 2010. The Information Content of Forward-looking Statements in Corporate Filings-A Naive Bayesian Machine Learning Approach. Journal of Accounting Research, 48: 1049-1102.

Li, Feng, Russell Lundholm, and Michaael Minnis. 2013. A Measure of Competition Based on 10-K Filings. Journal of Accounting Research, 51: 399-436.

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan. 2021. Measuring Corporate Culture Using Machine Learning. The Review of Financial Studies, 34: 3265-3315.

Liao, Li, Xiumin Martin, Ni Wang, and Zhengwei Wang. 2020. The Carrot Effect of Informing Borrowers about Credit Reporting: Two Randomized Field Experiments. Working paper.

Liao, Li, Zhengwei Wang, Jia Xiang, Hongjun Yan, and Jun Yang. 2020. User Interface and First-hand Experience in Retail Investing. Review of Financial Studies, forthcoming.

Loughran, Tim, and Bill Mcdonald. 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. Journal of Finance, 66(1): 35-65.

Modic, David, and Lea E. G Stephen. 2013. Scam Compliance and the Psychology of Persuasion. Available at SSRN: https://ssrn.com/abstract=2364464 or http://dx.doi.org/10.2139/ssrn.2364464.

Rantala, Ville. 2019. How Do Investment Ideas Spread Through Social Interaction? Evidence From a Ponzi Scheme. The Journal of Finance, 74(5): 2349-2389.

Stango, Victor, and Jonathan Zinman. 2009. Exponential Growth Bias and Household Finance. The Journal of Finance, 64(6): 2807-2849.

Tetlock, Paul C. 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. Journal of Finance, 62: 1139-1168.

Tetlock, Paul C. 2010. Does Public Financial News Resolve Asymmetric Information? Review of Financial Studies, 23: 3520-3557.

Tetlock, Paul C. 2015. The Role of Media in Finance. In S. P. Anderson, D. Stromberg, and J. Waldfogel (Eds.), Handbook of Media Economics, Vol 1B, Chapter 18, pp. 701-721. Oxford: Elsevier.

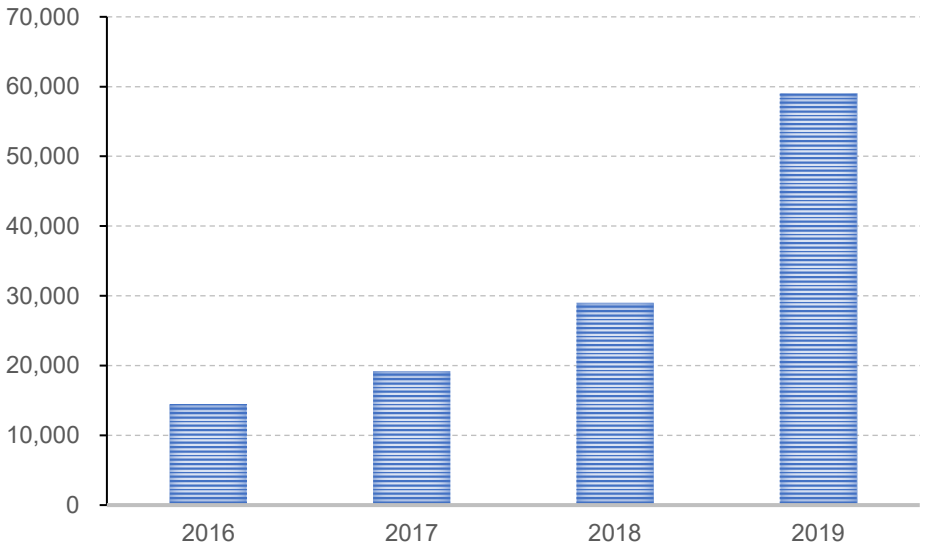
Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy. 2008. More Than Words: Quantifying Language to Measure Firms' Fundamentals. Journal of Finance, 63: 1437-1467.

Van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie. 2011. Financial Literacy and Stock Market Participation. Journal of Financial Economics, 101(2): 449-472.

Van Rooij, Maarten C. J., Annamaria Lusardi, and Rob J. M. Alessie. 2012. Financial Literacy, Retirement Planning and Household Wealth. The Economic Journal, 122(560): 449-478.

Yu, Frank, and Xiaoyun Yu. 2011. Corporate Lobbying and Fraud Detection. Journal of Financial and Quantitative Analysis, 46(6): 1865-1891.

Panel A: Number of cyber-telecom criminal cases



Panel B: Average monetary loss per case due to cyber-telecom fraud

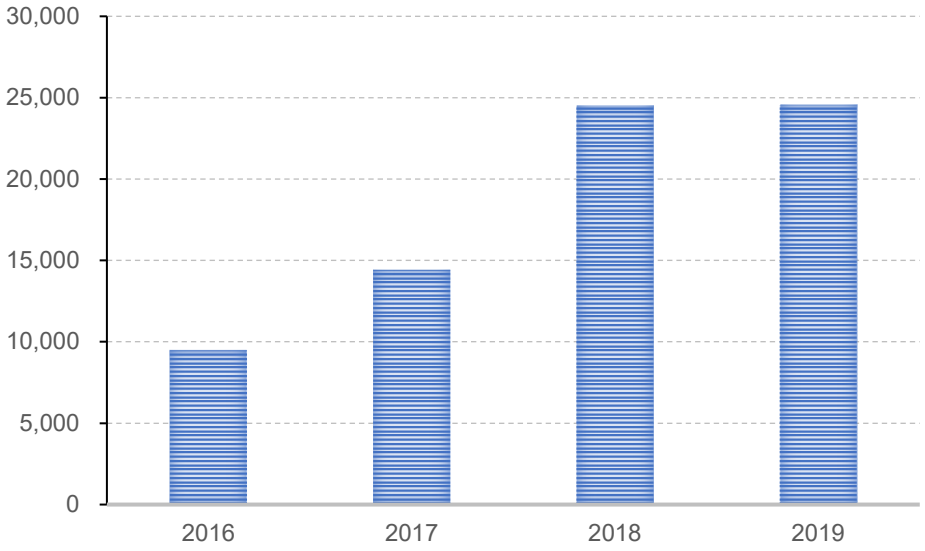


Figure 1: Numbers of Cyber-Telecom Criminal Cases and Loss per Case Each Year

Notes: Panel A shows the annual number of cyber-telecom criminal cases from 2016 to 2019. The data are from the China Justice Big Data Research Institute report and Legal Daily. Panel B shows the average per capita loss due to cyber-telecom fraud for the same period. The data come from the 360 Internet Safety Center.

