Default Data Manipulation in Marketplace Lending

Yutong Hu* Shasha Li†

Latest draft here

April 2025‡

ABSTRACT

We show that online marketplace lending platforms selectively disclose credit risk, often reporting near-zero default rates, which significantly reduces market efficiency and stability. Using data from a leading online lending platform in China, we reveal default rates under-reported by at least 3.66 percent points monthly, through offline partnerships and risk control funds designed to promise safe returns to investors. This phenomenon is prevalent across the industry. Our further investigation finds that these practices undermine market efficiency, and reduce existing investors' attentiveness while attracting new, less-informed investors with fragile confidence. Moreover, we conducted an event study focused on the Ezubao scandal—a major negative shock to the industry—to examine its impact on market stability, using data from the platform's secondary market transactions. Evidence suggests that hidden default rates significantly raise the risk of investor runs, making the platform more susceptible to liquidity risk and potential collapse when facing shocks.

Keywords: Marketplace Lending, Information Disclosure, Market Efficiency, Investor Run

JEL classification: G14, G23, G33, G50, G51, D82

*London Business School, yhu@london.edu.

†Halle Institute for Economic Research (IWH) and University of Magdeburg, shasha.li@iwh-halle.de.

‡This research is inspired by conversations with Julian Franks. The authors thank Tim Adam (discussant), Patrick Breithaupt (discussant), Svetlana Bryzgalova, Artur Doshchyn, Rosa Ferrer, Julian Franks, Hira Ghaffar (discussant), Alberto Manconi, Anna Pavlova, Darcy Pu, Stefano Rossi, Nicolas Serrano-Velarde, Xu Zhang, Haikun Zhu, Yueteng Zhu (discussant) and conference and seminar participants at 2024 Wharton Financial Regulation Conference, 2024 EEA-ESEM Congress, 2024 EARIE, 2024 ZEW ICT, 2024 EFT Edinburgh, 2023 EWMES, 2023 EEA WinE Mentoring, 2022 HU-IWH Joint Workshop, 2022 CFRC, and IWH for their helpful comments and suggestions. All errors remain ours.

1. Introduction

Financial technology (Fintech) lending platforms, such as Prosper in the United States, provide online marketplaces that facilitate the matching of loan demand and supply. They brought many previously unbanked individuals into the lending sector, which are typically at high credit risk (Tang, 2019). Despite the perceived high risk, most online lending platforms around the world report surprisingly low default rates, often close to zero. For instance, Zopa, the world's first-ever Peer-to-Peer (P2P) lending company in the U.K. claims the default rate since it launched in 2005 is 0.6 percent, which is extremely low compared to the default rates of bank loans which generally exceeds 2 percent[1]. Given that Fintech lending companies typically have less professional expertise and experience than traditional banks, it seems unlikely that these platforms are better at managing loan quality and recovering bad loans.

Puzzled by this phenomenon, we examine the information disclosure practices related to loan default rates in Fintech marketplace lending platforms. We guess that Fintech platforms may be concealing actual loan performance and seek to understand what's behind this and its economic impacts. In this paper, we focus on the Chinese market, which was once the world's largest Fintech lending market before its collapse post-COVID. Using data primarily from a prominent P2P lending platform in China, we provide empirical evidence of default data hiding and estimate the actual default rates. Furthermore, we investigate how this data-hiding practice affects price efficiency and market stability, highlighting its broader implications. We believe the insights gained from studying data from the Chinese market can have general implications for understanding Fintech lending practices globally, as many platforms in other countries exhibit similar practices. Our study contributes to the ongoing discussion on transparency, optimal information disclosure, and the regulatory frameworks necessary for digital financial markets.

Fintech companies are incentivized to disclose key data selectively, and there is a lack of regulation on information transparency. Like banks, which may understate risk or manipulate credit ratings to retain informational rents (Giannetti et al., 2017; Begley et al., 2017), Fintech platforms may also manipulate data for profit or control. Transparency in delinquency and pricing has long

[1] In 2022, the default rate of small and medium-sized businesses debt in the UK is around 2 percent according to the Bank of England (BoE) report.

been an issue in the Fintech lending sector[2]. The lack of regulatory guidance on key data disclosure and the complexity of business models pose risks to financial stability, as seen in the Wirecard fraud in Germany and Ezubao scandal in China. The rapid growth of Fintech and its integration into the financial sector makes stronger oversight and a deeper understanding of associated risks more urgent, especially as banks increasingly embrace Fintech collaborations and digitalization after COVID.

In addition to the lack of regulatory guidance, there is a notable gap in the literature regarding how lending platforms disclose credit risk information for loans traded on their platforms, as well as the corresponding economic impacts. While much of the existing research extensively explores the crucial role of information disclosure design in Fintech lending (e.g., Iyer et al., 2016; Vallee and Zeng, 2019; Franks et al., 2021), the majority of studies focus on how market participants disclose credit risk information themselves, rather than on the platforms that act as market designers and shape the information environment. Furthermore, where research on platforms' disclosure exists, it tends to be theoretical, with a lack of empirical data to support conclusions and discuss economic mechanisms thoroughly. In this paper, we aim to fill this gap by empirically examining the information disclosure practices of online lending platforms in China, analyzing their economic consequences, and further exploring the reasons behind these practices and their underlying mechanisms.

Our primary dataset consists of P2P loan applications, repayment flows, and secondary market transfer records collected from Renrendai, a leading online lending platform in China. It covers all loan applications from January 2012 to December 2016, along with repayment and transfer data up to September 2018. To enhance the analysis, we supplement it with loan-level data from other Chinese lending platforms such as Paipaidai for comparison and a more comprehensive understanding. Renrendai is particularly valuable for addressing our research question due to its uniquely detailed data on loan transactions, repayments, and secondary markets, and its representativeness of the broader industry.

We begin by examining Renrendai's loan repayment performance data and estimate that it underreported default rates by at least 3.66 percent on average per month. Despite an average

[2]In 2016, the U.S. Treasury Department called for greater transparency in online lending, including clear loan performance metrics and standardized data. In July 2022, the Committee on Small Business highlighted the need for better transparency in small business lending, where Fintech plays a significant role.

annual interest rate of about 10 percent—almost twice that of traditional banks—Renrendai's reported default rate was around 3 percent, dropping to near zero after November 2012. Our investigation reveals that this drop coincided with Renrendai's partnership with Ucredit, an offline lending service provider, which assisted loan applications and offered guarantees through risk control funds. This arrangement allowed Ucredit to cover defaults from its risk control fund, making the loans applied through Ucredit appear risk-free on Renrendai's platform. Notably, Renrendai did not disclose these defaults, leading to an artificially low reported default rate.

We then employ regression models to estimate actual default rates based on borrower and loan characteristics, focusing on short pre- and post-partnership windows. To account for changes in borrower composition, we apply Propensity Score Matching (PSM), isolating the impact of Ucredit's involvement. Our analysis includes both the first introduction of offline partnerships and a pooled sample of offline branch openings across multiple cities. The findings show considerable underreporting, with actual default rates estimated to exceed the reported rates by 3.66 to 30.7 percentage points, depending on the time frame and sample used for analysis.

Next, we explore the economic consequences of this manipulation, particularly its effect on market efficiency and stability. To test the impacts on market efficiency, we follow Franks et al. (2021) to regress the default dummy on LGD-adjusted interest rate. In an efficient market, there's a unit coefficient of LGD-adjusted interest rates according to investors' participation constraints. However, our results show that the coefficients are close to zero, indicating significant market inefficiency. Remarkably, the introduction of offline loans further exacerbates this inefficiency, as the coefficient continues to decrease.

To assess market stability, we use the 2015 Ezubao scandal as a natural experiment. Ezubao, a major Chinese P2P lending platform, defrauded 900,000 investors of $7.6 billion, triggering a significant shock to the P2P industry. We compare loan transaction outcomes on Renrendai and Paipaidai during this period. Unlike Renrendai, Paipaidai operates purely online without loan guarantees, serving as our control group. Our findings show that Renrendai loans reacted more quickly to the initial news of the scandal, with interest rates rising immediately, whereas Paipaidai's response came later after the news had gained peak attention. Additionally, we analyze Renrendai's secondary market trading during the scandal, observing increased investor runs and falling share prices. The panic caused by the Ezubao scandal led investors in Renrendai to sell loans, with a significant rise in the number of active sellers and longer withdrawal wait times. This investor run highlights the market's fragility, suggesting that hiding default data amplifies the platform's vulnerability to investor runs and potential collapse.

Lastly, we use detailed micro-level data to explore how hiding information distorts the market and makes it more fragile. After the introduction of offline partnerships, we observe that new investors tend to be more cautious. They prefer placing manual bids over using the platform's automated tools and typically invest smaller amounts, suggesting they lack confidence in the platform. Interestingly, existing investors become less engaged, showing reduced attention to their investments.

The real-world developments further underscore the concerns raised in our analysis. In 2021, Renrendai faced significant overdue debt problems and liquidity risks, exacerbated by the economic disruptions caused by the COVID-19 pandemic. A large number of investors attempted to withdraw their funds but were unable to recover their investments, while the platform struggled to attract sufficient new investors. This situation led to allegations of non-transparency regarding the platform's operational status and financial condition, highlighting the risks associated with the manipulation of default data.

The paper is organized as follows: Section 2 presents a literature review. Sections 3 and 4 describe the background and data. Section 5 presents empirical analysis and discussions. The final section provides the concluding remarks.

2. Literature Review

This paper contributes to several strands of literature. First, this paper contributes in general to the literature on information disclosure of digital platforms, and more specifically to that of Fintech platforms. The platforms's goal is to maximize its profits, which may not always align with user welfare. For example, Rayo and Segal (2010) use advertising platform as context, examines a sender-receiver disclosure game where the sender maximizes profits by selecting an information disclosure rule for prospects with unknown payoffs, showing that without monetary transfers, partial disclosure is optimal, while full disclosure becomes optimal when transfers are introduced, with applications to internet advertising platforms.

Previous studies have extensively discussed the importance of disclosed information for market participants' decision-making (Duarte et al., 2012)

The previous literature mainly about how market participants have a use of the information already disclosed on the platforms but does not discuss thoroughly how platforms should disclose information to market participants. If any of these kinds of discussion in literature exists, it's about theoretical exploration. Di Porto and Zuppetta (2021) discussed qualitatively how digital platforms should disclose information to market participants for P2B transactions and for regulation concerns. They call for disclosure of information for easy understanding (via algorithm) and they call for co-regulation (These algorithms should be tested in a collaborative process involving regulators and stakeholders to make sure they work as intended.) Despite some very brief and overview explorationWang et al. (2021), there's a limited discussion to show empirically and quantitatively how digital platforms should disclose information and what are the economic consequences. Though (Vallee and Zeng, 2019) discussed Marketplace lending platforms optimally increase their own prescreening but decrease information provision to investors to maximize loan volume. Avoyan et al. (2024) pointed out Optimal disclosure on crowdfunding platforms involves a tradeoff between investor protection and innovation (mainly theory work).

highlights how information distribution affects investor performance on lending platforms. We are the first paper to fill in this gap.

Bimpikis and Mantegazza (2023) highlight that platforms can strategically disclose partial demand information to suppliers to optimize entry and pricing decisions, potentially boosting platform profits but at the expense of consumer welfare through higher prices.

We are the first to discuss about the optimal information disclosure for market stability purpose

—Old

This paper contributes to several strands of literature. First, our work relates to the literature on the role of information in the Fintech industry. Previous work, such as (Iyer et al., 2016), shows that nonstandard information can significantly improve default predictions. Wang et al. (2021) analyze data from Chinese P2P lending platforms and demonstrate that increased information disclosure is linked to a reduction in default rates. Franks et al. (2021) find that a British platform's shift from auctions to customer self-set prices increased vulnerability to liquidity shocks and information inefficiencies, driven by noncrowd players. Liao et al. (2021) a Chinese platform and find that investors appear to primarily focus on interest rates to make their investment decisions and largely ignore the credit ratings. Unlike previous studies that mostly focus on borrower self-disclosure during the application process, we examine platforms' disclosure on loan repayment performance.

Second, this research speaks to information manipulation, especially financial data manipulation literature. Most of the studies focus on fraudulent financial reporting (Kirkos et al., 2007; Kaminski et al., 2004; Gillett and Uddin, 2005; Ngai et al., 2011), and some study the lending industry, especially bank lending. For example, Giannetti et al. (2017) find banks manipulate credit ratings before sharing with competitors. Murfin (2012) find that banks write tighter loan contracts after perceiving higher default rates. Wang and Xia (2014) find that securitization-active banks exert less effort on ex-post monitoring. To the best of our knowledge, this paper is among the first to look into data manipulation in Fintech lending. The results from our study suggest a need for improved information disclosure in Fintechs.

Third, this paper contributes to the literature on market stability and run risk in financial markets. Much of the existing literature in this field is theoretical, like Diamond and Dybvig (1983) seminal work on panic runs in banking. Recent research, such as Carletti et al. (2023), discusses the impact of loan guarantees on bank stability. Our study focuses on the role of default performance hiding, primarily through loan guarantee funds in fintech lending platforms, in market fragility. Previous empirical studies on panic runs have primarily centered on the traditional banking sector and mutual funds. Iyer and Puri (2012) studies depositor behaviors in bank runs triggered by panic. Chen et al. (2020) provides evidence of fragility caused by liquidity transformation in the US banking sector. Cookson et al. (2023) delves into social media-fueled bank runs. Additionally, there is a body of literature exploring panic runs in other financial markets, including mutual funds (e.g., Schmidt et al., 2016; Jin et al., 2022). Our paper is the first to investigate empirically investor runs in the Fintech lending market, contributing to the ongoing discussions on fintech regulation.

3. Background

In this section, we describe the platforms and present background information in detail.

3.1. The Platforms

3.1.1. Renrendai P2P marketplace lending platform

Renrendai, founded in May 2010, is one of China's leading P2P marketplace lending platforms. Since its foundation, the loan book of Renrendai has grown rapidly. According to Renrendai's 2012 annual report, the platform's annual online trading volume grew by 803% over two years, and the total amount was 354 million RMB (about $52.8 million). At the end of 2012, Renrendai integrated with UCredit (Youxin in Chinese), a company focusing on offline debt services and founded by the same co-founders of Renrendai. Renrendai switched from purely allowing online loan applications to allowing listings creation on the platform through both the online and offline channels. Renrendai planned to go public at the end of 2012, but the attempt failed. In January 2014, Renrendai successfully financed $130 million, the biggest equity investment in the Chinese P2P lending market. Renrendai used the fund to improve internal operations, enhance risk control capabilities, and recruit talented employees. By the end of 2018, the cumulative trading volume of Renrendai exceeded 76.4 billion RMB (about $11.11 billion).

On Renrendai online lending platform, a borrower who is a Chinese citizen between the ages of 22 to 55 can apply for P2P loans without collateral by providing documents including a credit report from the central bank, an income certificate, a work certificate, and a resident identity card. The borrower can also voluntarily provide additional information such as property ownership certificate, marriage certificate, education background, or credit report from a third agent (e.g., Sesame credit score) to support the loan application.

The platform prescreens P2P loan applications and assigns passed borrowers credit ratings of AA (low risk), A, B, C, D, F, and HR (high risk). The credit rating determines the financing cost and the maximum loan amount. A borrower with a better credit rating can borrow more with a lower fee. After Renrendai verifies the applicant's eligibility, online investors can bid on the shares of the loan request at 50 RMB per share.

To apply for loans, eligible borrowers specify the contract terms, including the loan amount, interest rate, and maturity, and create loan request listings with detailed information online. The loan request pages contain borrowers' self-reported information such as marital status, age, educational background, working years, working industry, company size, and borrowers' historical performance on the platform, including the number of successful applications and the repayment record. Renrendai updates a borrower's credit rating based on the application record and repayment record on the platform and new information provided.

Online investors (i.e., lenders) observe the posted listings with detailed information, including loan contract terms, borrower characteristics, borrower historical credit performance on the platform, and the loan type. After assessing the credit risk, they can offer bids (i.e., lend money) if they agree to the contract terms. Each investor can invest part of a loan amount with a minimum loan part of RMB 50 and in multiple loans. Each listing is visible on the platform for a maximum of seven days. If a listing is not fully funded after seven days, the loan application fails and will be closed. The platform also offers automatic bidding facilities to lenders.

Once the requested loan amount is fully met and the loan is issued, electronic loan agreements are automatically reached between the borrower and online lenders. The Renrendai platform charges the borrower a service fee from 0% to 5% depending on the credit rating and a monthly management fee of 0.3% of the loan amount. Borrowers repay monthly in an equal amount. Early or late repayments may incur punishment fees.

Renrendai's Risk Control Fund and bad debt collection Renrendai recovers the bad debt and protects investors from credit risk by using the risk control fund, also called risk control fund policy, which was first introduced by Hongling Capital in 2011 to the Chinese P2P lending industry and later became very popular in the sample period[3]. In the event of delinquency, Renrendai guarantees to repay the principal to lenders using Risk Control Fund and then tries to manage a resolution by collection calls or messages, in-person visits conducted by partner debt collection agencies, or litigating on behalf of all lenders. The Risk Control Fund (feng xian pei fu jin in Chinese) aims to secure investors' returns and will step in if borrowers are late in their repayments (30 days behind). Renrendai uses the Risk Control Fund money to cover lenders of the defaulted loans for the remaining capital outstanding.

The money from Risk Control Fund comes from Renrendai's initial injection of RMB 210 million and loan servicing fees. Upon approval, borrowers pay a credit rating-based fee, proportional to the loan amount, to the risk control funds. Higher risk of borrowers pay a higher percentage of the

[3]Zopa, the first ever P2P lending company in the world, has similar risk control Fund Policy. Please see https://www.zopa.com/invest/risk/risk control-policy.

loan amount as loan servicing fees and the fees are pooled to the Risk Control Fund. The early, late, bad, or failed repayments of online "Credit" type loans are documented in repayment flows.

Renrendai's cooperation with an offline company From November 2012, Renrendai shifted its business model from pure online-to-online lending to a mix of online-to-online and online-to-offline lending, by cooperating with an offline company Ucredit. At that time, Renrendai integrated with Ucredit, to form a parent company named Renren Ucredit Group.

After that, in addition to borrowers directly applying for the loan online (denoted as "Credit" type, Xin Yong Ren Zheng Biao in Chinese), borrowers can also apply for Renrendai P2P loans through the sister company UCredit's offline branches (denoted as "Field" type, Shi Di Ren Zheng Biao in Chinese). An offline borrower submits loan application materials to the offline offices of UCredit. After verification and prescreening, UCredit's loan officers help create the loan request on the Renrendai platform on behalf of the borrowers, and the listing is denoted as "Field" type. The credit rating for all "Field" type borrowers is A, and the interest rate is usually a fixed rate offered by the loan officer.

Ucredit has its own risk control fund to deal with default risk incurred by "Field" type Renrendai loans. The actual repayment status of "Field" type loans is not disclosed online. All "Field" type loans in the sample period are repaid on time. On the Renrendai website, the "Field" type loans have no delinquency record. In other words, investors investing in "Field" loans are surely protected from the default risk if there is enough risk control fund.

3.1.2. Paipaidai P2P Marketplace Lending Platform: the control group

Paipaidai, founded in June 2007, is China's pioneering P2P marketplace lending platform. Different from Renrendai, Paipaidai operates only online, without any offline branches. Also, Paipaidai does not provide loan guarantees and has no risk control fund or risk control fund.[4] It means that in the event of a loan default, the platform does not repay for the borrower to compensate lenders' losses.

[4] Only very few AA loans are guaranteed (marked as "Pei" in Chinese), and we exclude these AA loans.

4. Data

This section introduces the datasets and presents summary statistics, followed by key motivating facts.

4.1. Loan Data

P2P borrowing data Our analysis in this paper is primarily based on P2P lending data from Renrendai, a leading platform in China. Additionally, we supplement the P2P lending dataset with data from another prominent platform called Paipaidai, which serves as a control group for comparison with Renrendai.

Our Renrendai dataset comprises 862,232 loan applications submitted on the Renrendai P2P lending platform between January 2012 and January 2017. This data is manually collected and accessible to all users on Renrendai's official website. For each borrower, we can observe the repayment flows until September 2018, the loan status (repaid or defaulted) for each loan, and the characteristics of the borrowers. Among Renrendai loan applications, around 40% of loan applications were accepted. At the end of May 2018, the cumulative amount of loan applied was 91.51 billion RMB (about $14.42 billion), and the cumulative amount of loans granted was 55.8 billion RMB (about $8.79 billion) on the Renrendai platform. At the same time, the total loan outstanding is 36.9 billion RMB (about $5.81 billion). In our sample from 2012 to the beginning of 2017, the monthly average growth rate for the amount of loan applied is 9.65%, and the standard deviation is 30.69%. The monthly growth rate for the loans granted is 49.82%, with a standard deviation of 203.11%. The average monthly loan applied is 938 million RMB (about $143 million) with a standard deviation of 804 million RMB (about $119 million). The average monthly loan granted is 353 million RMB (about $52 million) with a standard deviation of 352 million RMB (about $52 million). Figure 1 shows the cumulative loan applied from October 2010 to May 2018. Figure 2 shows the total monthly loan applied.

Table 1 reports the summary statistics for the loan and borrower characteristics. In our sample, the maturities of the loans range from 3 months to 48 months, with a median of 36 months. The maturities for 50% of the loans are 36 months. For the rest of the loans, most of them have maturities of 12, 18, or 24 months. The data also tracks the repayment flows and loan status for each loan. The borrower of a loan repays in Equated Monthly Installments, and the platform documents the repayment status of each monthly flow. In this study, a loan is defined as default when any monthly payment is past due by three months or more. Default rates are measured at the loan application time. According to our definition, the average monthly default rate is 2.51% with a standard deviation of 2.28%. The annual interest rate for the loans granted ranges from 6.6% to 24.4% with an average of 11% and a standard deviation of 1%. The average loan size of a granted loan is 71,000 RMB (about $10,506). The smallest loan size in our sample is 3,000 RMB (about $456), and the largest loan size is 500,000 RMB (about $73,986).

Most of the borrowers on Renrendai are individual borrowers. Our data for borrower characteristics include information about the borrower's age, working status, education, marital status, assets in possession, debts owed, and monthly income. The borrowers' ages in our sample range from 23 to 55, and the average age is 38.19. 50% of the borrowers have at least one flat property, and only 33% of the borrowers have at least one car. The borrowers have a median monthly income of 10001-20000 RMB (about $1575 to $3150). When applying for loans, borrowers need to provide personal information and related documents to allow the platform to assess the risks of the borrowers. The Renrendai platform gives each borrower a credit limit based on the information provided. The credit limits for the borrowers range from 0 RMB to 50 million RMB (about $7.6 million), with an average of 68,000 RMB (about $10,709).

Investor lending data Our investor-level lending data records the amount and time of each lending bid on the Renrendai platform. The data set contains around 40 million lending observations starting from 2010 to 2017. The data set documents investment actions made by 439,801 investors. Each user id uniquely identifies an investor on the platform. The loan ID of each lending record matches to the loan ID in the borrowing data set.

The data set also tracks the channel through which each bid is made. There are three methods for investors to make the bids, which are auto bidding, finance plan bidding, and normal bidding. If an investor uses auto bidding, it means the investor authorizes the Renrendai platform to automatically make bids if a loan meets certain requirements set by the investor in advance. A finance plan bidding appears if an investor participates in a specific investment plan set by the Renrendai platform. Then the platform manages the funds in the finance plan and bids on behalf of the investor. A normal bid is directly made by an investor without any intervention from the Renrendai platform.

4.2. Motivating facts: sudden drop in the default rate on Renrendai platform

In the loan data published by Renrendai, we find a sudden drop in the default rate reported by the platform starting from the end of 2012. In November 2012, Renrendai announced the integration with UCredit and established a parent company called Renren Ucredit Group. UCredit is a financial services company founded by the same co-founders of Renrendai and focuses on offline debt services. The observed default rates on Renrendai dropped dramatically since 2012 Nov. We find that the sudden drop in default rates was mainly due to the zero default rates of "Field" type borrowers from UCredit.

As explained in Section 3.1.1, Renrendai holds a Risk Control Fund to secure lenders' investment. For each loan granted, the platform charges the borrower service fees equaling 0-5% of the loan amount granted. The service fee percentage depends on the borrower's credit rating. The service fees are held in the Risk Control Fund. If repayment is 30 days past due, Renrendai will repay the lender using Risk Control Fund, and the creditorship is transferred from retail online lenders to the platform. The platform tries to collect the money back from the borrower. If the borrower refuses to repay the debt, he or she will be prosecuted. The offline sister company Ucredit also has its own Risk Control Fund to protect investors against credit risk.

The P2P loan data published on the Renrendai website contains both online loans ("Credit" type loans) and offline loans ("Field" type loans)[5]. When a borrower defaults and the repayment is made from the Risk Control Fund, the loan status is reported differently for "Credit" type and "Field" type borrowers. For the "Field" type loans, the default record will not appear in the data published on the official website. This method is completely legal, but that will lead to "manipulations" in the default rates data. Lenders observe the data to help them make investment decisions. However, the lenders cannot tell whether a normally repaid loan is repaid by the borrower on time or repaid by the Risk Control Fund, so the "manipulations" in the reported default rates can affect the market efficiency on the platform.

We first plot the monthly default rates for loans with different maturities in Figure 3. In our

[5] The loan application processes for both types of borrowers are described in Section 3.1

data, 50% of the loans have a maturity of 36 months. For the rest of the loans, 7.81% loans have a maturity of 12 months; 16.53% of the loans have a maturity of 18 months; 18.15% of the loans have a maturity of 24 months. In Figure 3, the default rates for 18-month and 24-month loans drop suddenly after November 2012. The default rates do not change much for the rest of the loans. The default rate for 9-month loans rises slightly after November 2012, but 9-month loans are only 0.72% of the total loans. For the loans that are processed through offline branches, 42.63% have a maturity of 18 months, and 50.17% have a maturity of 24 months.

In Figure 4, we plot the time series of monthly default rates for online borrowers and offline borrowers separately. The plot shows that, for borrowers who post the listings through offline channels, the default rates are always zero from November 2012 to May 2018, which is unusual. In the meantime, the reported mean default rate for online borrowers is 13.97% with a standard deviation of 0.7%. The plots imply that, although the platform uses the Risk Control Fund to repay defaulted loans for both offline and online loans, the platform is not reporting loan default records on online and offline borrowers in a consistent way.

In the next section, we will show that if the platform consistently reported the default records for both types of borrowers, the actual default rates for the offline borrowers should be much higher than zero. Figure 5 shows the monthly numbers of loans granted to online and offline borrowers. By comparison, the number of loan applications posted through offline branches is always roughly 20 times the number of loans applied directly online. The plot indicates that most of the default records that are visible to the investors are under-reported. In the next section, we also explore the impact of the default rate "manipulation" on market efficiency and market stability.

5. Empirical Analysis

In this section, we outline the empirical analysis. First, we detect default data manipulation on Renrendai. Second, we assess its impact on market efficiency and stability. Finally, we analyze secondary market data to explore how data-hiding influences investor behavior.

5.1. Detect Default Data Manipulation

5.1.1. Method

Renrendai uses the borrowers' characteristics to decide the loan risk level. We assume that if the default rate data is reported consistently before and after Renrendai cooperating with UCredit, borrowers' characteristics should consistently predict loan default rates before and after the event. We set November 21 2012 as the breaking point when Renrendai announced the integration with UCredit and introduced "Field" type loans to the platform. To focus on the event effect of Renrendai integrating with UCredit, we begin with 2-month windows before November 21 2012 to estimate the default prediction model and assume it to be the true model. The reason for using a 2-month window is that the true model will not likely change within a short period. In the Appendix A2, we expand the window to 6 months to have more observations. After training the model on the observations in the pre-window, we use the estimated results to forecast the default rates after November 2012 in a post-window of equal length.

The regression specification is:

$$\begin{aligned} Default_{it} = \quad & \beta_0 + \beta_1 \times CompanySize_{it} + \beta_2 \times MaritalStatus_{it} + \beta_3 \times Education_{it} \\ & + \beta_4 \times WorkExperience_{it} + \beta_5 \times Age_{it} + \beta_6 \times Gender_{it} + \beta_7 \times Property_{it} \\ & + \beta_8 \times HouseMortgage_{it} + \beta_9 \times Income_{it} + \beta_{10} \times Car_{it} + \beta_{11} \times CarMortgage_{it} \\ & + \beta_{12} \times LoanAmount_i + \beta_{13} \times LoanMaturity_i + \beta_{14} \times InterestRate_i + \eta_{it} \end{aligned} \tag{1}$$

where $Default_{it}$ is a dummy variable equal to 1 if the borrower defaults and 0 otherwise. The explanatory variables are borrower characteristics. $CompanySize_{it}$ is a categorical variable indicating the number of employers in the borrower's working company. $MaritalStatus_{it}$ is a dummy variable equal to 1 if the borrower is married and 0 otherwise. $Education_{it}$ is a categorical variable indicating the education level for the borrower. $WorkExperience_{it}$ is a categorical variable indicating the number of years the borrower has worked. Age_{it} and $Gender_{it}$ are the age and gender of the borrower. $Property_{it}$ is a dummy variable equal to 1 if the borrower owns at least one property and 0 otherwise. $HouseMortgage_{it}$ is a dummy variable equal to 1 if the borrower has at least one house mortgage and 0 otherwise. $Income_{it}$ is a categorical variable indicating the borrower's income level. Car_{it} is a dummy variable equal to 1 if the borrower owns at least one car and 0 otherwise. $CarMortgage_{it}$ is a dummy variable equal to 1 if the borrower has at least one car mortgage and 0 otherwise. $LoanAmount_i$ is the amount of loan applied, and $LoanMaturity_i$ is the loan's maturity. $InterestRate_i$ is the interest rate of the loan. η_{it} is the error term. The detailed descriptions of the borrower characteristics can be found under Table 1.

Note that we do not control for the credit rating given by Renrendai in our regression because offline borrowers are all assigned A ratings by the platform. The ratings assigned to the offline borrowers cannot distinguish the creditworthiness of the borrowers. While detecting the default rate manipulation, we control for weekly or monthly fixed effects to rule out the possibility that change in default rates could be due to industry trends or consumer awareness of the P2P lending industry, which may change over time.

We also consider that the quality of borrowers may change before and after the two companies' consolidation, which could cause changes in the default rates. The first half of Table A1 compares the changes in borrower characteristics in the pre-window and post-window. According to the different test results on the borrower characteristics before and after the event, we find that, on average, borrowers in the post window have lower education levels, work in smaller companies, and have less property. In the meantime, the borrowers in the post window have slightly higher income on average.

To rule out the effect of changed borrower characteristics, we use Propensity Score Matching (PSM) to match the observations before and after the event based on the borrowers' characteristics. PSM matches the borrowers in the post window to similar borrowers in the pre window. Using matching results, we can proxy whether a borrower in the post window with certain characteristics is likely to default if there is no treatment effect of the data "manipulation" in the borrowers' default rates after the company merge. Then we estimate the default rate prediction model based on the matched sample. Next, we use the estimated model to predict the default rates for the borrowers in the post window and compare the estimated default rates with the default rates reported. We first match the borrowers to implement PSM based on all the characteristics included in regression (1). We then change the matching criteria to ensure the robustness of the Propensity Score Matching results.

5.1.2. Default Rate Estimations

Table 2 shows the OLS regression results using 2-month windows before and after the event. Column (1) shows the regression results on the loans from October 2012 to November 21 2012. Column (2) shows the regression results on the loans from the end of November 2012 to January 2013. The table shows that the significance of the coefficients is slightly different before and after the event. The loan amount and the indicator of whether a borrower owns a property lose their prediction power in the post window. In the post window, the education level and borrower income are significant predictors of default rates. Interestingly, the positive coefficients on borrower income imply that higher-income borrowers are more likely to default on the Renrendai platform. The coefficients of interest rate are always intuitively and positively predict default rates.

In Figure 6, we plot the weekly default rates in the data together with the predicted weekly default rates. The left figure is based on the regression results in Column (1) of Table 2, and the right figure shows PSM results in Column (3). We use weekly default rates for the 2-month window figures because the window is too short for using monthly rates. For other plots with longer windows, we consistently use monthly default rates. The predicted default rates are based on the regression results for the pre-window data. The red vertical line indicates the event time, November 21 2012. The plot shows that after November 21 2012, the predicted default rates are consistently higher than the published default rates. To compare the published default rates and the predicted default rates in the 2-month post window, we implement a t-test with the null hypothesis that the predicted default rates equal the published default rates. The mean of the difference between the predicted default rates and published default rates in the post window is reported in Table 2. The results indicate that the predicted weekly default rates are, on average, 7.8 percentage points significantly higher than the published default rates.

Column (3) of Table 2 reports the regression results on the propensity-score-matched sample using all borrower and loan characteristics included in Equation (1). The second graph in Figure 6 plots the predicted default rates based on the PSM results versus the published default rates. The predicted default rates are consistently higher than the reported default rates after November 21, 2012. The results of the difference test show that, in the post window, the predicted default rates are, on average, 30.7 percentage points significantly higher than the published default rates.

In addition to OLS regressions, we also use logit regressions and LASSO to predict default rates using the same set of borrower and loan characteristics. Both the logit and LASSO regression results show that the predicted default rates are consistently higher than the default rates reported by the lending platform. The results support our conclusion that the Renrendai platform did not fully reveal the default records.

We then implement a series of analyses to ensure the robustness of our findings. Table A2 shows the OLS regression results using samples in the 6-month pre- and post-windows. In Figure 7, we plot the monthly default rates in the data together with the predicted monthly default rates. Both the plot and the difference test results show that after November 2012, the predicted default rates were consistently higher than the published default rates.

To rule out that changing the matching criteria could influence the PSM regression results, we change the propensity score matching criteria in this section to check for the robustness of the PSM regression results. Table A3 and Table A4 in the Appendix show the regression results on samples matched using different criteria. For example, for column (1) of Table A3, which is labeled as "No Company Size", the matched sample is matched on all the other nine borrower characteristics except for *Company Size*. The t-tests on the differences between predicted and reported default rates also consistently show that the predicted default rates are always higher than the reported data.