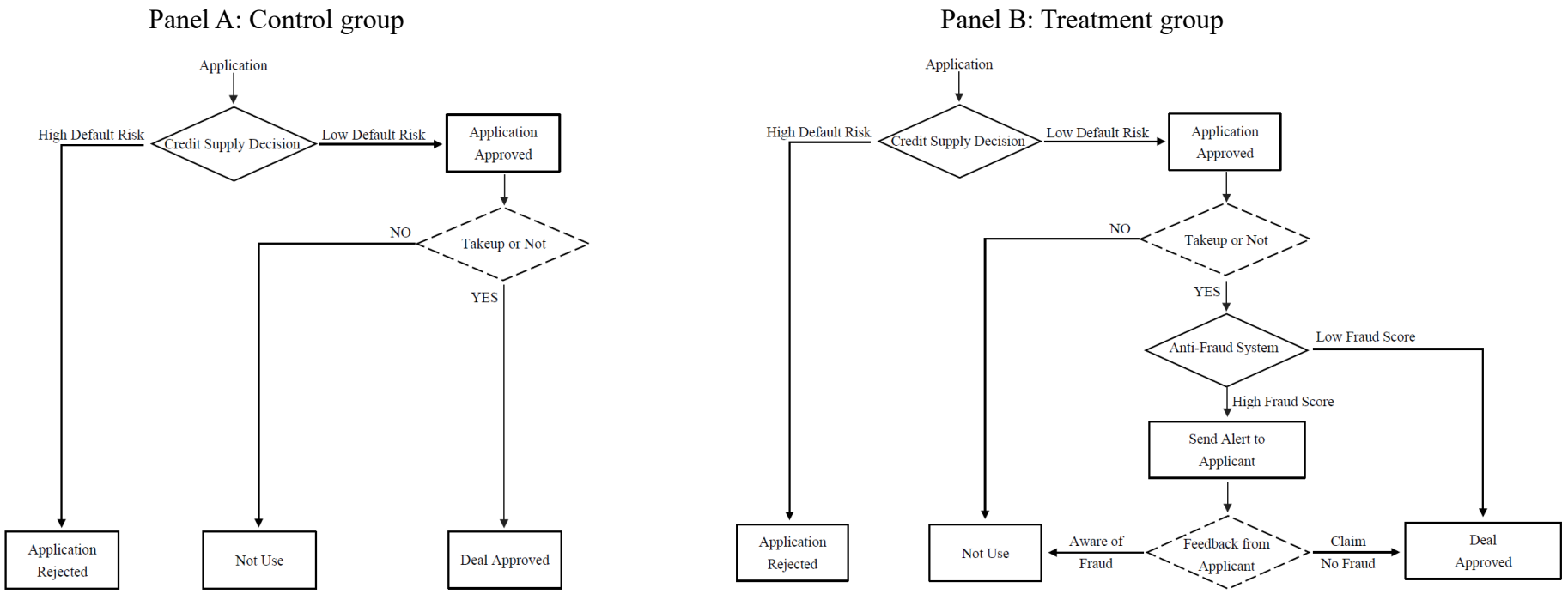
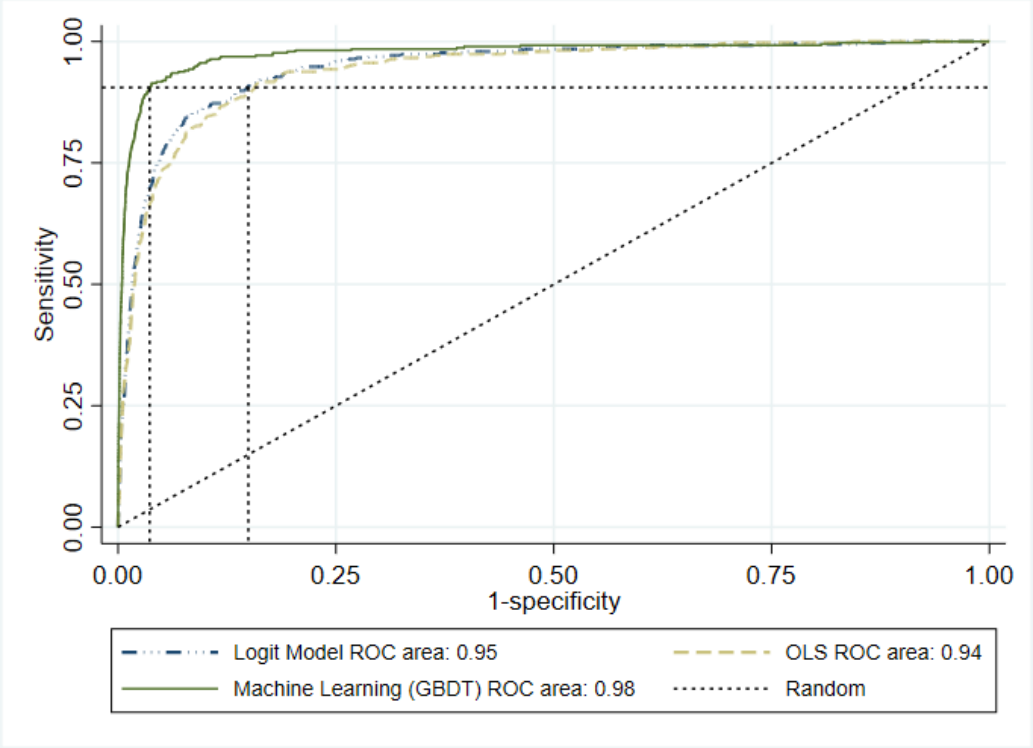


Figure 2: The Anti-Fraud Experiment

Notes: This figure illustrates the intervention received by the treatment group during the anti-fraud experiment. In the control group, there is no intervention from the Platform to prevent cyber-telecom fraudulently induced borrowing at the loan usage stage. In the treatment group, loan usage requests first undergo anti-fraud screening. The anti-fraud system assigns a score to each loan application, with a higher score indicating a higher probability of cyber-telecom fraudulently induced borrowing. For applicants with low scores, loan usage requests are approved without intervention. For applicants with high scores, the Platform sends alerts and obtains feedback from them. Most of the cyber-telecom fraudulently induced applicants then generally recognize the fraud and withdraw their loan usage requests. Applicants who insist that they are not undertaking fraudulently induced borrowing also have their loan usage requests approved. The sample contains all loan usage requests for the control and treatment groups.