We also apply the Rosenbaum Bound Test to the difference between the reported default rates and predicted default rates using estimation in Column (3) of Table 2. Rosenbaum Bounds examines whether the average treatment effects on the treated are sensitive to hidden bias. Table A5 shows the Rosenbaum Bounds Test results. The table shows that the lower and upper bounds for predicted default rates minus reported default rates are significantly positive. The sign of this difference is insensitive to different levels of hidden bias measured by gamma.

In general, both the baseline regression results and the PSM regression results support that the default rates reported by Renrendai after November 21 2012 are substantially lower than they should be if they report the default rates as truthfully as before the company consolidation.

5.1.3. Offline Versus Online Borrowers

Before November 2012, the platform only allows borrowers to submit loan applications online directly. After Renrendai consolidated with UCredit in November 2012, Renrendai started to allow borrowers to submit loan applications through offline branches. In Figure 4, we show that after November 21, 2012, the default rates for offline borrowers are reported to be zero, while the average reported default rate for online borrowers is 13.97%. The plot implies that, while the default records for offline borrowers are "manipulated", the default records for online borrowers are relatively more truthfully reported.

In this section, we no longer examine the default rate "manipulation" for all of the borrowers at the same time. Instead, we assume that the default rates for online borrowers are correctly reported and focus on detecting default rate "manipulation" on the offline borrowers. We use the default records for the online borrowers and borrower characteristics to proxy the true default records for the offline borrowers. We first assume that online borrowers' default rate prediction model is true for offline borrowers if there is no data "manipulation". Since offline borrowers start to exist after November 2012, we only focus on the loan data after November 2012 in this section.

First, we run the regression specified in (1) on all the online borrowers after November 2012, and use the regression results to predict the default rates for the offline borrowers. Then we conduct PSM analysis and match the offline borrowers to the online borrowers based on the borrower characteristics and use the default records for the matched online borrowers to proxy the true default records for the offline borrowers.

Table 3 reports the mean difference between the reported and predicted default rates using different time windows and controlling different fixed effects. The mean difference between the predicted and reported default rates are all significantly positive and larger than 15 percentage points. The evidence shows that assuming the default rates for online borrowers are truthfully reported, Renrendai significantly under-reports the monthly default rates for offline borrowers after November 2012 by at least 15 percentage points roughly.

5.1.4. Offline Branches

Renrendai's sister company Ucredit started to establish offline branches all over China at the end of 2012. Borrowers can choose to submit their loan requests with the help of Ucredit offline branches if the branches are locally available. The offline loan officers create loan request listings on the Renrendai platform on behalf of verified borrowers. Of course, borrowers can still directly request P2P loans online by themselves. We find that loans applied via offline branches generally have default rates close to zero. On China Judgements Online, which is the official site of the Chinese government that publishes court judgment documents, we find a bunch of judgment documents regarding lawsuits against Renrendai borrowers for defaults on loans. However, those borrowers have no default records in the data published by Renrendai.

In the previous subsections, we have found evidence that the default rates for offline borrowers are largely manipulated downward after Renrendai's partnership with Ucredit. In this subsection, we further explore the impact of offline branch establishments on the default rates and quantify how much the default rates are under-reported because of the introduction of offline branches.

We use a difference-in-difference approach to investigate how the opening of offline branches affects the default rates for the borrowers near the corresponding offline branches. We first gather the opening time and city-level locations for Renrendai offline branches established from the end of 2012 to the beginning of 2017, which are in the same time range as our loan application level data. For the cities that have multiple offline branches opening at different times, we only consider the first offline branch opened in the same city.

For each offline branch, we assume that the offline branch only affects the borrowers in the same city, and borrowers in other cities are not affected due to the cost of transportation. So we set the loans applied by the borrowers in the same city as the treatment group. To investigate the effects of the offline branch opening, we include the observations in the same city and within two months before and after the offline branch opening time.

For the control group, we use the loans originated in the same time period and applied by the borrowers in the same province but not the same city as the offline branch. Because of transportation costs, an offline branch has highly limited or almost no impact on the borrowers that are in different cities. In the meantime, the borrowers in other cities of the same province are in similar economic environments to those in the city with the offline branch. Within the six months before and after the offline branch opening date, it is rational to assume that the borrowers in the control group experienced the same economic and political shocks as the borrowers in the treatment group, except for the shock of opening the offline branch.

The following shows the difference-in-difference regression equation:

$$\begin{aligned} Default_{it} = \quad & \beta_0 + \beta_1 \times CompanySize_{it} + \beta_2 \times MaritalStatus_{it} + \beta_3 \times Education_{it} \\ & + \beta_4 \times WorkExperience_{it} + \beta_5 \times Age_{it} + \beta_6 \times Property_{it} + \beta_7 \times HouseMortgage_{it} \\ & + \beta_8 \times Income_{it} + \beta_9 \times Car_{it} + \beta_{10} \times CarMortgage_{it} \\ & + \beta_{11} \times LoanAmount_i + \beta_{12} \times LoanMaturity_i + \beta_{13} \times Distance_{ij} \\ & + \beta_{14} \times Treat_{ij} + \beta_{15} \times Post_{ij} + \beta_{16} \times Treat_{ij} \times Post_{ij} \\ & + \beta_{17} \times RelativeMonth_{ij} + \eta_{it} \end{aligned} \tag{2}$$

$Treat_{ij}$ equals 1 if observation i is in the treatment group regarding offline branch j and equals 0 if the observation is in the control group. $Post_{ij}$ equals 1 if loan i originated after the corresponding offline branch j opens and equals 0 if it originated before the offline branch opens. $RelativeMonth_{ij}$ is the number of months between the loan origination date and the offline branch j's opening date. Negative values indicate that the observations are in the before period, and positive values indicate that the observations are in the post period. For example, if loan i is originated one month before offline branch j opens, then $RelativeMonth_{ij}$ equals -1. If loan i is originated two months after offline branch j opens, then $RelativeMonth_{ij}$ equals 2. $Distance_{ij}$ measures the driving distance from the borrower i's location to the corresponding offline branch j. We control for the fixed effects of the loans' origination time, the fixed effects of offline branches' opening time, and the fixed effects of borrowers' living cities. Standard errors are clustered at loan borrowers' living cities level.

The regression results for Equation (2) are reported in Column (1) of Table 4. The coefficient for the interaction term is negative and highly significant. The magnitude and sign of the coefficient indicate that the opening of an offline branch is on average associated with a 7.2 percentage point decrease in the reported default rate for the loans applied by the borrowers in the same city.

To estimate how much the default rates are under-reported for the treatment group, we assume that the borrower and loan characteristics predict the true model in the 2-month windows before offline branches open. We run the regression based on equation (2) excluding $Treat_{ij}$, $Post_{ij}$, and the interaction term. Then, we use the regression results to predict the default rates for the observations in the post-window. According to our assumption, the prediction results reflect the default rates without data manipulation for the observations in the post-window. The regression results are shown in Column (2) of Table 4. After comparing the predicted default rates to the reported default rates in the post-window, we find that default rates are, on average, significantly under-reported by 3.66 percentage points monthly.

We also extended the sample using 6-month windows before and after the introduction of each offline branch. The results are reported in Appendix A6. Our findings are robust using 6-month windows. The opening of offline branches is on average associated with an 8.7 percentage points decrease in the reported default rate for the loans applied by the borrowers in the same city. The default rates are significantly under-reported by 4.98 percentage points monthly.

In the next section, we will explore how the manipulation of the default rates impacts the market efficiency on the platform.

5.2. Market Efficiency Test

As shown above, the Renrendai platform hides information about default rates and largely under-report the credit risk of P2P loans. This section further explores the consequences of default data manipulation. In particular, we would like to investigate the impact of default data hiding on market efficiency.

The direction of the effect is not straightforward. On the one hand, disclosing bad debts may trigger a fear run, and thus, hiding negative information would improve the market efficiency if market participants have fragile beliefs. On the other hand, platforms hide the actual financial condition and credit risk, leading to less reliable information available to market participants and thus may bias their decision-making.

To test market efficiency, we use the Efficient Markets Hypothesis (EMH) specification following Franks et al. (2021) derived from the lender's participation constraint.

We start from the six months before and six months after the November 2012 window period. We use 205,278 monthly repayment performances of propensity-score-matched loans originated on Renrendai between 2012 June and 2013 May to test market efficiency. Among them, 7,996 observations are in the pre-change period, and 197,282 observations are in the post-change period with offline loans. P2P loans drop out of the sample one month after defaulting or when they mature. Note that we observe repayment performance until September 2018.

Furthermore, we take advantage of the gradual establishment of Ucredit branches in Chinese cities to see whether and understand how market efficiency of the online lending market is affected by the introduction of "Field" type loans. During the sample period, 75 Chinese cities opened Ucredit offline branches to acquire offline borrowers for Renrendai. We keep a 6-month before and 6-month after window relative to the introduction of the offline branch for each city of these 75 cities separately and construct a control group for each city by keeping the loan applications from the same province but without an offline branch in the borrower's city in the same time.

5.2.1. The EMH Specification

As in Franks et al. (2021), the main Efficient Market Hypothesis specification is derived from the lenders' participation constraint. A risk-neutral lender makes participation decisions based on the following conditions,

$$1 + \rho = (1 - \pi_i^e)(1 + r_i) + \pi_i^e(1 - LGD_i^e)(1 + r_i) \tag{3}$$

where ρ and π represent risk-free rate and loan i's probability of default respectively. After liberalizing, we can get the one-to-one relationship between the Loss Given Default (LGD)-adjusted interest rate and the expected probability of default, $\pi_i^e \approx \alpha^\star + r_i^\star$ where $\alpha^\star = -\frac{\rho}{LGD_i^e}$ and $r_i^\star = \frac{r_i}{LGD_i^e}$. In other words, if the market is efficient, a 1% increase in LGD-adjusted interest should predict a 1% higher probability of default.

The constraint gives us the benchmark regression equations for the market efficiency test. The explanatory variables include the borrower's credit ratings and the LGD-adjusted loan interest rate, and the dependent variable is a credit default dummy. We want to see whether the interest rate can predict default performance and whether the coefficient of the LGD-adjusted interest rate is close to one.

$$
\begin{aligned}
Default_{it} = \quad & \beta \times r_i^{\star} + \theta \times FE\_Rating_i + \gamma \times FE\_MIssue_i \\
& + \delta \times X_{it} + \lambda \times FE\_City_{it} + \epsilon_{it}
\end{aligned} \tag{4}
$$

where $r_i^{\star}$ is the LGD-adjusted interest rate. FE_Rating and FE_MIssue are fixed effects (FEs) for the borrower's credit ratings and the month of issuing the loan i. X_{it} captures the platform's market performance at the loan's origination month t, including last month's trading volume in terms of the number of loan applications and the total value of the applied amount. We also incorporate borrowers' city-fixed effects to eliminate all time-invariant unobservable or observable confounding factors related to the city's economic conditions and other factors. The standard errors are robust and clustered at the borrower level.

As we can observe monthly repayment performance data but only until September 2018, we adjust the specification to deal with the possible truncation problem,

$$
\begin{aligned}
Default_{it} = \quad & \beta \times r_i^{\star} + \theta \times FE\_Rating_i + \gamma \times FE\_MIssue_i \\
& + \eta \times FE\_SLife_{it} + \delta \times X_{it} + \lambda \times FE\_City_{it} + \epsilon_{it}
\end{aligned} \tag{5}
$$

where $r_i^{\star}$ is the LGD-adjusted interest rate. FE_Rating and FE_MIssue are fixed effects (FEs) for the borrower's credit ratings and the month of issuing the loan i. FE_SLife_{it} are fixed effects of loan i's life cycle, into three equal stages, *Early*, *Mid* and *Late*. For example, for an 18-month loan, the first six months are on its *Early* stage, the second six months are on its *Mid* stage, and the remaining six months are on its *Late* stage. X_{it} captures the platform's market performance at the loan's performance month t, including last month's trading volume in terms of the number of loan applications and the total value of the applied amount. We also incorporate borrowers' city-fixed effects to eliminate all time-invariant unobservable or observable confounding factors related to the city's economic conditions and other factors. The standard errors are robust and clustered at the borrower level.

If the market is efficient, we expect to have a unit coefficient of the primary variable of interest,

β, in regression models (4) and (5). If the market is efficient, the LGD-adjusted interest rate can predict the loan's default performance.

5.2.2. *LGD-adjusted Interest Rate, Periodic Default Dummy, and the First Window Period*

To begin with, we focus on the 6-month before and 6-month after November 2012 window period and use 205,278 monthly repayment performance of propensity-score-matched loans originated on Renrendai between June 2012 and May 2013 to test market efficiency. As in Franks et al. (2021), the stacked regression methodology we use takes into consideration the timing of repayment and the possible truncation problem in the data[6].

We augment the regression model (5) with an interaction term, $Post \times r_i^\star$ where $Post$ dummy represents the time after (including) November 2012 when Renrendai started to cooperate with the offline sister company Ucredit.

$$\begin{aligned} Default_{it} = \quad & \beta_0 \times r_i^\star + \beta_1 Post \times r_i^\star + \theta \times FE\_Rating_i + \gamma \times FE\_MIssue_i \\ & + \lambda \times FE\_SLife_{it} + \delta \times X_{it} + \lambda \times FE\_City_{it} + \epsilon_{it} \end{aligned} \tag{6}$$

Table 8 shows the results of the market efficiency test using equations (6). The market is not efficient. To reduce the concern that the LGDs are computed rather than observed, we test the robustness by bootstrapping standard errors, and we obtain the same main results as shown in Table 9.

LGD Construction Before testing equation (6) we need an estimated LGD_i^e for each loan i. To construct LGD_i^e, we start by estimating the pre- and post-default recovery rate. First, we estimate the default recovery rate by running the following regression with a periodic default dummy as the dependent variable,

$$Default_{it} = \theta \times FE\_Rating_i + \gamma \times FE\_MIssue_i + \lambda \times FE\_SLife_{it} + \epsilon_{it} \tag{7}$$

where $Default_{it}$ is loan i's default dummy at performance month t. As before, FE_Rating and

[6]See Soyeshi (1995) and Cameron and Trivedi (2005) for a more comprehensive discussion about this methodology

FE_MIssue are fixed effects (FEs) for the borrower's credit ratings and the month of issuing loan i. FE_SLife_{it} is loan i's life cycle fixed effects.

Table 5 reports results. It shows that bad repayment performance is almost evenly distributed across three loan life cycle stages. AA or A rating loans have a significantly lower default probability.

Following Franks et al. (2021), we can compute the annualized unconditional default probabilities and the pre-default recovery rates as shown in Table 7. Panel one is the estimated default probabilities for different ratings. AA or A rating borrowers have a default probability of 1.36%, while the HR rating borrowers have the highest likelihood of 15.7%.

Because P2P loans on Renrendai are fully amortized with equal monthly payments, loans that default early have already repaid one-sixth of the debt, that default middle has repaid 50%, and that default late has repaid around 83.3%. Thus we can calculate the pre-default recovery rates based on the life-cycle patterns we get from the regression (7), shown in the last column of Table 7. For example, for an HR-scored loan, we obtain the life-cycle pattern of default probabilities $(33.08\%, 33.59\%, 33.33\%))$ by normalizing the unconditional probabilities $(1.3\%, 1.32\%, 1.31\%))$ by the overall probability of default 3.93%. That is, the P2P loan has a probability of 33.08% that default early, a probability of 33.59% that default middle, and a probability of 33.33% that default late. We multiply the vector of the life-cycle pattern probabilities by a vector of conditional recovery rates $(\frac{1}{6}, \frac{3}{6}, \frac{5}{6})$. The first column of Panel two on Table 7 shows the results of pre-default recovery rates for all credit ratings.

$\langle$*insert Table 5 and 7 here*$\rangle$

Second, to get the post-default recovery rate, we now focus on 147 default events. We run the regressions for the sample in the whole window,

$$RRecoveryPost_i = \alpha + \theta \times FE\_Rating_i + \psi \times MRecovery_i + \gamma \times FE\_MIssue_i + \epsilon_i \tag{8}$$

where $RRecoveryPost_i$ is loan i's post-default recovery rate, which equals to post-default recovered value at the end of the sample divided by the balance remaining at the point of default. $MRecovery_i$ is the natural logarithm of the length of the recovery period, the number of months from the point of default to observation time in September 2018. FE_Rating and FE_MIssue are fixed effects (FEs) for the borrower's credit ratings and the month of issuing the loan i[7].

Table 6 reports results. Post-default recovery rates on Renrendai are around 13.7% to 1, with B-rating loans having the lowest recovery rates and AA or A-rating loans fully recovered.

⟨*insert Table 6 here*⟩

Results of Market Efficiency Test Table 8 shows the results of the market efficiency test (equation (6)). Table 8 indicates that loan prices cannot significantly predict the default event, and the magnitude is far from 1. Though not statistically significant, the negative coefficients of the interaction term of LGD-adjusted interest rate and post dummy suggest that the deviation from information efficiency increases after introducing manipulated loans originated through an offline sister company. The default manipulation drives the market further away from information efficiency. Thus, this online market is not informationally efficient ex-ante and even worse in terms of market efficiency after default data manipulation. The liquidity measures significantly predict the active loans ' default instead of borrowers' credit rating or LGD-adjusted interest rate. We will discuss the effects of liquidity later.

Liquidity In this paragraph, we study the role of liquidity shock in market efficiency. Franks et al. (2021) finds that liquidity shocks can drive interest rates away from fundamental values. As they did, we augment the benchmark regression 6 with liquidity measures and other control variables such as active bid share. Columns (2), (4), and (5) of Table 8 show the results.

We have two liquidity measures. One is the finishing time of the bids for a loan, denoted as $finit$, calculated by the difference between the first and last bidding. The shorter the finishing time, the more liquid the market is. The other proxy for liquidity is agg_weekly_borrow, the total value of the loans originated in the seven days that the loan request listing i is open, normalized by last month's loan book. In addition, we also check the role of automatic bids and incorporate a loan's active bidding share in the augmented regressions.

Similarly, we find that liquidity shocks further drive the prices away from the market-efficient level. The significant negative coefficient of finishing time implies a high default probability when

[7] $MRecovery_i$ is omitted in the regression because of co-linearity

there is high liquidity. The interest rate should be adjusted upward to restore market efficiency. In other words, liquidity shock drives down the interest rate on loans listed in that period below their market-efficient level. The direction of the effects of liquidity is different from that of Franks et al. (2021). One possible explanation is that higher liquidity comes from the higher risk-taking of the platform, and Renrendai introduced too many low-credit-quality borrowers to the market.

5.2.3. Offline Branches and Window Periods relative to cities

The sister companies of Renrendai opened offline branches in different Chinese cities at different times. In this section, we take advantage of this staggered introduction of P2P loans from offline branches to identify the impact of default data manipulation on online market outcomes, especially market efficiency. We construct a pooled sample for 6 months before and 6 months after the offline branch introduction window period as described in subsection 4.1.5. The pooled sample consists of separate sub-samples constructed relative to each treatment city with offline branches opening between January 2012 to December 2016. For each treatment city, we keep the observations in the window period, which is 6 months before and 6 months after establishing the offline branch in that city. In each sub-sample, the control sample consists of all applications from cities in the same province of the treated city but without offline branches open in the same window period. We pool all the sub-samples together to get the final sample.

Market Efficiency Tests Online "Credit" type Renrendai loans have relatively shorter maturity than offline "Field" type Renrendai loans, and we can observe full repayment performance of all "Credit" type Renrendai loans at the observation time of September 2018. Thus we can do the MEH test at the loan level directly. We adopt the Different-In-Difference analysis by augmenting the main MEH test specification (4) and run the following regression,

$$\begin{aligned} Default_{it} &= \beta_0 \times r_i^\star + \beta_1 \times PostEstab \times Treat \times r_i^\star \\ &\quad + \beta_2 \times PostEstab \times r_i^\star + \beta_3 \times Treat \times r_i^\star \\ &\quad + \theta \times FE\_Rating_i + \gamma \times FE\_EstabDate_i + \delta \times X_{it} + \lambda \times FE\_City_{it} + \epsilon_{it} \end{aligned} \tag{9}$$

where the new dummy variable $PostEstab$ represents the time after the introduction of the offline branch in the borrower's city and $Treat$ represents the borrower has access to offline branches opening in his/her city. $FE\_EstabDate_i$ is the corresponding sub-samples opening date of the offline branch in that treatment city. The variable of interest is $PostEstab \times Treat \times r_i^\star$, and we want to see whether its coefficient is significantly different from zero and in which direction.

LGD As we can already observe the full performance of online "Credit" type loans, it is possible to calculate the average LGDs of different credit ratings. We use the calculated credit rating based on average LGDs to get the LGD-adjusted interest rates $r_i^\star$. The average LGDs of AA, A, B, C, D, E and HR level loans are 61.64%, 77.38%, 62.5%, 59.07%, 49.96%, 53.04% and 59.95% respectively.

Market Efficiency and Offline Branch Openings Table 10 reports the results of MEH tests using equation (9). Negative coefficients of $PostEstab \times Treat \times r_i^\star$ in the seventh row of columns (1) and (2) show that after introducing offline branches of Renrendai sister's companies, the price of pure online loans contains less information about the loan's probability of default. The online lending market has become less efficient. In the online sample, the LGD-adjusted interest rate can significantly predict the default performance of the loan. However, the interest rate level is far from market efficiency as the coefficient of LGD-adjusted interest rate is around 0.03 instead of 1.

In order to check the persistence and dynamic change of the effects, we replace $PostEstab$ with the factor variable $RelativeMonth$ described in section 4.1.5. $RelativeMonth$ takes integer values from -6 to 6. If $RelativeMonth$ equals -1, the loan origination time is one month before the corresponding treatment city's offline branch opening.

$$
\begin{aligned}
Default_{it} \quad = \quad & \beta_0 \times r_i^\star + \beta_1 \times FE\_RelativeMonth \times Treat \times r_i^\star \\
& + \beta_2 \times FE\_RelativeMonth \times r_i^\star + \beta_3 \times Treat \times r_i^\star \\
& + \theta \times FE\_Rating_i + \gamma \times FE\_EstabDate_i + \delta \times X_{it} + \lambda \times FE\_City_{it} + \epsilon_{it}
\end{aligned} \tag{10}
$$

Figure 8 plots the coefficients of each $FE\_RelativeMonth \times Treat \times r_i^\star$ level in the regression model (10) and the corresponding confidence intervals at the level of 1%. The treatment effects only appear after introducing offline branches, and the effect is significant and negative, driving the online lending market further away from market efficiency. As any combination of coefficients of $r_i^\star$ related variables is different from one, the market price of loans is not at the efficient level.

5.2.4. Other results

Pure Online Borrowers Using the pooled sample, we use the following regression model to check the dynamic change in market outcomes after the introduction of offline branches.

$$\begin{aligned} Y_{it} &= \beta_1 \times FE\_RelativeMonth \times Treat \\ &+\beta_2 \times FE\_RelativeMonth + \beta_3 \times Treat \\ &+\gamma \times FE\_EstabDate_i + \delta \times X_{it} + \lambda \times FE\_City_{it} + \epsilon_{it} \end{aligned} \tag{11}$$

Where $RelativeMonth_{ij}$ is the number of months between the loan origination date and the offline branch j's opening date. Negative values indicate that the observations are in the before period, and positive values indicate that the observations are in the post period. $FE\_EstabDate_i$ is the loans' origination time fixed effect.

We begin with the successful loan application sample to test the model (11). Figure 9 plots the coefficients of each $FE\_RelativeMonth \times Treat$ level in the regression model (11) with the outcome variable the fraction of "Credit" type loans and the corresponding confidence intervals at the level of 1%. The figure shows a shrinking fraction of "Credit" type loans after the introduction of offline branches. However, Figure 10 shows no clear and significant decrease in the number of "Credit" type loans. It suggests the shrinking fraction is likely caused by the extensions of "Field" type loans. In addition, we use the latest credit scores of borrowers in the performance time of September 2018 to help understand whether there is a borrower composition change in the online market. Figure 11 shows a significant change in borrower credit quality and a decline in credit score on average for the cities with offline branches.

After that, we use the sample of all applications, including the failures to test model (11) and find similar results for the quality of borrowers (see Figure 12). In addition, Figure 13 and 14 suggest that online applications become less attractive with the introduction of offline type loans, and both the success rate and the finishing ratio are lower for applications from cities with offline branches.

5.3. Market Stability Test: 2015 Ezubao Scandal

In this part, we investigate the impact of strategically disclosing default rates on market stability. We exploit the 2015 Ezubao scandal, a substantial negative news shock to the P2P industry, as a natural experiment (Albrecht et al., 2017; Cheng et al., 2022). As previously mentioned, Renrendai utilizes risk control funds and offline branch-assisted loans to obscure actual default rates. We compare the loan transaction outcomes on Renrendai, where default data is obscured through risk control funds and offline loans, to Paipaidai, another platform that operates purely online and has no risk control funds during the Ezubao scandal shock period. Furthermore, we take a close look at the secondary P2P loan transfer market on Renrendai. We analyze trading activities in the secondary market to understand how default rate hiding may affect market stability.

2015 Ezubao scandal In December 2015, another major P2P lending platform in China, Ezubao, was exposed as a Ponzi scheme, failing to repay as much as Rmb38bn ($5.5bn) to investors. Between December 7 and December 13, 2015, a surge of news emerged, reporting that Ezubao was suspected of violating the law and is currently under investigation. Figure 17 portrays the Baidu search index for "Ezubao" over time. Notably, this visualization elucidates that the Ezubao scandal began to gain dramatic public attention around the middle of December 2015. Similar to Cheng et al. (2022), we test how Ezubao shock affects Renrendai loan tradings.

Ezubao achieved popularity by making false promises of high returns and leveraging extensive advertising channels such as state-run media. Many investors regarded this platform as a promising investment opportunity before the truth unfolded. The scandal and the subsequent investigation eventually led to Ezubao's shutdown in December 2015. The Ezubao scandal sent negative shock-waves through the P2P lending market. Renrendai investors reacted to this negative shock by running away from the platform through the secondary market on the same platform.

Renrendai vs Paipaidai Following Cheng et al. (2022), we take December 16 of 2015 as the event day for Ezubao scandal. This aligns with the peak observed in the Baidu index for "Ezubao" news search (Figure 17). The basic regression specifications for the event studies are as below,

$$Y_{i,t} = \sum_{-30 \leq p \leq 30} \beta_p D_{i,t-p} + \alpha_i + \delta_t + X_{i,t} + \epsilon_{i,t} \tag{12}$$

where $D_{i,t-p}$ is a time indicator variable, meaning that the Ezubao scandal event took place p periods before loan i's calendar time t. β_p captures the dynamic effects of Ezubao scandal shock. α_i is the loan fixed effect, and δ_t is the calendar time fixed effect. $X_{i,t}$ is the loan i's characteristics at time t. $Y_{i,t}$ are the market outcomes such as loan interest rate, and trading volume.

To begin with, we run this regression separately for loans successfully granted on Renrendai and Paipaidai. Paipaidai, another prominent P2P lending platform in China, operates exclusively online and does not offer guarantees for defaulted cases. Therefore, Paipaidai serves as the control group in this analysis. Results in Figure 18 show that the interest rates on Renrendai loans respond much earlier than those on Paipaidai loans to the negative shock resulting from the Ezubao scandal. The vertical line represents December 16, 2015, at the search peak of Ezubao scandal on the Baidu search engine. Renrendai loans began to have interest rate increases immediately following the first news report of the Ezubao scandal on December 3, 2015, about two weeks before the peak of news coverage and search. The loan prices on Renrendai also had a second surge after one week of the search peak. Whereas Paipaidai loans remained unaffected until December 22, 2015, a week after the news search peak. Meanwhile, the magnitudes of interest rate change were lower for Renrendai loans, whereas the changes in interest rate for Paipaidai were more volatile. In the event of another major platform scandal (Ezubao Scandal), investors may perceive a higher risk of not getting their money back, leading them to demand higher risk compensation, thereby resulting in observed higher interest rates. The results could be explained by the platform attracting particular types of investors through data hiding. The extremely low default rate on Renrendai likely attracts investors with fragile confidence, similar to the economic mechanism described by Gennaioli et al. (2012).

Renrendai secondary loan market Next, we study the responses of the Renrendai secondary market to the 2015 Ezubao scandal. The secondary market on the Renrendai platform started operations on October 14, 2013, allowing investors to transfer their acquired loans to other investors before reaching maturity (Bao et al., 2023). After holding loan shares for 90 days, investors have the option to sell their owned shares at either a discount or premium within the Renrendai platform's secondary market. Each share is assigned a principal value of 50 RMB. Interested investors can participate by bidding on the shares available for transfer.

Figure 15 illustrates the weekly trading volume in the secondary market on the Renrendai platform we study, showing a substantial surge in 2015 and reaching its peak during the summer of 2016. This surge suggests that investors engaged in a panic run from the platform in response to the Ezubao scandal. Similarly, Figure 16 depicts the weekly average share price in the secondary market, which also saw a dramatic drop in the summer of 2016. The big drop in share price reflected diminished investor confidence in the Renrendai platform during the industry's negative shock. In response to the Ezubao shock, investors pursue exits at much lower share prices.

We redo regression (12) for the Renrendai secondary market outcomes, separately analyzing the online and offline originated loans. Our analysis reveals that secondary market loan transfers are higher for online loans on Renrendai compared to offline loans after the Ezubao scandal. Thus though the whole Renrendai platform is fragile, offline loans are considered as safer on this same platform.

Moreover, we examine the impact of the Ezubao shock on liquidity and investor runs using secondary market data. Figures 20 and 19 demonstrate that following a major negative shock to the industry, investors on the platform rushed to exit. Since investors could only exit by selling shares in the secondary market, the platform faced liquidity risk when there were insufficient buyers. Figure 20 shows a dramatic increase in the number of sellers in the secondary market a quarter after the Ezubao shock. Similarly, Figure 19 reveals that investors had to wait significantly longer—about nine days longer—than usual to recover their funds after the shock.

The mass exodus of investors from the Renrendai platform during major negative market events, such as the Ezubao scandal, highlights the inherent fragility of the market. This fragility may stem from the platform's lack of transparency about actual default rates, which creates a false sense of security for investors and change investor behaviors. This opacity prevents investors from accurately assessing potential risks and making optimal decisions.

5.4. Investor Characteristics

To understand better, we delve deeper into how the establishment of offline branches, combined with the introduction of selective default rate disclosure, impacts the overall characteristics of investors on the Renrendai platform. To examinate investor behavior, we examine three key metrics aggregated at both the loan and weekly levels. This analysis sheds light on how the disclosure of loan default risks shapes market dynamics.

Investment Frequency: The average number of bids made by an investor per month, which reflects their level of activity on the Renrendai platform.

Average Monthly Investment Amount: The average monetary investment per month, serving as a proxy for investor confidence in the profitability of investing on Renrendai.

Proportion of Normal Bids (Mannual Investments): The percentage of funds invested manually without the use of Renrendai's automated finance plan bidding services. Automated bidding involves pre-specified parameters, such as maturity and promised annual interest rate, which direct investments into eligible loans automatically. In contrast, Normal Bids indicate greater investor confidence in their decision-making and a higher level of engagement in gathering information about the platform.

To investigate these effects, we use the same loan sample described in Section 5.1.4. We treat the establishment of each offline branch as event points. The treatment group comprises loans originated within the same city as an offline branch during the six-month window before and after its establishment. The control group includes loans originated within the same province but in different cities over the same period. For each investor, we aggregate all transactions to calculate average monthly investment frequency, average monthly investment amount, and the proportion of Normal Bids. These metrics are matched to the corresponding loans in the sample. As this analysis aims to understand how offline branches influence investor behavior, we focus exclusively on transactions made through Normal Bids, which reflect direct investment decisions. A difference-in-differences (DiD) analysis is conducted using the following regression model:

$$\begin{aligned} Y_{it} = \quad & \alpha + \beta \times BorrowerCharacteristics_{it} + \gamma \times LoanCharacteriscits_{it} \\ & +\theta_1 \times Distance_{ij} + \theta_2 \times RelativeMonth_{ij} \\ & +\delta_1 \times Treat_{ij} + \delta_2 \times Post_{ij} + \delta_3 \times Treat_{ij} \times Post_{ij} + \eta_{it} \end{aligned} \tag{13}$$

Y_{it} is the target investor characteristics to be investigated, including investment frequency, average monthly investment amount, and proportion of Normal Bids, aggregated at loan level. Individual investment frequency is the average times of investment made per month, calculated as the total number of transactions an investor made divided by the number of her active month on Renrendai. Individual average monthly investment amount is an investor's total monetary investment divided by the number of her active month on Renrendai. Individual proportion of Normal Bids is the percentage of money invested through Normal Bids in an investor's total monetary investment. Y_{it} is the value-weighted average at loan level for each of the three investor characteristics. $BorrowerCharacteristics_{it}$ include all of the borrower characteristics introduced in Section 5.1.1. $LoanCharacteristics_{it}$ include log of loan amount, loan maturity, and interest rate. Other variables are the same as defined in Section 5.1.4.

The regression results, reported in Table 11, reveal significant and negative coefficients for the interaction terms across all three columns. The results in Column (1) analyzes how the introductions of offline branches influence investors' investment frequency. The results indicate that after the establishment of offline branches, investors on the platform become less active, with a significant reduction in investment frequency. Column (2) shows the regression results on the log of average monthly investment amount. The coefficient for the interaction term suggests that average monthly investment amounts decrease following the establishment of offline branches. The decrease in average monthly investments reflects investors' conservative attitude toward the reliability of the disclosed loan information after the introduction of offline branches. The last column investigates the influence of introducing offline branches on the average proportion of Normal Bid made by the investors. The results show an increased reliance on automated bidding services post-establishment, indicating diminished confidence in personal investment decisions and reduced engagement in loan evaluation.

We then further investigate the characteristics of new investors attracted to Renrendai after the establishment of offline branches. Using the same sample as the DiD analysis but averaging investor characteristics at the weekly level, we compare new and existing investors in treatment and control groups. For example, in Figure 11, we plot investment frequency for existing investors versus newly attracted investors after the establishment of offline branches. New investors in the treatment group constitutes 1.2% of all investors in the sample. We use relative number of weeks to the establishing date instead of weeks on the real timeline. To make comparison, we also plot the investment frequency for newly attracted investors in the same province but in different cities to the corresponding offline branch.

Figure 11 shows that new investors demonstrate lower activity levels compared to existing investors. A decline in investment frequency is also observed among existing investors. Additionally, investment frequency of new investors in treatment versus control groups are very close, suggesting that reduced transparency in default data primarily impacts existing investors' investment frequency.