Figure 3: Characteristics of the Treatment and Control Groups

Notes: This figure illustrates the distribution of gender, age, education, income, loan amount, and deal approval rates for the loan usage applications for the treatment and control groups. The x-axis corresponds to the percentage of applicants in each subcategory. The panels show the distribution of the applicants (a) by gender; (b) across four age groups; (c) across different education levels; (d) across five income groups; (e) across different loan amounts; and (f) by percentage of approved and rejected loan usage requests.



Ho: Area(Logit Model) = Area(Machine Learning Prediction Model)
$Chi^2(2) = 55.47$ $Prob > chi^2 = 0.0000$

Figure 4: ROC Curves: Machine Learning, Logit and Linear Probability Model

Notes: This figure illustrates the discriminatory power of different models based on the receiver operating characteristic curve (ROC curve) and the area under the curve (AUC). The ROC curves of the machine learning prediction based on the GBDT algorithm (green), of the logit model (blue) and of the linear probability model (LPM, yellow) are shown in the figure. The sample only includes the treatment group, and the specification of the logit model and the linear probability model used is the same as that in column 2 of Table 3. The result of the chi-square test for the difference in AUC between the logit model and the machine learning prediction model is also presented.

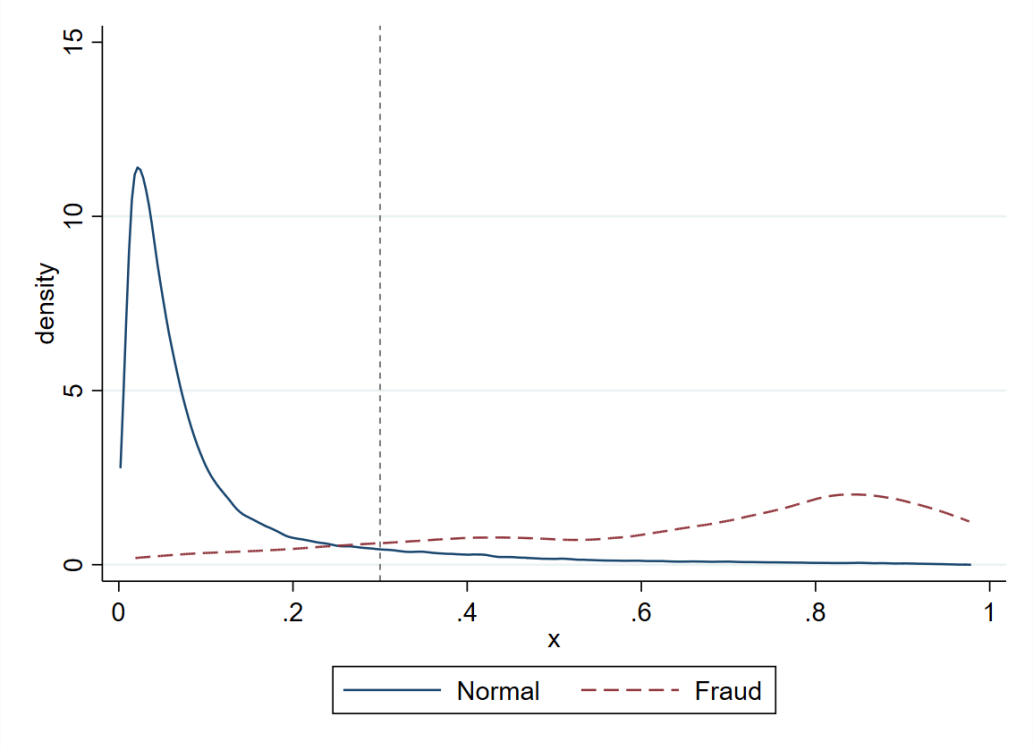
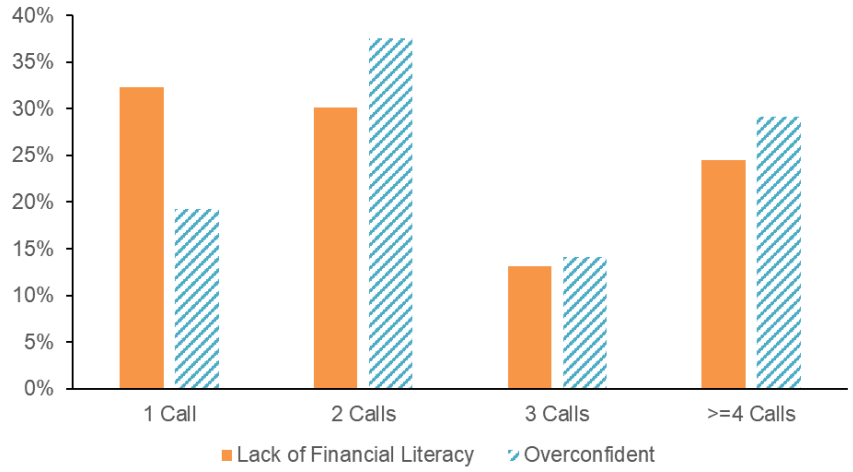


Figure 5: Back Test for the Anti-Fraud System in the Control Group

Notes: This figure presents the distribution of fraud scores in the control group, derived from the anti-fraud system in the back test procedure. The fraud score of normal applicants is represented by the blue line, and that of cyber-telecom fraudulently induced applicants is represented by the red line. The vertical dotted line is the cut-off point, with a fraud score > 0.3 corresponding to a high probability of cyber-telecom fraudulently induced borrowing.