Figure 22 plots average monthly investment amount on the y-axis at the same setting as Figure 11. The plot shows that new investors contribute smaller monthly investments than existing investors, indicative of more conservative attitudes and reduced confidence in the platform's profitability. The decline in investment by existing investors further highlights the broader impact of reduced transparency. The monthly investment of new investors in treatment and control groups are close, which shows that the introduction of offline branches drag down the overall monthly investment mainly through affecting existing investors.

Figure 23 plots the proportion of Normal Bid in an investor's overall investment.New investors exhibit a higher proportion of Normal Bids, suggesting higher reliability on their own decision-making and more willingness to independently evaluate loans compared to existing investors. Comparing the new investors in treatment and control group, the results imply that the decrease in default data transparency also induced new investors to rely more on their own decision. However, new investors in the treatment group constitute only 1.2% of the sample. The downward trending plot for existing investors show that the decrease in default data transparency leads old investors to rely more on automated bidding system, because they are less confident in their own judgment on the borrowers' credibility.

The results in this section imply that the establishment of offline branches, which is linked to a decrease in default data transparency, reduces existing investors' attentiveness while attracting new, less-informed investors with fragile confidence.

6. Conclusion

To sum up, this paper examines the impact of data manipulation on the financial stability and market efficiency of Fintech lending platforms, specifically focusing on Renrendai, a prominent P2P lending platform in China. Our findings reveal that Renrendai underreported its default rates by a significant margin, particularly after its partnership with Ucredit, an offline lending service provider. This created an illusion of a risk-free market that misled investors.

We further explore the economic consequences of this data manipulation on market efficiency and stability. Our analysis shows that the lack of transparency distorted market pricing and contributed to inefficiencies, with data-hiding practices exacerbating the problem. Moreover, our findings suggest that this opacity makes the market more susceptible to liquidity shocks, as evidenced by investor runs during the 2015 Ezubao scandal. After the introduction of offline partnerships, new investors became more conservative, preferring manual bids and smaller investments, while existing investors showed decreasing engagement. This change in investor behavior underscores the fragile nature of the market when key risk information is hidden.

Our research emphasizes the broader implications of information transparency in the Fintech lending sector, particularly for platforms that operate without sufficient regulatory oversight. We argue that a clearer understanding of default data is essential for maintaining market stability and trust. By shedding light on the consequences of data manipulation, we contribute to the ongoing discourse on the need for regulatory frameworks to ensure transparency and mitigate potential risks to the financial system in the rapidly evolving digital finance landscape.

References

Albrecht, C., Morales, V., Baldwin, J. K., Scott, S. D., 2017. Ezubao: a chinese ponzi scheme with a twist. Journal of Financial Crime 24, 256–259.

Avoyan, A., Khubulashvili, R., Mekerishvili, G., 2024. Optimal disclosure on crowdfunding platforms. Available at SSRN .

Bao, T., Ding, Y., Gopal, R., Möhlmann, M., 2023. Throwing good money after bad: Risk mitigation strategies in the p2p lending platforms. Information Systems Frontiers pp. 1–21.

Begley, T. A., Purnanandam, A., Zheng, K., 2017. The strategic underreporting of bank risk. The Review of Financial Studies 30, 3376–3415.

Bimpikis, K., Mantegazza, G., 2023. Strategic release of information in platforms: Entry, competition, and welfare. Operations Research 71, 1619–1635.

Carletti, E., Leonello, A., Marquez, R., 2023. Loan guarantees, bank underwriting policies and financial stability. Journal of Financial Economics 149, 260–295.

Chen, Q., Goldstein, I., Huang, Z., Vashishtha, R., 2020. Liquidity transformation and fragility in the us banking sector. Tech. rep., National Bureau of Economic Research.

Cheng, W. G., de Oliveira Leite, R., Caldieraro, F., 2022. Financial contagion in internet lending platforms: Who pays the price? Finance Research Letters 45, 102187.

Cookson, J. A., Fox, C., Gil-Bazo, J., Imbet, J. F., Schiller, C., 2023. Social media as a bank run catalyst. Available at SSRN 4422754 .

Di Porto, F., Zuppetta, M., 2021. Co-regulating algorithmic disclosure for digital platforms. Policy and society 40, 272–293.

Diamond, D. W., Dybvig, P. H., 1983. Bank runs, deposit insurance, and liquidity. Journal of political economy 91, 401–419.

Duarte, J., Siegel, S., Young, L., 2012. Trust and credit: The role of appearance in peer-to-peer lending. The Review of Financial Studies 25, 2455–2484.

Franks, J., Serrano-Velarde, N., Sussman, O., 2021. Marketplace lending, information aggregation, and liquidity. The Review of Financial Studies 34, 2318–2361.

Gennaioli, N., Shleifer, A., Vishny, R., 2012. Neglected risks, financial innovation, and financial fragility. Journal of financial economics 104, 452–468.

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

Gillett, P. R., Uddin, N., 2005. Cfo intentions of fraudulent financial reporting. Auditing: A Journal of Practice & Theory 24, 55–75.

Iyer, R., Khwaja, A. I., Luttmer, E. F., Shue, K., 2016. Screening peers softly: Inferring the quality of small borrowers. Management Science 62, 1554–1577.

Iyer, R., Puri, M., 2012. Understanding bank runs: The importance of depositor-bank relationships and networks. American Economic Review 102, 1414–1445.

Jin, D., Kacperczyk, M., Kahraman, B., Suntheim, F., 2022. Swing pricing and fragility in open-end mutual funds. The Review of Financial Studies 35, 1–50.

Kaminski, K. A., Wetzel, T. S., Guan, L., 2004. Can financial ratios detect fraudulent financial reporting? Managerial Auditing Journal .

Kirkos, E., Spathis, C., Manolopoulos, Y., 2007. Data mining techniques for the detection of fraudulent financial statements. Expert systems with applications 32, 995–1003.

Liao, L., Wang, Z., Xiang, J., Yan, H., Yang, J., 2021. User interface and firsthand experience in retail investing. The Review of Financial Studies 34, 4486–4523.

Murfin, J., 2012. The supply-side determinants of loan contract strictness. The Journal of Finance 67, 1565–1601.

Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., Sun, X., 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision support systems 50, 559–569.

Rayo, L., Segal, I., 2010. Optimal information disclosure. Journal of political Economy 118, 949–987.

Schmidt, L., Timmermann, A., Wermers, R., 2016. Runs on money market mutual funds. American Economic Review 106, 2625–2657.

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

Vallee, B., Zeng, Y., 2019. Marketplace lending: A new banking paradigm? The Review of Financial Studies 32, 1939–1982.

Wang, Q., Su, Z., Chen, X., 2021. Information disclosure and the default risk of online peer-to-peer lending platform. Finance Research Letters 38, 101509.

Wang, Y., Xia, H., 2014. Do lenders still monitor when they can securitize loans? The Review of Financial Studies 27, 2354–2391.

Figures and Tables

Figures

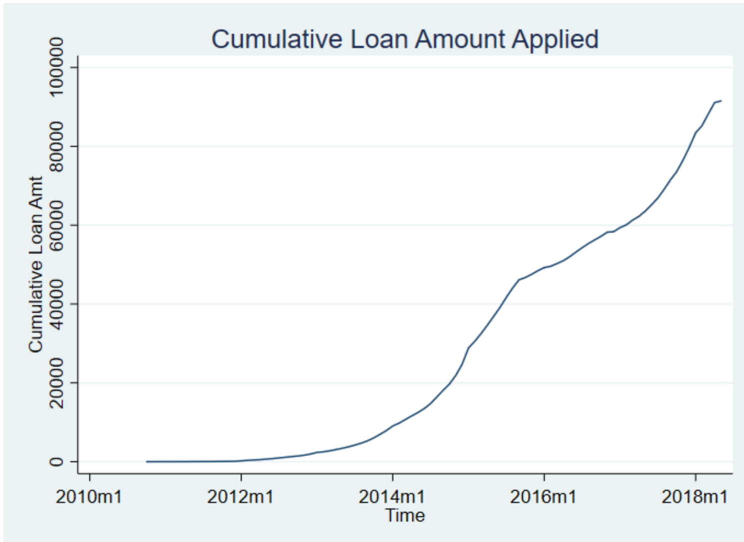

Fig. 1. Cumulative Loan Amount

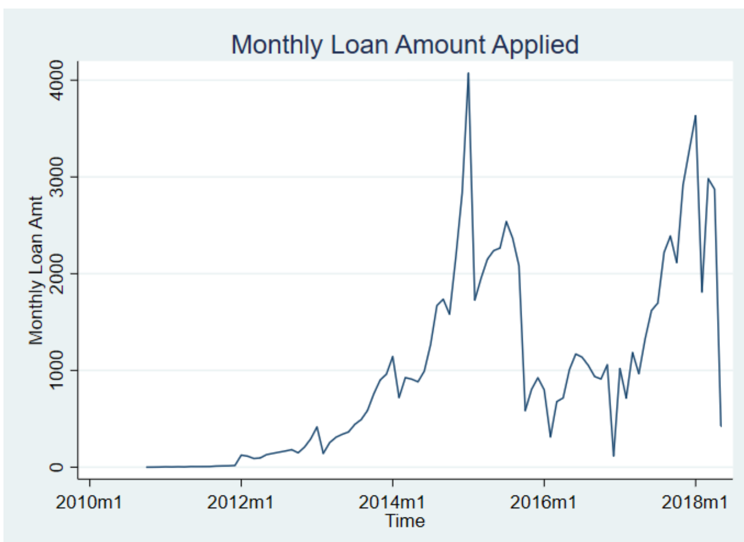

Fig. 2. Monthly Loan Amount

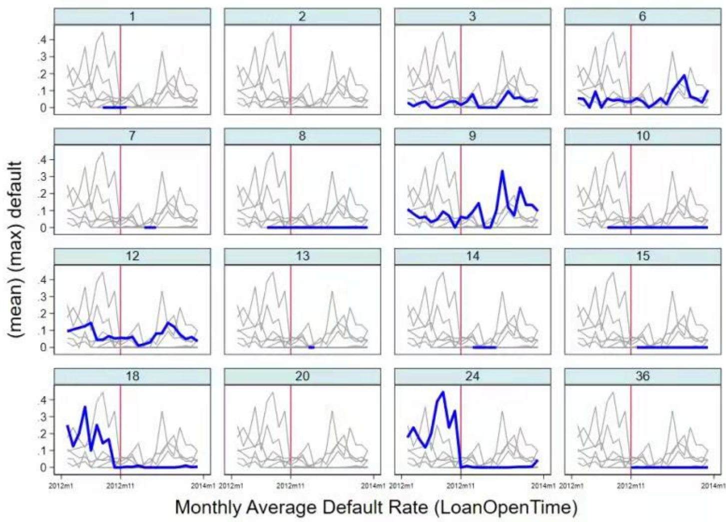

Fig. 3. Monthly Default Rates for Loans with Different Maturities

The title for each subplot is the maturity of the loan type measured in months. For example, "18" in the subtitle means the loan maturity is 18 months. The blue line plots the loan default rates with maturity specified in the subtitle. The grey lines plot default rates with the other maturities.

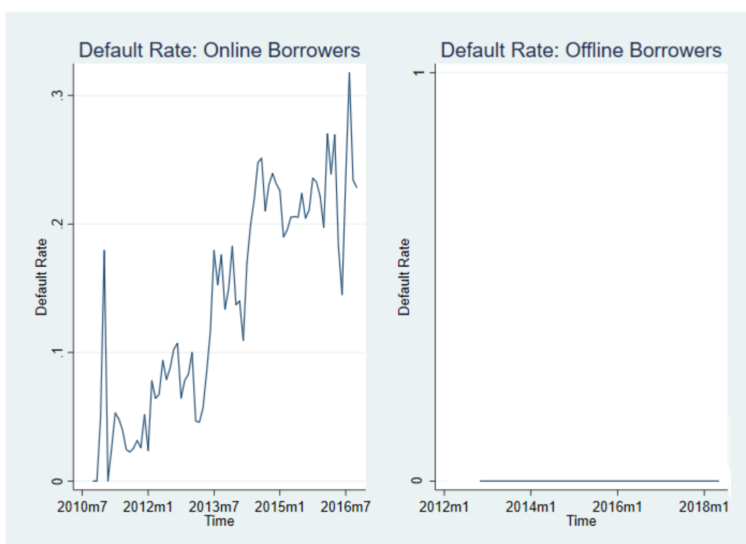

Fig. 4. Monthly Default Rates

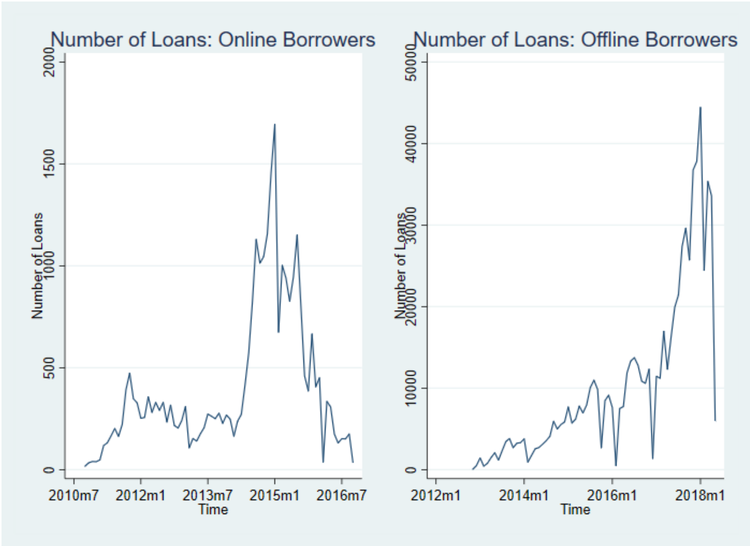

Fig. 5. Monthly Number of Loans

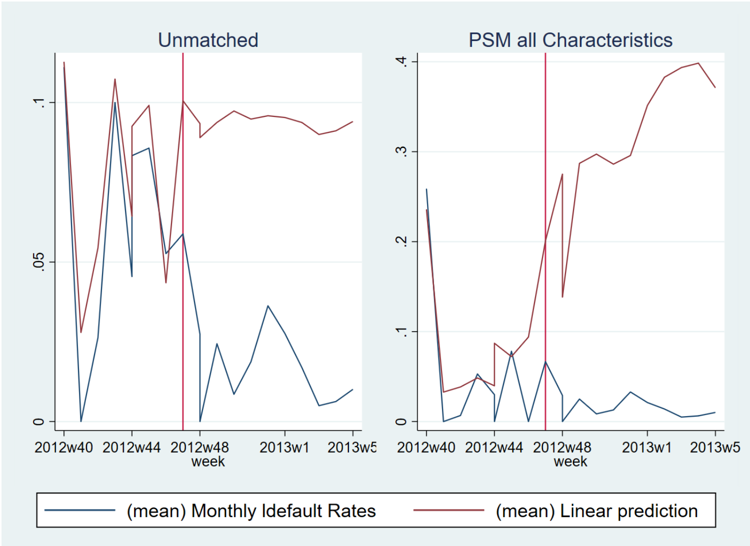

Fig. 6. Predicted v.s Published Default Rates, 2-month Window

This figure reports the default rates predicted using 2-month window regression results v.s. published default rates. The red vertical line indicates November 2012, the event time.

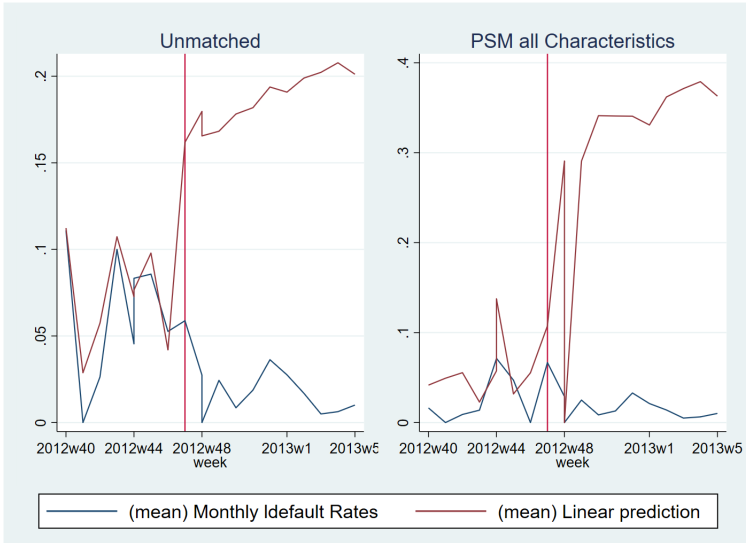

Fig. 7. Predicted v.s Published Default Rates, 6-month Window

This figure reports the default rates predicted using 6-month window regression results v.s. published default rates. The red vertical line indicates November 2012, the event time.

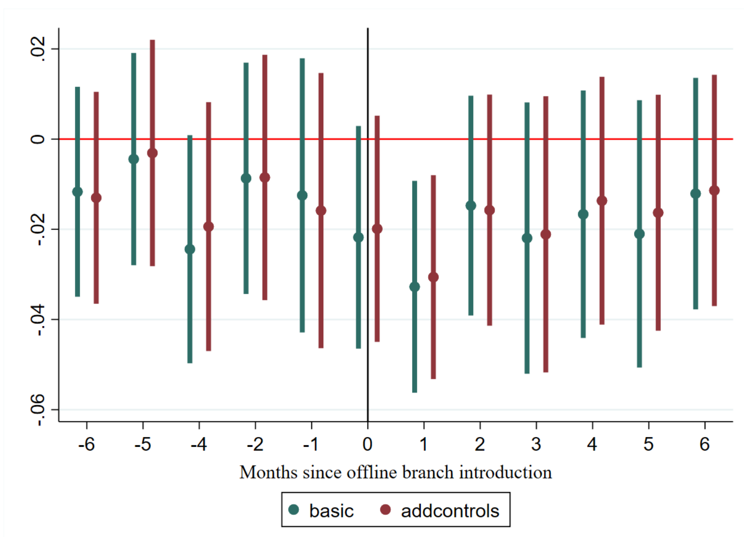

Fig. 8. Market Efficiency Test: Offline Branches' Establishment

Figure 8 plots the coefficients of each $FE\_RelativeMonth \times Treat \times r_i^{\star}$ level in the regression model (10) and the corresponding confidence intervals at the level of 1%.

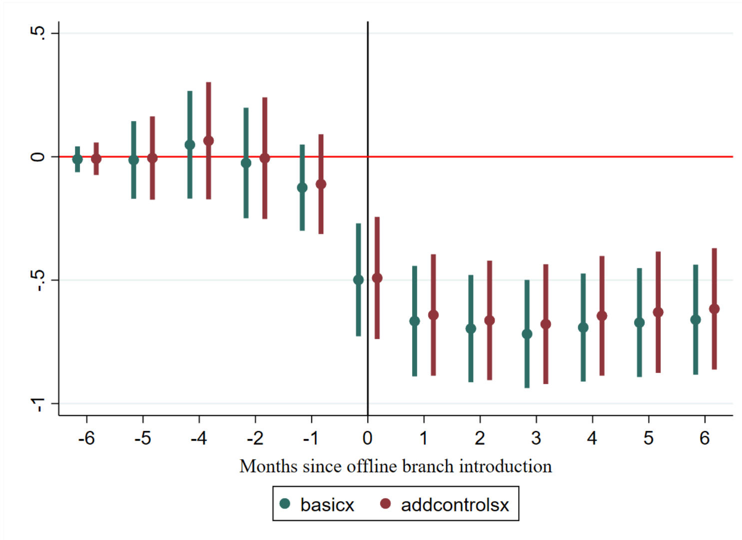

Fig. 9. "Credit" Type Loan Fraction: Offline Branches' Establishment

Figure 9 plots the coefficients of each $FE\_RelativeMonth \times Treat$ level in the regression model (11) where the outcome variable is the fraction of "Credit" Type Loans and the corresponding confidence intervals at the level of 1%.

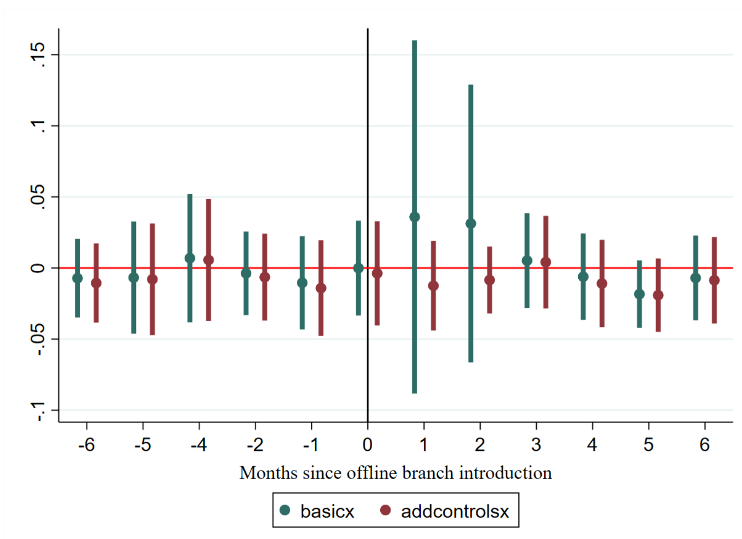

Fig. 10. The Number of "Credit" Type Loan: Offline Branches' Establishment

Figure 10 plots the coefficients of each $FE\_RelativeMonth \times Treat$ level in the regression model (11) where the outcome variable is the number of "Credit" Type Loans and the corresponding confidence intervals at the level of 1%.

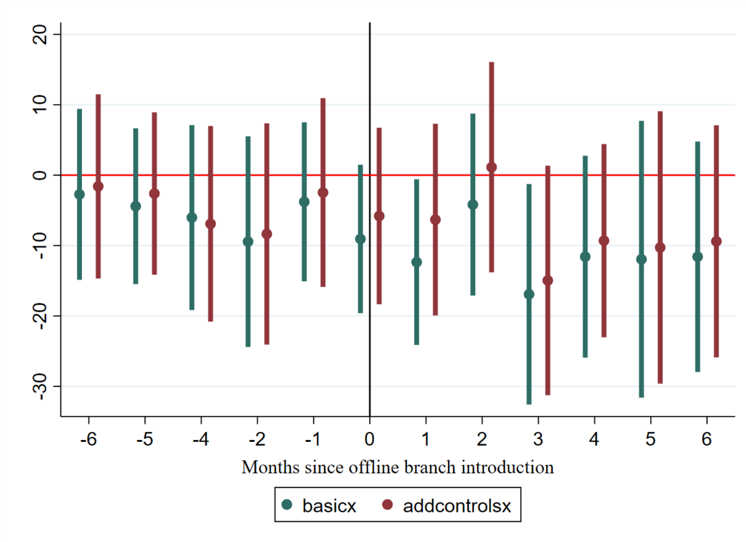

Fig. 11. Latest Credit Score: Offline Branches' Establishment

Figure 11 plots the coefficients of each $FE\_RelativeMonth \times Treat$ level in the regression model (11) where the outcome variable is the latest credit score of the borrower at the performance time September 2018 and the corresponding confidence intervals at the level of 1%.

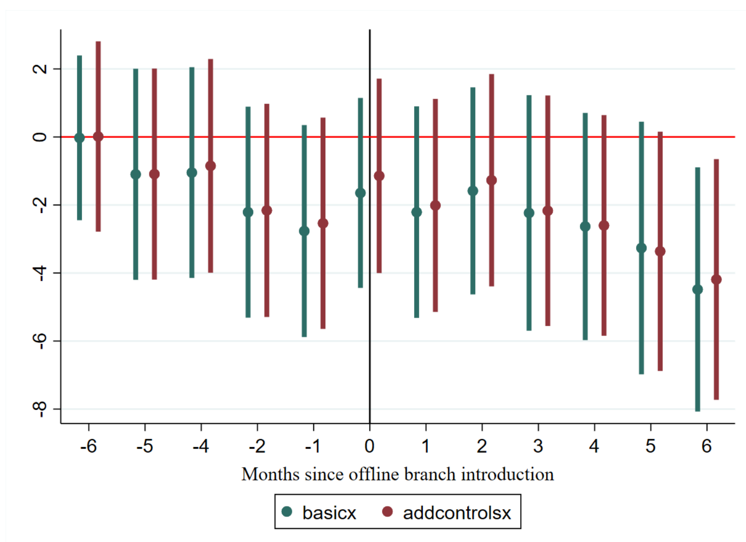

Fig. 12. Latest Credit Score: Offline Branches' Establishment, all applications

Figure 12 plots the coefficients of each $FE\_RelativeMonth \times Treat$ level in the regression model (11) where the outcome variable is the latest credit score of the borrower at the performance time of September 2018, and the sample includes all applications. The corresponding confidence intervals are at the level of 95%.

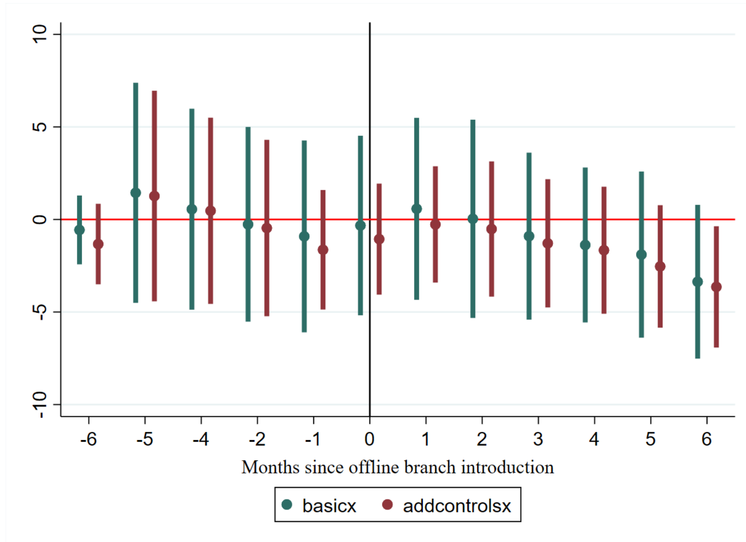

Fig. 13. Finishing Ratio: Offline Branches' Establishment, all applications

Figure 13 plots the coefficients of each $FE\_RelativeMonth \times Treat$ level in the regression model (11) where the outcome variable is the finishing ratio of loan applications and the sample includes all applications. The corresponding confidence intervals are at the level of 95%.

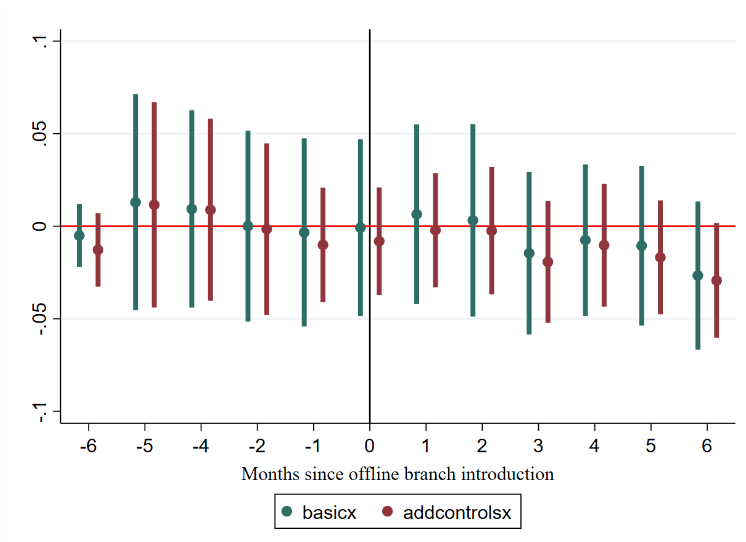

Fig. 14. Success Rate: Offline Branches' Establishment, all applications

Figure 13 plots the coefficients of each $FE\_RelativeMonth \times Treat$ level in the regression model (11) where the outcome variable is the success rate of loan applications, and the sample includes all applications. The corresponding confidence intervals are at the level of 95%.

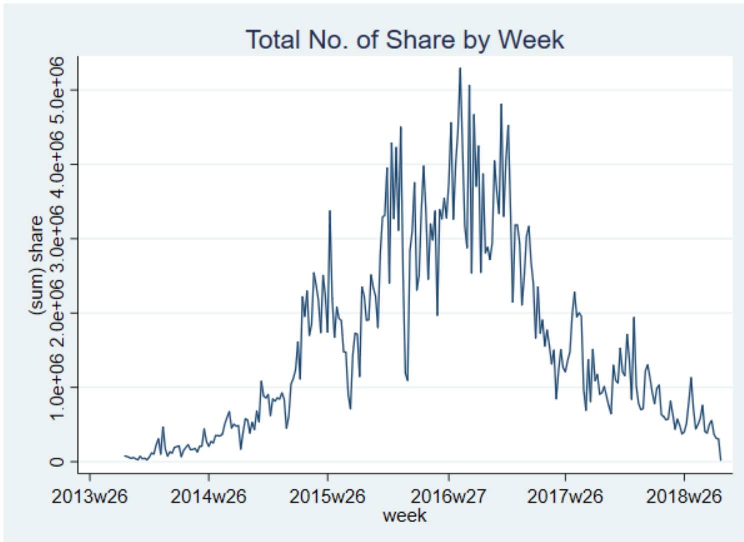

Fig. 15. Secondary Market Loan Trading Volume, by week

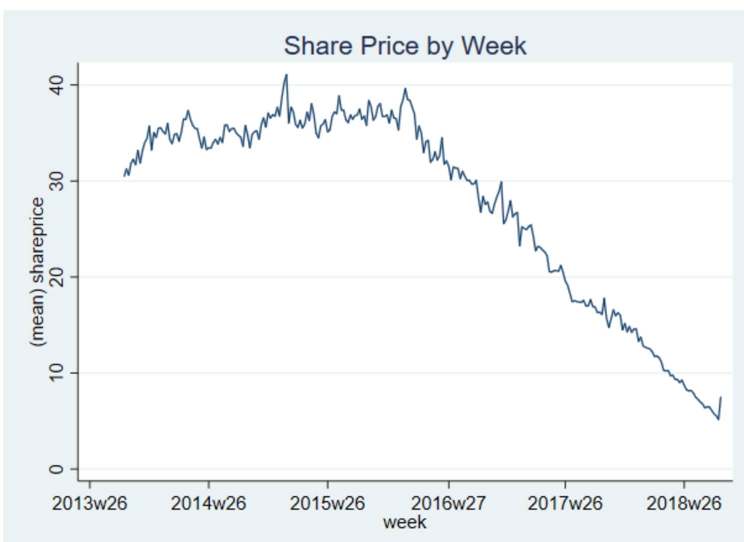

Fig. 16. Secondary Market Loan Share Price, by week

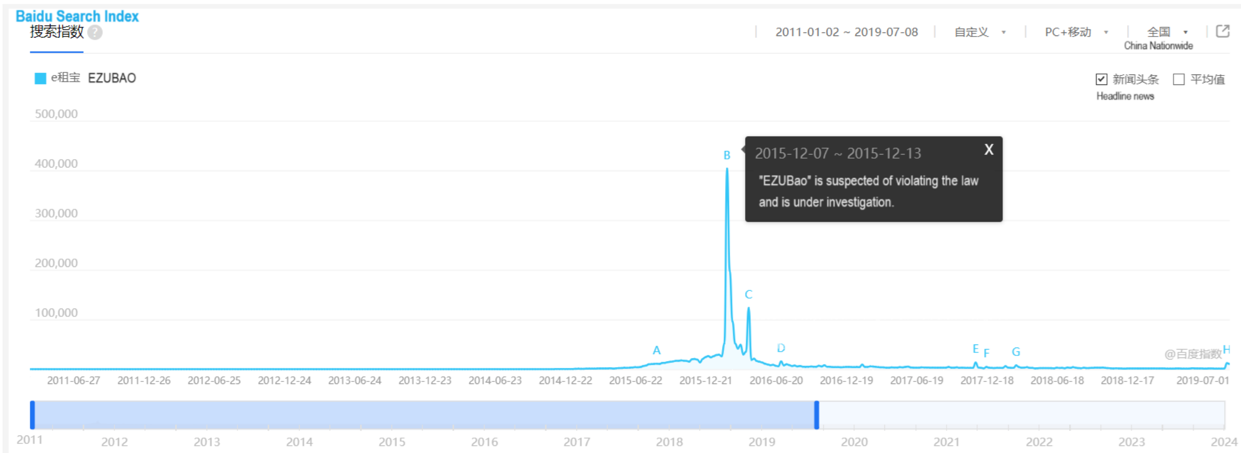

Fig. 17. Baidu Search Index “Ezubao”

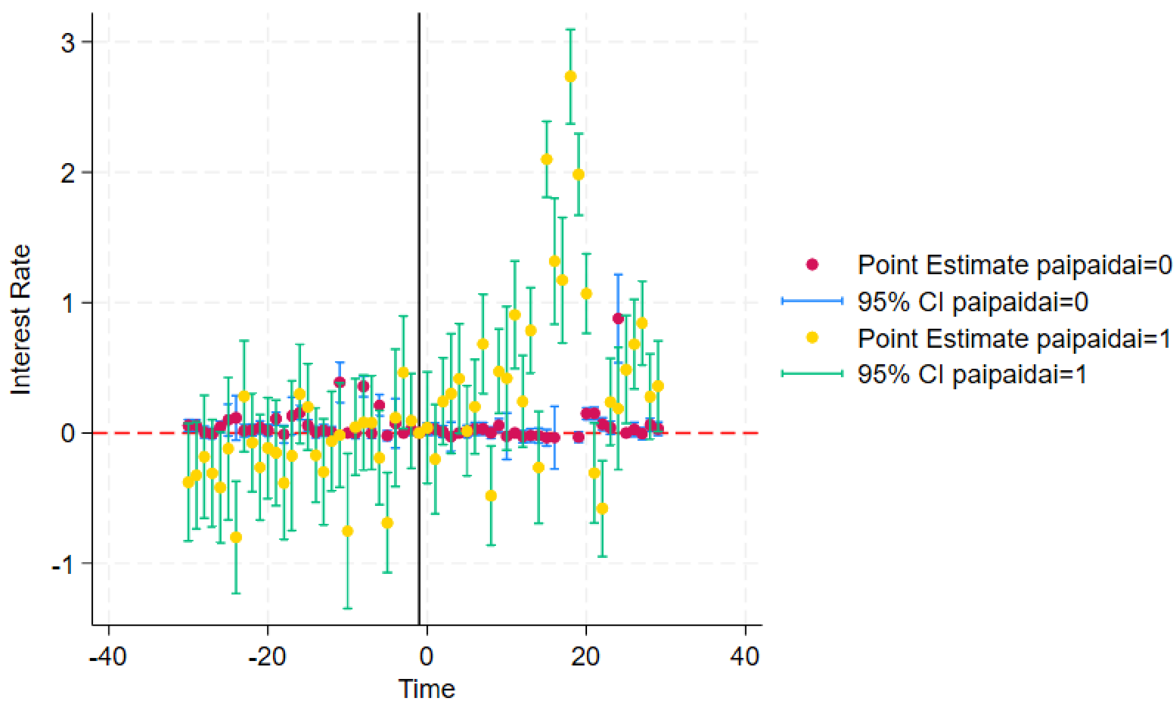

Fig. 18. Event Study “Ezubao”, Paipaidai vs Renrendai

This figure plots the coefficients for the event study of the Ezubao scandal. The red dots and blue intervals represent Renrendai (paipaidai=0), while the yellow dots and green intervals represent Paipaidai (paipaidai=1). Time 0 represents December 16, 2015, the attention peak of the Ezubao scandal news when Baidu search activity for Ezubao reached its highest point. The first news reports of the Ezubao scandal emerged on December 3, 2015. The corresponding confidence intervals are at the level of 95%.