Panel A: Number of warning calls made to potential victims



Panel B: Total length of warning calls to potential victims

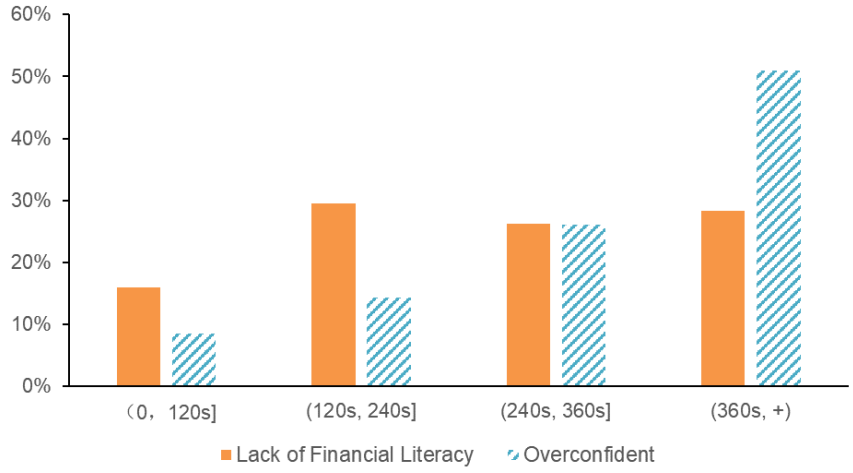


Figure 4: The Effect of Financial Literacy on the Difficulty of Dissuading a Potential Victim from Engaging in Fraudulently Induced Borrowing

Notes: This figure illustrates the difficulty in persuading two types of potential victims of fraudulently induced borrowing to change their behaviors, measured by the number of warning calls made and the total length of these warning calls. Potential victims are divided into two groups: those who are unaware that they are applying for a loan and those who are aware that they are making a loan application. The first subgroup lacks financial literacy and the second subgroup exhibits overconfidence (i.e., they mistakenly believe that they can earn very high returns through an investment or lottery).

Table 1: Fraudulently Induced Borrowing Across Gender and Age Groups

Panel A. Male versus Female

	Male, **Control**	Male, **Treatment**	Female, **Control**	Female, **Treatment**
No. of observations	126,847	141,112	60,332	72,467
No. of *Fraud* Cases	46	80	269	305
No. of *Use*	44	7	257	28
Ratio of *Fraud* to All Applications	0.0363%	0.0567%	0.4459%	0.4209%
Prob. of *Use* conditional on *Fraud*	95.65%	8.75%	95.54%	9.18%
Average *Loss* conditional on *Use*	25,545.45	8,357.14	29,628.40	27,142.86
Average *Loss* per potential victim	24,434.78	731.25	28,306.69	2,491.80

Panel B. Different Age Groups in the Control Sample

	Age < 26, **Control**	Age [26, 30], **Control**	Age [31, 35], **Control**	Age > 35, **Control**
No. of observations	50,469	54,905	39,812	41,993
No. of Fraud Cases	138	100	43	34
No. of *Use*	131	95	43	32
Ratio of *Fraud* to All Applications	0.2734%	0.1821%	0.1080%	0.0810%
Prob. of *Use* conditional on *Fraud*	94.93%	95.00%	100.00%	94.12%
Average *Loss* conditional on *Use*	5,577.10	27,766.32	35,713.95	37,950.00
Average *Loss* per potential victim	24,279.71	26,378.00	35,713.95	35,717.65

Panel C. Different Age Groups in the Treatment Sample

	Age < 26, **Treatment**	Age [26, 30], **Treatment**	Age [31, 35], **Treatment**	Age > 35, **Treatment**
No. of observations	607,06	57,838	42,900	52,135
No. of Fraud Cases	201	104	48	32
No. of *Use*	17	3	11	4
Ratio of *Fraud* to All Applications	0.3311%	0.1798%	0.1119%	0.0614%
Prob. of *Use* conditional on *Fraud*	8.46%	2.88%	22.92%	12.50%
Average *Loss* conditional on *Use*	16,029.41	45,666.67	12,636.36	67,500.00
Average *Loss* per potential victim	1,355.72	1,317.31	2,895.83	8,437.50

Notes: This table shows information on cyber-telecom fraudulently induced borrowing corresponding to different gender and age subsamples in the control and treatment groups. The Platform labels a loan usage application as a fraud case based on (a) applicants' feedback to alert phone calls, and (b) post-borrowing feedback from borrowers. *Fraud* takes a value of 1 if an applicant recognizes the fraud and withdraws their loan usage application after receiving the alert message, or if an applicant manipulated by fraudsters subsequently reports a fraud case to the Platform. *Use* (credit use following a fraud-induced loan application) takes a value of 1 if a cyber-telecom

fraudulently induced applicant successfully takes out a loan, and 0 otherwise. The sample probability of fraud-induced loan usage application (*Fraud*) is the number of fraud-induced loan usage requests divided by the sample size, and the sample probability of successful credit use following a fraud-induced loan application (*Use*) is the number of use incidences divided by the sample size. Average loss per case is calculated as average loss due to cyber-telecom fraud when the victim was successfully perpetrated by the fraudster, and average loss conditional on *Fraud* is total loss divided by the number of fraud cases, whether the fraudster was successful or not.

Table 2: Case Characteristics of the *Fraud* Subsamples

Panel A. Control: *Fraud=1 Subsample*

	N	mean	p50	SD	min	max
Use	315	95.56%	1	0.21	0	1
Female	315	0.85	1	0.35	0	1
Age	315	27.81	26	5.81	20	50
Loan Amount	315	29,070.79	20,000	25,426.10	1,000	100,000
Loss conditional on Use	301	29,031.56	20,000	25,207.56	1,000	100,000

Panel B. Treatment: *Fraud=1 Subsample, Missed by Machine Learning*

	N	mean	p50	SD	min	max
Use	39	53.85%	1	0.51	0	1
Female	39	0.67	1	0.48	0	1
Age	39	28.44	28	5.99	21	52
Loan Amount	39	13,069.23	6,700	16,244.48	500	75,000
Loss conditional on Use	21	14,666.67	10,000	17,736.38	500	75,000

Panel C. Treatment: *Fraud=1 Subsample, Correctly Identified by Machine Learning*

	N	mean	p50	SD	min	max
Use	346	4%	0	0.20	0	1
Female	346	0.81	1	0.40	0	1
Age	346	26.98	25	5.25	20	50
Loan Amount	346	13,404.62	10,000	16,059.02	500	100,000
Loss conditional on Use	14	36,464.29	17,250	34,743.82	3000	100,000