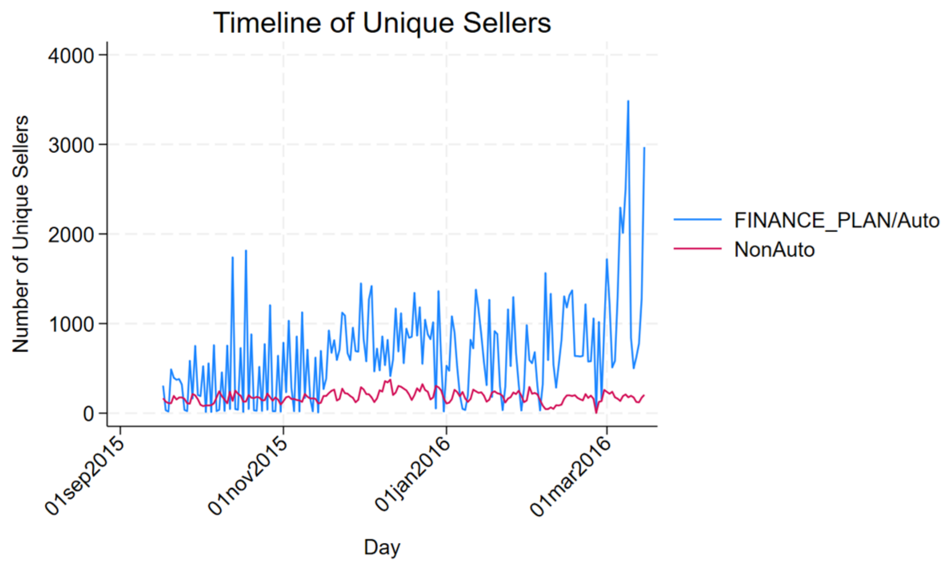

Fig. 19. Event Study "Ezubao", Number of Sellers in Renrendai Secondary Market

Timeline of the number of sellers in the Renrendai secondary market before and after the Ezubao scandal (2015 December).

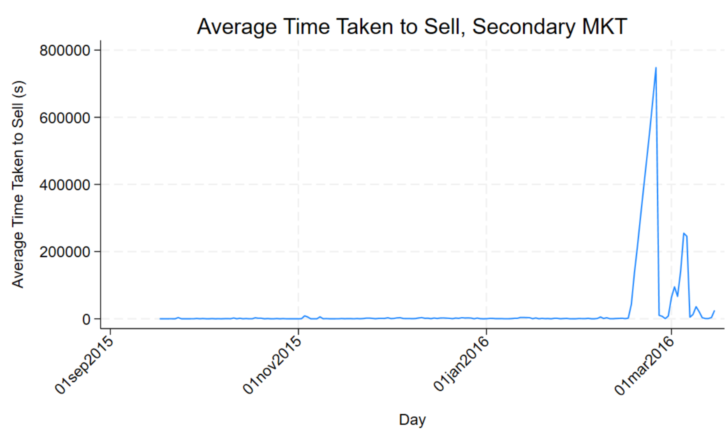

Fig. 20. Event Study "Ezubao", Time Taken to Sell in Renrendai Secondary Market

Timeline of time taken to sell (in seconds) in the Renrendai secondary market before and after the Ezubao scandal (2015 December).

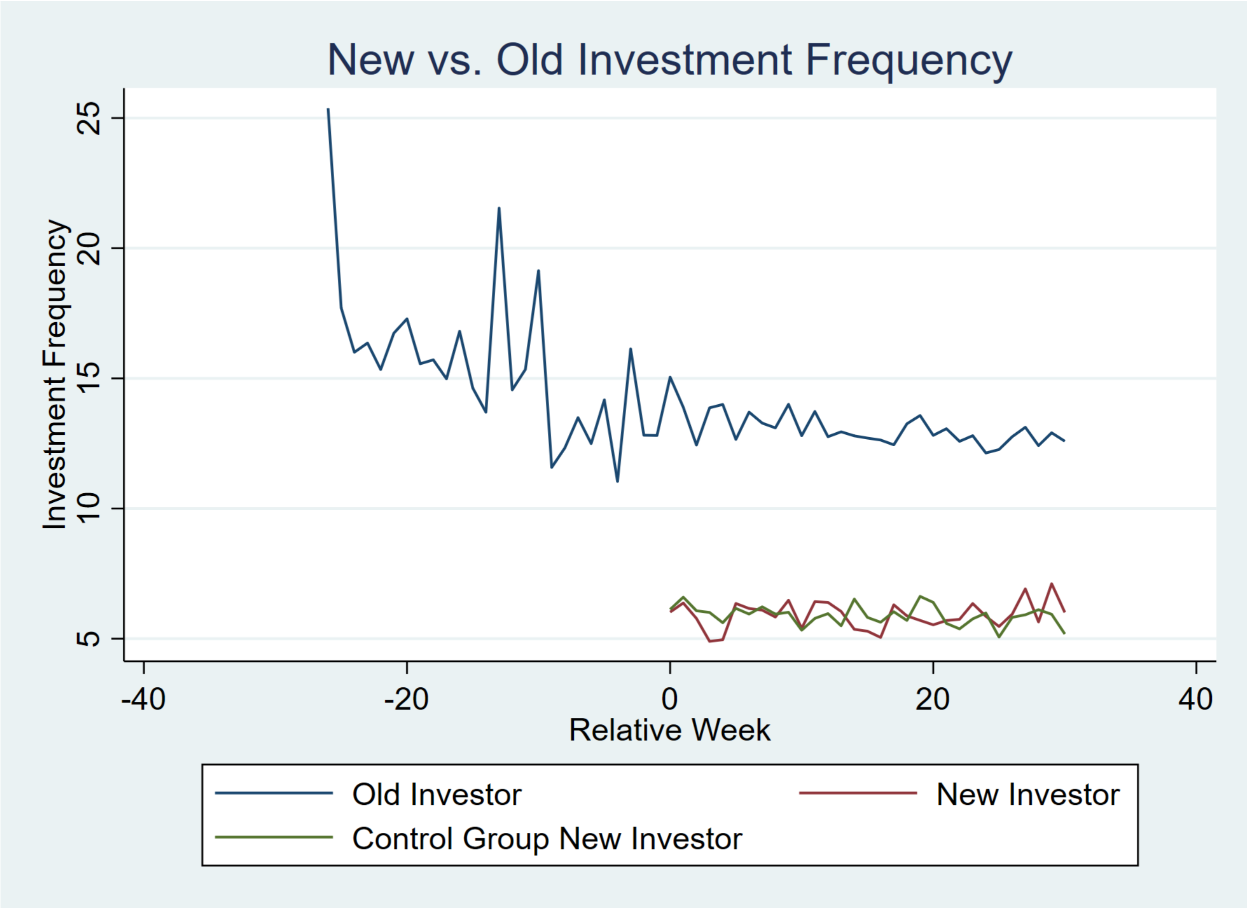

Fig. 21. Offline Branch: Investment Frequency

This figure plots the investment monthly frequency of existing investors on Renrendai platform before the introduction of offline branches versus investment frequency of new investors in the same city of the offline branches who are attracted after the offline branches are established. We also plotted the investment frequency of new investors in the control group, and those investors are in the same province as the offline branch but in different cities.

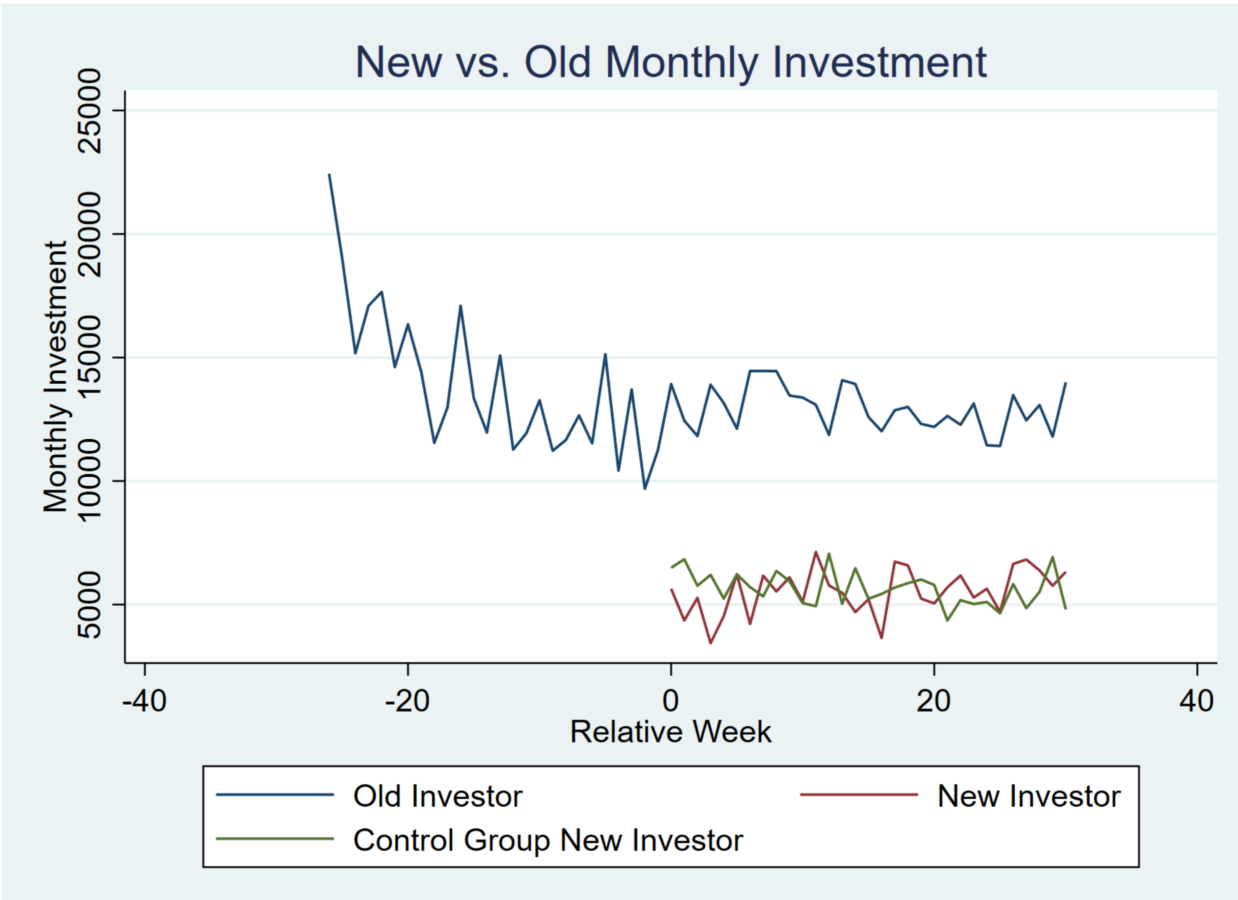

Fig. 22. Offline Branch: Monthly Investment

This figure plots the average monthly investment amount of existing investors on Renrendai platform before the introduction of offline branches versus the average monthly investment amount of new investors in the same city of the offline branches who are attracted after the offline branches are established. We also plotted the average monthly investment amount of new investors in the control group, and those investors are in the same province as the offline branch but in different cities.

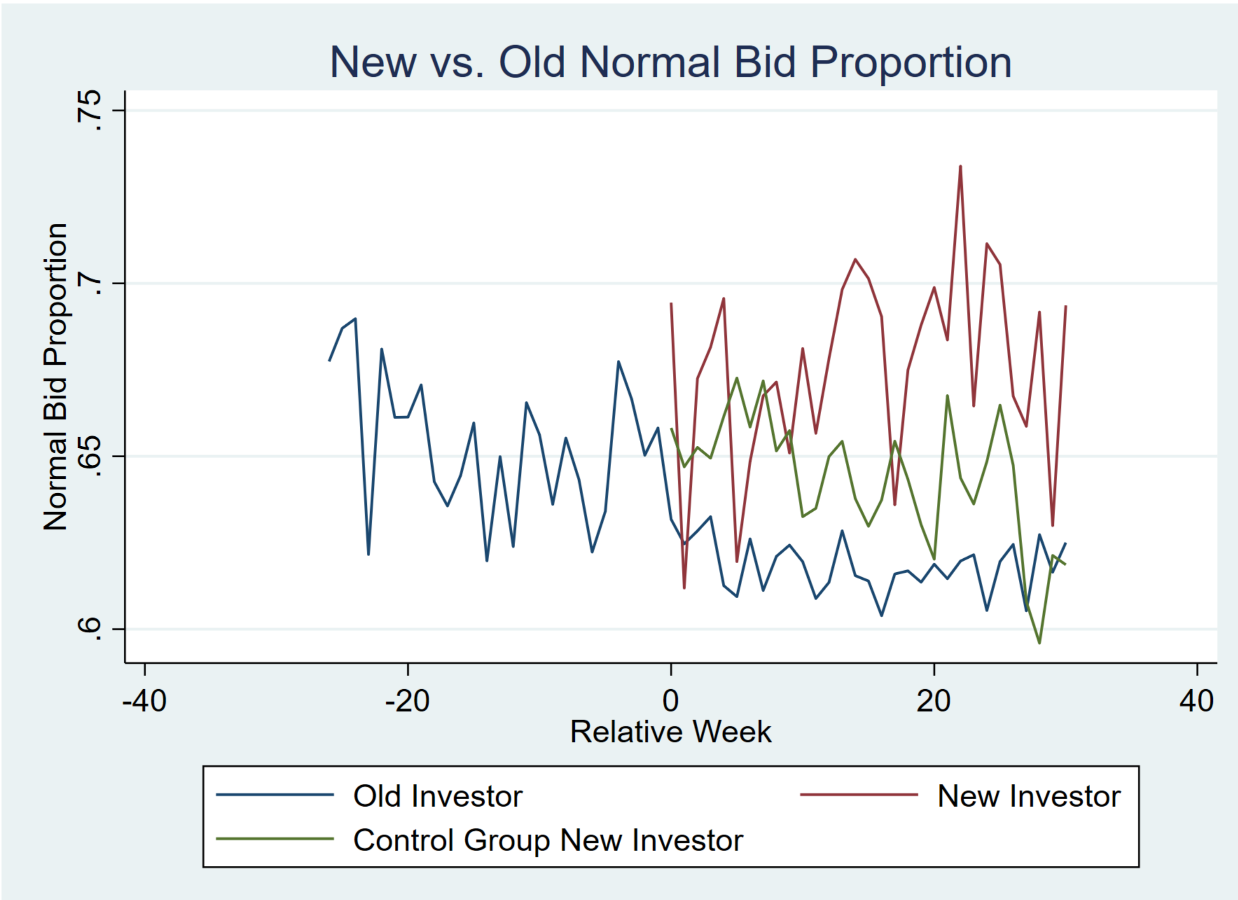

Fig. 23. Offline Branch: Normal Bid Proportion

This figure plots the proportion of Normal Bid in overall investment for existing investors before the introduction of offline branches versus the proportion of Normal Bid for new investors in the same city of the offline branches who are attracted after the offline branches are established. We also plotted the proportion of Normal Bid for new investors in the control group, and those investors are in the same province as the offline branch but in different cities.