Panel D. Treatment: *Fraud=1 Subsample, Correctly Identified by Machine Learning, Use=1*

	N	mean	p50	SD	min	max
Use	14	100%	1	0	1	1
Female	14	0.93	1	0.27	0	1
Age	14	28.79	25.5	6.39	22	40
Loan Amount	14	33,607.14	17,250	33,392.29	3,000	100,000
Loss conditional on Use	14	36,464.29	17,250	34,743.82	3,000	100,000

Notes: This table presents the characteristics of the subsample of applicants under the influence of cyber-telecom financial fraud. Panel A shows the characteristics of the control group, and Panels B, C, and D show the characteristics of the treatment group. Panel B shows the characteristics of applicants whose applications under the influence of fraudsters were missed by the machine learning prediction system, and Panel C shows the characteristics of applicants whose applications under the influence of fraudsters were captured by the machine learning prediction system. Panel D shows the characteristics of applicants whose applications under the influence of fraudsters were captured by the machine learning prediction system, but when applicants fail to recognize the fraud and takes out the loan before transferring the money to the

fraudsters nevertheless. *Use* takes a value of 1 if a cyber-telecom fraudulently induced applicant successfully takes out a loan, and 0 otherwise. *Loss* conditional on *Use* is set to missing for applications when *Use* = 0.

Table 3: Who are More Likely to be Defrauded?

	Logit Model		Probit Model	
	(1)	(2)	(3)	(4)
	Fraud	*Fraud*	*Fraud*	*Fraud*
Female	1.6358***		0.5995***	
	(0.1019)		(0.0353)	
Age (> 35 ref.)				
18–25	1.5854***		0.5780***	
	(0.1515)		(0.0554)	
25–35	0.8335***		0.2934***	
	(0.1388)		(0.0494)	
Age and Gender Group (> 35 Male ref.)				
18–25 Female		3.4907***		1.1895***
		(0.3320)		(0.1007)
25–35 Female		2.9093***		0.9744***
		(0.3242)		(0.0962)
> 35 Female		2.0816***		0.6656***
		(0.3446)		(0.1046)
18–25 Male		2.2737***		0.7203***
		(0.3426)		(0.1040)
25–35 Female		0.7656**		0.2266**
		(0.3621)		(0.1075)
Income (0–4K ref.)				
4K–6K	0.4227***	0.3366**	0.1647***	0.1312**
	(0.1384)	(0.1402)	(0.0552)	(0.0560)
6K–8K	0.7088***	0.6405***	0.2610***	0.2352***
	(0.1578)	(0.1585)	(0.0625)	(0.0629)
8K–10K	0.7499***	0.6780***	0.2787***	0.2541***
	(0.2004)	(0.2011)	(0.0783)	(0.0787)
> 10K	0.4778	0.4377	0.1619	0.1390
	(0.3598)	(0.3602)	(0.1339)	(0.1352)
Loan Amount (0–5K ref.)				
5K–10K	0.5227***	0.5349***	0.2081***	0.2112***
	(0.1174)	(0.1176)	(0.0437)	(0.0438)
> 10K	1.2154***	1.2294***	0.4842***	0.4882***
	(0.0987)	(0.0991)	(0.0374)	(0.0376)
Total Loan in Previous 12 Months (> 20K ref.)				
0	1.8891***	1.8790***	0.6406***	0.6416***
	(0.2513)	(0.2511)	(0.0799)	(0.0802)
<= 20K	0.2505	0.2404	0.0851	0.0841
	(0.1756)	(0.1753)	(0.0583)	(0.0584)
Number of Loan Accounts in Previous 12 Months (> 1 ref.)				
0	1.1660***	1.1649***	0.3775***	0.3795***

	(0.2379)	(0.2377)	(0.0765)	(0.0767)
1	0.3823	0.3728	0.1531*	0.1540*
	(0.2625)	(0.2624)	(0.0823)	(0.0826)
Historical Consumer Loan (> 10K ref.)				
0	0.9271***	0.9271***	0.3562***	0.3573***
	(0.1392)	(0.1393)	(0.0493)	(0.0495)
<= 10K	0.4059***	0.4036***	0.1497***	0.1502***
	(0.1542)	(0.1542)	(0.0537)	(0.0537)
Days after the Last Credit Inquiry (> 0 ref.)				
0	2.2499***	2.2469***	0.8165***	0.8168***
	(0.0978)	(0.0977)	(0.0329)	(0.0330)
No Previous Credit Inquiry	1.1600***	1.1575***	0.3636***	0.3646***
	(0.1896)	(0.1897)	(0.0710)	(0.0712)
Credit Card Usage (> 20% ref.)				
<= 20%	1.3426***	1.3353***	0.4768***	0.4775***
	(0.1200)	(0.1201)	(0.0421)	(0.0423)
No Credit Card	1.1229***	1.1160***	0.3788***	0.3787***
	(0.1268)	(0.1269)	(0.0448)	(0.0449)
Constant	-12.9331***	-13.2197***	-5.2695***	-5.2990***
	(0.3316)	(0.4329)	(0.1187)	(0.1410)
pseudo R^2	0.306	0.308	0.306	0.308
N	400,758	400,758	400,758	400,758

Notes: This table illustrates the types of persons who are at risk of becoming victims of cyber-telecom fraudulently induced borrowing, based on analyses using a logit model and a probit model at the loan usage request level. The dependent variable is *Fraud*, which takes a value of 1 if an applicant recognizes the fraud and withdraws their loan usage application after receiving the alert message, or if an applicant manipulated by fraudsters subsequently reports a fraud case to the Platform. The explanatory variables include age, gender, income, loan amount, total credit in the last 12 months (excluding mortgages), number of loan accounts in the last 12 months, historical consumer loan amounts (settled and outstanding), days after the last credit report inquiry, and credit card utilization rate. The last five variables are obtained from external credit records. The baseline for age is 35 years; that for income is under RMB4,000; and that for loan amount is under RMB5,000. The baseline for total credit in the last 12 months is a credit exceeding RMB20,000; that for the number of loan accounts in the last 12 months is 1; that for historic consumer loan amounts is a loan exceeding RMB10,000; that for days after the last credit report inquiry is 1 day; and that for credit utilization rate is a rate of over 20%.

Table 4: Treatment Effect of the Intervention

		Treatment	Control		
(i)	Sample Size	213,584	187,179		
(ii)	No. of Fraud-induced applications (*Fraud*)	385	315		
(iii)	No. of *Identified*	13,298	0		
(iv)	No. of *Correctly Identified*	346	0		
(v)	No. of *Use*	35	301	Diff.	t-stat
(vi)	Average *Loss* conditional on *Use*	23,385.71	29,031.56	-5,645.85	1.2414
(vii)	Prob. of *Fraud* (ii)/(i)	0.18%	0.17%	0.01%	0.9053
(viii)	Model Accuracy（iv)/(iii)	2.60%	-	-	-
(ix)	Detection Rate (iv)/(ii)	89.87%	0	89.87%	52.7885
(x)	Prob. of *Use* Conditional on *Fraud* (v)/(ii)	9.09%	95.56%	-86.46%	44.7253
(xi)	Average *Loss* per potential victim (vi)*(v)/(ii)	2,125.97	27,741.27	-25,615.29	17.9860

Notes: This table summarizes the treatment effect of the machine learning based anti-fraud system on the probability of credit use under the influence of financial fraud and the monetary value of borrower losses, showing the differences between the two groups and the corresponding t-statistics. The Platform labels an application as fraud-induced based on (a) applicants' feedback to alert phone calls, and (b) post-borrowing feedback from borrowers. *Fraud* takes a value of 1 if an applicant recognizes the fraud and withdraws their loan usage application after receiving the alert message, or if an applicant manipulated by fraudsters subsequently reports a fraud case to the Platform. *Identified* takes a value of 1 if an application receives a fraud score higher than the cut-off point from the anti-fraud system. *Correctly Identified* takes a value of 1 if *Fraud* = 1 and *Identified* = 1. *Use* takes a value of 1 if a cyber-telecom fraudulently induced applicant successfully takes out a loan, and 0 otherwise. The probability of *Fraud* is the number of fraud-induced loan usage requests divided by the sample size. Model accuracy is defined as the number of fraud-induced loan usage requests correctly identified, divided by the total number of applications identified. The detection rate is the percentage of fraud-induced loan usage requests identified by the anti-fraud system, that is, the number of correctly identified loan usage requests divided by the total number of fraud-induced loan usage requests. Probability of *Use* conditional on *Fraud* is the number of use loan usage requests divided by the number of fraud-induced loan usage requests. Average *Loss* per potential victim is calculated as total loss/number of fraud cases.

Table 5: Probability of Getting Defrauded: Treatment versus Control Groups

		Logit Model		Probit Model	
		(1)	(2)	(3)	(4)
		Fraud	*Fraud*	*Fraud*	*Fraud*
Treatment		0.0412	0.1352	0.0135	0.0362
		(0.0778)	(0.0834)	(0.0256)	(0.0326)
Female		2.2383***	1.6371***	0.7001***	0.5996***
		(0.0992)	(0.1020)	(0.0295)	(0.0354)
Age (> 35 ref.)					
	18–25	1.5066***	1.5760***	0.4939***	0.5752***
		(0.1493)	(0.1511)	(0.0469)	(0.0554)
	25–35	0.7238***	0.8370***	0.2236***	0.2938***
		(0.1374)	(0.1388)	(0.0422)	(0.0494)
Income (0–4K ref.)					
	4K–6K	0.2907**	0.4306***	0.0927**	0.1660***
		(0.1359)	(0.1386)	(0.0465)	(0.0552)
	6K–8K	0.3710**	0.6881***	0.1258**	0.2554***
		(0.1498)	(0.1583)	(0.0510)	(0.0627)
	8K–10K	0.2551	0.7028***	0.0890	0.2663***
		(0.1901)	(0.2025)	(0.0630)	(0.0791)
	> 10K	-0.0664	0.4376	-0.0115	0.1506
		(0.3497)	(0.3607)	(0.1092)	(0.1343)
Apply Amount (0–5K ref.)					
	5K–10K		0.5258***		0.2093***
			(0.1174)		(0.0438)
	> 10K		1.2287***		0.4883***
			(0.0991)		(0.0377)
Total Loan in Previous 12 Months (> 20K ref.)					
	0		1.8237***		0.6238***
			(0.2543)		(0.0813)
	<=20k		0.2467		0.0841
			(0.1755)		(0.0583)
Number of Loan Accounts in Previous 12 Months (> 1 ref.)					
	0		1.0793***		0.3551***
			(0.2442)		(0.0792)
	1		0.2897		0.1292
			(0.2688)		(0.0852)
Historical Consumer Loan (> 10K ref.)					
	0		0.9311***		0.3571***
			(0.1390)		(0.0493)
	<= 10K		0.4021***		0.1491***
			(0.1542)		(0.0537)
Days after the Last Credit Inquiry (> 0 ref.)					

0		2.2538***		0.8172***
		(0.0978)		(0.0329)
No Previous Credit Inquiry		1.1656***		0.3659***
		(0.1897)		(0.0710)
Credit Card Usage (>20% ref.)				
<=20%		1.3454***		0.4771***
		(0.1200)		(0.0421)
No Credit Card		1.1180***		0.3767***
		(0.1268)		(0.0448)
Constant	-8.8924***	-12.9370***	-3.7132***	-5.2699***
	(0.2170)	(0.3317)	(0.0697)	(0.1188)
pseudo R^2	0.086	0.306	0.086	0.307
N	400,758	400,758	400,758	400,758

Notes: This table compares the probability of fraud-induced loan applications for the control and treatment groups using a logit model and a probit model at the loan usage request level. The dependent variable is *Fraud*, which takes a value of 1 if an applicant recognizes the fraud and withdraws their loan usage application after receiving the alert message, or if an applicant manipulated by fraudsters subsequently reports a fraud case to the Platform. *Treatment* equals 1 for the treatment group and 0 for the control group. The control variables include age, gender, income, loan amount, total credit in the last 12 months (excluding mortgages), number of loan accounts in the last 12 months, historical consumer loan amounts (settled and outstanding), days after the last credit report inquiry, and credit card utilization rate. Loan amount ranges from RMB500 to RMB100,000. The last five variables are obtained from external credit records. The baseline for age is 35 years; that for income is under RMB4,000; and that for loan amount is under RMB5,000. The baseline for total credit in the last 12 months is a credit exceeding RMB20,000; that for number of loan accounts in the last 12 months is 1; that for historic consumer loan amounts is a loan exceeding RMB10,000; that for days after the last credit report inquiry is 1 day; and that for credit utilization rate is a rate of over 20%.

Table 6: Probability of *Use* and Amount of Monetary *Loss*: Treatment versus Control Groups

	Logit Model	Probit Model	OLS	
	(1)	(2)	(3)	(4)
	Fraud and *Use*	*Fraud* and *Use*	*Loss*	*Loss* in the *Fraud* Subsample
Treatment	-2.2519***	-0.7392***	-67.7454***	-20,534.2195***
	(0.1851)	(0.0572)	(4.5530)	(1,503.4666)
Female	1.7000***	0.5889***	46.4968***	2,349.4556
	(0.1569)	(0.0508)	(3.7936)	(1,769.7576)
Age (> 35 ref.)				
18–25	1.1173***	0.4139***	28.0312***	-1,739.7427
	(0.2265)	(0.0801)	(5.9396)	(2,645.8520)
25–35	0.6394***	0.2317***	12.3312***	-3,511.8943
	(0.1925)	(0.0668)	(4.4744)	(2,412.5695)
Income (0–4K ref.)				
4K–6K	-0.0199	-0.0086	7.9893	2,221.5675
	(0.1845)	(0.0722)	(8.9074)	(2,322.7280)
6K–8K	-0.1604	-0.0596	0.2506	3,965.6339
	(0.2261)	(0.0859)	(9.6259)	(2,602.2605)
8K–10K	-0.0132	0.0197	2.0386	7,054.0262**
	(0.3048)	(0.1112)	(10.7915)	(3,321.5702)
> 10K	-0.1452	-0.0689	6.8006	29,237.1216***
	(0.5060)	(0.1813)	(13.4785)	(6,034.7430)
Loan Amount (0–5K ref.)				
5K–10K	0.6580***	0.2440***	2.5177	-1,340.8638
	(0.2130)	(0.0706)	(4.5214)	(2,006.8266)
> 10K	1.9897***	0.7054***	78.6832***	12,906.9618***
	(0.1632)	(0.0560)	(4.4387)	(1,740.7156)
Total Loan in Previous 12 Months (> 20K ref.)				
0	1.6636***	0.5736***	71.8578***	-7,316.1111*
	(0.2930)	(0.0914)	(7.4399)	(3912.8004)
<= 20K	-0.1254	-0.0210	2.7296	-2,748.1109
	(0.2562)	(0.0811)	(4.7243)	(2,990.0214)
Number of Loan Accounts in Previous 12 Months (> 1 ref.)				
0	1.3678***	0.4621***	38.4626***	-5641.1233
	(0.2807)	(0.0885)	(6.4259)	(4291.9842)
1	0.8418**	0.3183***	51.1999***	-5243.9883
	(0.3274)	(0.0994)	(5.6857)	(4,846.7489)
Historical Consumer Loan (> 10K ref.)				
0	1.1246***	0.3883***	60.2081***	365.1158

	(0.2097)	(0.0690)	(6.6423)	(2,387.5527)
<= 10K	0.8912***	0.2958***	12.3341**	190.7980
	(0.2291)	(0.0750)	(5.4280)	(2,618.6180)
Days after the Last Credit Inquiry (> 0 ref.)				
0	1.6925***	0.6020***	77.7822***	-2,568.7877
	(0.1317)	(0.0446)	(5.2460)	(1,699.0023)
No Previous Credit Inquiry	1.1118***	0.3554***	78.3702***	-2.9870
	(0.2353)	(0.0891)	(14.2573)	(3,226.8287)
Credit Card Usage (> 20% ref.)				
<= 20%	1.1401***	0.3952***	44.4846***	815.5284
	(0.1599)	(0.0550)	(4.8520)	(2,080.8229)
No Credit Card	0.7529***	0.2500***	-2.1100	-958.2578
	(0.1768)	(0.0611)	(4.6032)	(2,219.1958)
Constant	-12.1552***	-4.9121***	-64.7304***	24,042.6425***
	(0.4233)	(0.1481)	(11.2289)	(5,723.2069)
R^2			0.004	0.450
pseudo R^2	0.336	0.331		
N	400,758	400,758	400,758	700

Notes: This table compares the probability of actual credit use following fraud-induced credit applications and customer losses for the control and treatment groups. We use a logit model and a probit model for credit use incidences and OLS for customer losses at the loan usage level. The outcome variables are *Use* and *Loss*. *Use* is a dummy variable with a value of 1 if a cyber-telecom fraudulently induced applicant successfully takes out a loan, and 0 otherwise. *Loss* is the amount of monetary loss caused by cyber-telecom fraud, which takes a value of 0 for applicants when *Use* = 0. *Fraud* takes a value of 1 if an applicant recognizes the fraud and withdraws their loan usage application after receiving the alert message, or if an applicant manipulated by fraudsters subsequently reports a fraud case to the Platform. *Treatment* equals 1 for the treatment group and 0 for the control group. The control variables include age, gender, income, loan amount, total credit in the last 12 months (excluding mortgages), number of loan accounts in the last 12 months, historical consumer loan amounts (settled and outstanding), days after the last credit report inquiry, and credit card utilization rate. Loan amount ranges from RMB500 to RMB100,000. The last five variables are obtained from external credit records. The baseline for age is 35 years; that for income is under RMB4,000; and that for loan amount is under RMB5,000. The baseline for total credit in the last 12 months is a credit exceeding RMB20,000; that for number of loan accounts in the last 12 months was one; that for historic consumer loan amounts is a loan exceeding RMB10,000; that for days after the last credit report inquiry is 1 day; and that for credit utilization rate is a rate of over 20%.

Table 7: The (Lack of) Cost of Overprotection? False Positives and Credit Use

		(1)
Sample: Treatment Group (excluding *Fraud*=1)		*CreditUse*
False Positive		0.0405*
		(0.0218)
Female		-0.0396***
		(0.0102)
Age (> 35 ref.)		
	18–25	0.0179
		(0.0150)
	25–35	-0.0280**
		(0.0116)
Income (0–4K ref.)		
	4K–6K	-0.0568**
		(0.0249)
	6K–8K	-0.1091***
		(0.0269)
	8K–10K	-0.0762***
		(0.0295)
	> 10K	-0.0327
		(0.0356)
Loan Amount (0–5K ref.)		
	5K–10K	0.0649***
		(0.0121)
	> 10K	0.1563***
		(0.0122)
Days after the Last Credit Inquiry (> 0 ref.)		
	0	0.0275**
		(0.0139)
	No Previous Credit Inquiry	0.0439
		(0.0403)
Total Loan in Previous 12 Months (> 20K ref.)		
	0	0.0444
		(0.0582)
	<= 20K	-0.0278**
		(0.0127)
Number of Loan Accounts in Previous 12 Months (> 1 ref.)		
	0	0.1070*
		(0.0582)
	1	0.1010*
		(0.0580)
Credit Card Usage (> 20% ref.)		
	<= 20%	-0.0012

		(0.0130)
	No Credit Card	-0.0056
		(0.0123)
Constant		0.2641***
		(0.0656)
pseudo R^2		0.001
N		213194

Notes: This table presents the relationship between receiving a false positive alert regarding the loan usage application being high risk of fraudulently induced and ex-post credit use. The estimation sample is the treatment sample, but with applications actually fraudulently induced excluded from the sample in order to focus on false positives. The dependent variable *CreditUse* is a dummy variable defined to be 1 if the applicant eventually used the credit service from the Platform, and 0 if the applicant did not use the credit service. The main explanatory of interest is a dummy variable *FalsePositive*, which is defined to be 1 if the applicant received an alert call from the Platform despite being not induced by financial fraudsters. The control variables include age, gender, income, loan amount, total credit in the last 12 months (excluding mortgages), days after the last credit report inquiry, credit card utilization rate, and number of loan accounts in the last 12 months. Loan amount ranges from RMB500 to RMB100,000. The last five variables are obtained from external credit records, which are missing for some applicants. The baseline for age is 35 years; that for income is under RMB4,000; and that for loan amount is under RMB5,000. The baseline for total credit in the last 12 months is a credit exceeding RMB20,000; that for number of loan accounts in the last 12 months is 1; that for historic consumer loan amounts is a loan exceeding RMB10,000; that for days after the last credit report inquiry is 1 day; and that for credit utilization rate is a rate of over 20%.

Table 8: Probability of Getting Defrauded: Frauds Targeting Lack of Financial Literacy versus Frauds Targeting Overconfidence

		(1) *Fraud_FL*	(2) *Fraud_OC*	Coefficient Diff.
Female		1.6518***	1.8672***	0.2300
		(0.1672)	(0.3524)	(0.3900)
Age (> 35 ref.)				
	18–25	2.2651***	0.7317*	-1.4978***
		(0.3131)	(0.4147)	(0.5198)
	25–35	1.2708***	0.5558	-0.7090
		(0.3104)	(0.3646)	(0.4789)
Income (0–4K ref.)				
	4K–6K	0.8858***	0.9326	0.0707
		(0.2461)	(0.6458)	(0.6911)
	6K–8K	1.5096***	1.4675**	-0.0082
		(0.2670)	(0.6743)	(0.7262)
	8K–10K	1.6704***	1.2289	-0.3940
		(0.3330)	(0.7644)	(0.8355)
	> 10K	1.2982**	0.8328	-0.4310
		(0.6490)	(1.2160)	(1.3788)
Loan Amount (0–5K ref.)				
	5K–10K	0.3594**	0.5840	0.2334
		(0.1679)	(0.3936)	(0.4278)
	> 10K	0.0809	1.3611***	1.2645***
		(0.1786)	(0.3349)	(0.3797)
Days after the Last Credit Inquiry (> 0 ref.)				
	0	3.5943***	2.0698***	-1.5039***
		(0.2484)	(0.2960)	(0.3861)
	No Previous Credit Inquiry	1.4302***	-0.5105	-1.9401*
		(0.5108)	(1.0366)	(1.1556)
Total Loan in Previous 12 Months (> 20K ref.)				
	0	1.0850	4.9728	3.8967
		(0.7252)	(5.8619)	(5.8864)
	<= 20K	0.5628*	0.6667	0.1035
		(0.3007)	(0.7304)	(0.7897)
Number of Loan Accounts in Previous 12 Months (> 1 ref.)				
	0	1.5337**	-1.3542	2.8853
		(0.7599)	(5.8950)	(5.9237)
	1	0.4274	-2.6704	3.0985
		(0.7451)	(5.8536)	(5.8806)
Credit Card Usage (> 20% ref.)				
	<= 20%	1.8622***	1.7021***	-0.1471

		(0.2426)	(0.4287)	(0.4925)
	No Credit Card	1.7737***	1.9154***	0.1572
		(0.2471)	(0.4450)	(0.5089)
Constant		-14.0791***	-17.5055***	-3.4970
		(0.8871)	(5.9406)	(5.9876)
pseudo R^2		0.347	0.280	
N		213,477	213,477	

Notes: This table presents the characteristics of two types of potential victims: those who lack financial literacy and those who are overconfident. We use the treatment sample in this table as the Platform does not keep fraud type records for the control sample. A logit model is used in columns (1)–(2). The outcome variable in column (1), *Fraud_FL*, is equal to 1 if a fraud case is due to a lack of financial literacy. Similarly, the outcome variable in column (2), *Fraud_OC*, is equal to 1 if a fraud case is due to overconfidence. We record a fraud case if an applicant recognizes the fraud and withdraws their loan usage application after receiving the alert message, or if an applicant manipulated by fraudsters subsequently reports a fraud case to the Platform. The explanatory variables include age, gender, income, loan amount, total credit in the last 12 months (excluding mortgages), days after the last credit report inquiry, credit card utilization rate, and number of loan accounts in the last 12 months. Loan amount ranges from RMB500 to RMB100,000. The last five variables are obtained from external credit records, which are missing for some applicants. The baseline for age is 35 years; that for income is under RMB4,000; and that for loan amount is under RMB5,000. The baseline for total credit in the last 12 months is a credit exceeding RMB20,000; that for number of loan accounts in the last 12 months is 1; that for historic consumer loan amounts is a loan exceeding RMB10,000; that for days after the last credit report inquiry is 1 day; and that for credit utilization rate is a rate of over 20%.