Tables

Table 1: Summary Statistics

	Mean	SD	Min	Max	N
Loan Characteristics					
Total loan applied per month (million RMB)	937.81	803.77	74.90	3,959.78	60
Total loan applied monthly growth (%)	9.65	30.69	-72.22	115.02	59
Total loan granted per month (million RMB)	353.49	352.42	1.57	1,258.16	60
Total loan granted monthly growth (%)	49.82	203.11	-94.11	1,465.56	59
Loan size (granted)(thousand RMB)	71	43	3.00	500.00	300626
Maturity (months)	31	9	3.00	48.00	300626
Annual interest rate (%)	11	1	6.60	24.40	300626
Borrower Characteristics					
Company Size	1	1	1.00	4.00	298667
Marital Status (0/1)	0.69	0.46	0.00	1.00	300626
Gender (0/1)	1.30	0.46	1.00	2.00	277960
Education	2.15	0.73	1.00	4.00	300619
Work Experience	2.08	1.23	1.00	4.00	299951
Age	38.19	7.71	23.00	55.00	300626
Property (0/1)	0.55	0.50	0.00	1.00	300626
House Mortgage (0/1)	0.34	0.48	0.00	1.00	300626
Income	4.36	1.20	1.00	7.00	300624
Car (0/1)	0.25	0.43	0.00	1.00	300626
Car Mortgage (0/1)	0.06	0.25	0.00	1.00	300626

This table reports the summary statistics on the loans applied between January 2012 to December 2016. Panel A reports the summary of loan characteristics. *Loan size* applied and granted is reported in thousand RMB. *Total loan* applied and granted per month is measured in million RMB. *Maturity* of loans is measured in months. *Default rate* is measured at loan opening time. Late and bad repayments are all viewed as defaults.

Panel B shows the summary statistics for borrower characteristics. *Company size* is a categorical variable describing the number of total employers in the borrower's working company. (*Company size* = 1: below 10 employers; 2: 10-100 employers; 3: 100-500 employers; 4: above 500 employers.) *Marital status* is a dummy variable equaling 0 for not married and 1 for married. *Education* is a categorical variable describing the borrower's level of education. (*Education* = 1: junior college; 2: high school; 3: undergraduate; 4: graduate or above.) *Work experience* is a categorical variable describing the number of years the borrower has worked. (*Work experience* = 1: 1 year and below; 2: 1-3 years; 3: 3-5 years; 4: 5 years and above.) *Property* is a dummy variable equaling one if the borrower owns at least one property and 0 if the borrower has no property. *Housing mortgage* is a dummy variable equaling one if the borrower has at least one housing mortgage outstanding and 0 otherwise. *Income* is a categorical variable describing the borrower's monthly income. (*Income* = 1: below 1000 RMB; 2: 1001-2000 RMB; 3: 2001-5000 RMB; 4: 5001-10000 RMB; 5: 10001-20000 RMB; 6: 20001-50000 RMB; 7: above 50000 RMB.) *Car* is a dummy variable equaling one if the borrower has a car and 0 if the borrower has no car. *Car mortgage* is a dummy variable equaling one if the borrower has at least one car mortgage outstanding and 0 otherwise. *Credit limit* describes the borrower's credit limit on the Renrendai lending platform.

Table 2: Regression Results: 2-month window

Regression Results Before and After the Event

VARIABLES	(1) Before	(2) After	(3) PSM
Company Size	-0.017	0.003	0.025
	(0.025)	(0.004)	(0.029)
Marriage	-0.012	0.001	-0.074
	(0.057)	(0.005)	(0.069)
Education	-0.004	-0.006*	0.045
	(0.009)	(0.003)	(0.029)
Work Experience	0.008	-0.003	0.098**
	(0.019)	(0.003)	(0.035)
Age	0.001	-0.000	0.006
	(0.002)	(0.000)	(0.004)
Gender	0.014	-0.004	0.092
	(0.023)	(0.003)	(0.061)
Property	0.060*	-0.003	-0.005
	(0.029)	(0.025)	(0.028)
House Mortgage	0.004	0.025	0.154
	(0.034)	(0.033)	(0.100)
Income	0.018	0.005*	-0.021
	(0.012)	(0.003)	(0.018)
Car	-0.034	-0.012	-0.036
	(0.031)	(0.008)	(0.051)
Car Mortgage	-0.063	-0.006	-0.053
	(0.046)	(0.041)	(0.061)
Log Loan Amount	-0.029*	-0.007	-0.032
	(0.015)	(0.005)	(0.026)
Loan Maturity	0.004	-0.002	0.025*
	(0.006)	(0.001)	(0.011)
Interest Rate	0.020**	0.023***	0.014**
	(0.008)	(0.007)	(0.004)
Constant	-0.094	-0.176*	-0.588
	(0.188)	(0.084)	(0.416)
Difference Test	0.078***	N/A	0.307***
	(0.003)	N/A	(0.003)
Logit Difference Test	0.079***	N/A	0.137***
	(0.003)	N/A	(0.024)
Lasso Difference Test	0.073***	N/A	0.136***
	(0.003)	N/A	(0.007)
Week FE	Yes	Yes	No
Province FE	Yes	Yes	Yes
Observations	195	2,197	2,196
R-squared	0.293	0.169	0.585

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

This table reports the regression results for the specification in Equation (1). Column (1) reports the loan regression results in the 2-month pre-window (October 2012 to November 2012). Column (2) reports the regression results on the loans in the 2-month post window (December 2012 to January 2013). Column (3) reports the regression results on the propensity score matched sample. Observations are matched on the 10 borrower characteristics included in Equation (1). *Difference Test* reports the mean difference between the predicted default rates and the published default rates in the post window (December 2012 to January 2013). Instead of using OLS, the *Logit Difference Test* uses the predicted default rates from logit regressions of default rates on the borrower and loan characteristics. Standard errors are clustered at the week level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3: Online v.s. Offline Borrowers

	2-month		6-month		1-year	
	Base	PSM	Base	PSM	Base	PSM
Time & Province FE	0.275***	0.524***	0.216***	0.196***	0.218***	0.254***
	(0.005)	(0.013)	(0.010)	(0.044)	(0.006)	(0.009)
Time FE	0.263***	0.453***	0.198***	0.159***	0.212***	0.150***
	(0.005)	(0.022)	(0.009)	(0.037)	(0.005)	(0.004)
Province FE	0.280***	0.389***	0.222***	0.186***	0.220***	0.251***
	(0.005)	(0.021)	(0.011)	(0.041)	(0.006)	(0.010)

The table reports the mean difference between the reported and predicted default rates using different time windows and controlling different fixed effects. the mean difference between the predicted default rates and the published default rates. The Base columns predict default rates for "field" type borrowers from the regression results on online borrowers. The PSM columns predict "field" type default rates from the regression results on the propensity score matched sample. The propensity score matched samples are created by matching "field" type borrowers to online borrowers in the same time windows. We also control for different fixed effects to furthur check the robustness of the results. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: Regression for Offline Branches: 2-month Window

	Offline Establishment	
	(1)	(2)
VARIABLES	DID	Pre-Window
Treat	0.065**	
	(0.028)	
Post	0.008***	
	(0.002)	
Relative Month	-0.001*	-0.001
	(0.001)	(0.002)
Treat * Post	-0.072***	
	(0.026)	
Company Size	0.022***	0.019***
	(0.005)	(0.005)
Marriage	-0.002*	0.000
	(0.001)	(0.002)
Education	-0.007**	-0.006**
	(0.003)	(0.002)
Work Experience	0.027***	0.024***
	(0.005)	(0.006)
Age	-0.000***	-0.000***
	(0.000)	(0.000)
Gender	-0.000	-0.001
	(0.001)	(0.002)
Property	0.006	-0.003
	(0.004)	(0.003)
Housing Mortgage	-0.010**	-0.003
	(0.004)	(0.003)
Income	0.005***	0.004***
	(0.001)	(0.001)
Car	-0.001	-0.001
	(0.001)	(0.002)
Car Mortgage	-0.002	0.000
	(0.002)	(0.002)
Log Loan Amount	-0.030***	-0.026***
	(0.004)	(0.004)
Loan Maturity	0.001**	0.001**
	(0.000)	(0.000)
Log Distance	0.000	-0.005*
	(0.001)	(0.003)
Constant	0.258***	0.288***
	(0.042)	(0.055)
Loan Time FE	Yes	Yes
Branch Open Time FE	Yes	Yes
Borrower Living City FE	Yes	Yes
Observations	74,567	23,935
R-squared	0.165	0.191

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Column (1) reports the difference-in-difference regression results based on Equation (2). Column (2) reports the regression results on all observations in the pre-window. Standard errors are clustered at month level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: Default Rate

	Default prob. , Online Matched	Default prob. , All
	default	default
CreditRating FEs		
AA/A	-0.0113***	-0.0120***
	(0.00225)	(0.00169)
B	-0.0120	-0.00563
	(0.00957)	(0.00857)
C	-0.00294	-0.00508
	(0.00794)	(0.00485)
D	-0.00622	-0.00203
	(0.00410)	(0.00411)
E	-0.00307	-0.00315
	(0.00412)	(0.00270)
Loan Life-cycle FEs		
Early	0.00535	0.0130***
	(0.00342)	(0.00268)
Mid	0.00817***	0.0132***
	(0.00302)	(0.00267)
Late	0.00888**	0.0131***
	(0.00349)	(0.00269)
Year-Month FE	Yes	Yes
N	9625	205278
R^2	0.0150	0.0132

This table reports the estimates of the default equation (7) for 276,418 monthly repayment performance of propensity score-matched loans originated on Renrendai between 2012 June and 2013 May. The repayment performance observation ended in September 2018. The first column is the result of a pure online matched sample, including manipulated offline loans and pure online loans. The second column is the result of all loans. The dependent variable, default dummy, equals one if loan i defaults in the performance month t. The regressions include credit rating fixed effects, and the three-stage loan's life cycle fixed effects. Standard errors are heteroscedasticity-robust, and clustered by borrower user id. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Recovery Rate Regression

	RecoveryRate, Post online	RecoveryRate, Post All
	(1)	(2)
	rrecov	rrecov
CreditRating FEs		
B	0.123	0.137
	(0.105)	(0.0964)
C	0.494^{**}	0.518^{***}
	(0.229)	(0.194)
D	0.511^{**}	0.573^{***}
	(0.250)	(0.140)
E	0.148	0.231^{**}
	(0.117)	(0.108)
HR	0.140	0.210^{**}
	(0.110)	(0.0850)
AA/A		1.044^{***}
		(0.152)
Year-Month FE	Yes	Yes
City FE	Yes	Yes
N	111	147
R^2	0.416	0.406

This table reports the estimates of the recovery equation (8) using all 1,135 default events for loans originated on Renrendai between 2012 June and 2013 May. The dependent variable equals the fraction of post-default recoveries to the balance remaining at the point of default. The regressions include credit rating fixed effects, and loan issue time fixed effects. Standard errors are heteroscedasticity-robust, and clustered by borrower user id. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Default Rate, Recovery Rate and Loss Given Default (LGD)

	Default prob., Online Matched	Default prob., All
	(1)	(2)
	default	default
AA/A	0	0.0136
B	0	0.0898
C	0.0543	0.0963
D	0.0150	0.133
E	0.0527	0.120
HR	0.0896	0.157

	RecoveryRate pre-default, All	RecoveryRate post-default, All	LGD
AA/A	0.511	1	0
B	0.502	0.137	0.429774
C	0.502	0.518	0.240036
D	0.501	0.573	0.213073
E	0.501	0.231	0.383731
HR	0.501	0.210	0.39421

Panel 1 reports annualized probabilities of default based on estimates of monthly default probabilities reported in Table 5. Panel 2 reports the pre-default recovery rate, post-default recovery rate, and estimated Loss Given Default (LGD) conditional on credit ratings.

Table 8: Market Efficiency Test

	Efficiency Test, All		Efficiency Test, Online		
	(1) default	(2) default	(3) default	(4) default	(5) default
interueststar	2.28e-10	1.88e-10	-3.41e-09	-3.96e-09	0.000973**
	(5.15e-10)	(5.25e-10)	(6.80e-09)	(6.86e-09)	(0.000442)
post	-0.00119	-0.00322	0.00865*	0.0142*	0.0154
	(0.00314)	(0.00348)	(0.00486)	(0.00792)	(0.0102)
intereststar × post	2.11e-11	5.94e-11	-4.37e-10	-6.33e-10	-0.000232
	(3.81e-10)	(3.82e-10)	(3.25e-10)	(4.67e-10)	(0.000195)
AA/A	-0.0156***	-0.0153***	0.0462	0.0579	
	(0.00365)	(0.00380)	(0.108)	(0.108)	
B	-0.00456	-0.00509	0.0000567	0.00102	-0.000296
	(0.00814)	(0.00819)	(0.00908)	(0.00908)	(0.00735)
C	-0.00712	-0.00753	-0.00841	-0.00732	-0.0353***
	(0.00524)	(0.00521)	(0.00983)	(0.0103)	(0.0132)
D	-0.000358	-0.000823	-0.00512	-0.00426	-0.0300**
	(0.00439)	(0.00442)	(0.00492)	(0.00524)	(0.0151)
E	-0.00245	-0.00289	-0.000348	0.000435	-0.00408
	(0.00292)	(0.00292)	(0.00466)	(0.00477)	(0.00327)
Early	0.0264**	0.0280**	0.00702	0.0403	-0.0108
	(0.0120)	(0.0121)	(0.00873)	(0.0360)	(0.0235)
Mid	0.0266**	0.0282**	0.00990	0.0433	-0.00682
	(0.0120)	(0.0120)	(0.00857)	(0.0371)	(0.0235)
Late	0.0266**	0.0281**	0.0113	0.0446	-0.00751
	(0.0120)	(0.0121)	(0.00891)	(0.0358)	(0.0234)
finit		-0.00000215**		-0.00000192*	-0.00000234**
		(0.000000873)		(0.00000104)	(0.00000106)
activebid		0.000391		-0.0403	-0.00799
		(0.000245)		(0.0395)	(0.0125)
agg_weekly_borrow		0.00108		0.0277	0.0296
		(0.00474)		(0.0575)	(0.0459)
Year-Month FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
CreditRating FE	Yes	Yes	Yes	Yes	Yes
N	204911	204911	9446	9446	11775
R^2	0.0152	0.0154	0.0200	0.0209	0.0199

This table reports the estimates of the market efficiency test equation (6) for the monthly repayment performance of the P2P loans on the Renrendai platform originated between 2012 June and 2013 May, with the repayment performance observation ending in September 2018. Columns (1) and (2) are the results of all samples in the chosen window. Columns (3) and (4) report the estimates for the online sample. Column (5) is the online sample, excluding AA and A rating loans. The dependent variable, default dummy, equals one if loan i defaults in the performance month t. The regressions include credit rating fixed effects, the three-stage loan's life cycle fixed effects, loan issue year-month fixed effects, and borrower city fixed effects. Control variables include liquidity measures and active bidding shares. Standard errors are heteroscedasticity-robust, and clustered by borrower user id. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Market Efficiency Test, Bootstrap

	Efficiency Test, All		Efficiency Test, Online		
	(1)	(2)	(3)	(4)	(5)
	default	default	default	default	default
intereststar	2.28e-10	1.88e-10	-3.41e-09	-3.96e-09	0.000973**
	(5.15e-10)	(5.16e-10)	(6.93e-09)	(6.86e-09)	(0.000442)
post	-0.00119	-0.00322	0.00865	0.0142*	0.0154
	(0.00314)	(0.00305)	(0.00803)	(0.00792)	(0.0102)
intereststar × post	2.11e-11	5.94e-11	-4.37e-10	-6.33e-10	-0.000232
	(3.81e-10)	(3.45e-10)	(3.97e-10)	(4.67e-10)	(0.000195)
AA/A	-0.0156***	-0.0153***	0.0462	0.0579	
	(0.00365)	(0.00295)	(0.107)	(0.108)	
B	-0.00456	-0.00509	0.0000567	0.00102	-0.000296
	(0.00814)	(0.00676)	(0.0143)	(0.00908)	(0.00735)
C	-0.00712	-0.00753*	-0.00841	-0.00732	-0.0353***
	(0.00524)	(0.00433)	(0.00650)	(0.0103)	(0.0132)
D	-0.000358	-0.000823	-0.00512	-0.00426	-0.0300**
	(0.00439)	(0.00487)	(0.00357)	(0.00524)	(0.0151)
E	-0.00245	-0.00289	-0.000348	0.000435	-0.00408
	(0.00292)	(0.00237)	(0.00361)	(0.00477)	(0.00327)
Early	0.0264**	0.0280	0.00702	0.0403	-0.0108
	(0.0120)	(0.0172)	(0.00929)	(0.0360)	(0.0235)
Mid	0.0266**	0.0282	0.00990	0.0433	-0.00682
	(0.0120)	(0.0171)	(0.0103)	(0.0371)	(0.0235)
Late	0.0266**	0.0281	0.0113	0.0446	-0.00751
	(0.0120)	(0.0172)	(0.00998)	(0.0358)	(0.0234)
finit		-0.00000215**		-0.00000192*	-0.00000234**
		(0.000000842)		(0.00000104)	(0.00000106)
activebid		0.000391		-0.0403	-0.00799
		(0.000324)		(0.0395)	(0.0125)
agg_weekly_borrow		0.00108		0.0277	0.0296
		(0.00371)		(0.0575)	(0.0459)
Year-Month FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
CreditRating FE	Yes	Yes	Yes	Yes	Yes
N	204911	204911	9446	9446	11775
R^2	0.0152	0.0154	0.0200	0.0209	0.0199

This table reports the estimates of the market efficiency test equation (6) for the monthly repayment performance of the P2P loans on the Renrendai platform originated between 2012 June and 2013 May, with the repayment performance observation ending in September 2018. Columns (1) and (2) are the results of all samples in the chosen window. Columns (3) and (4) report the estimates for the online sample. Column (5) is the online sample, excluding AA and A rating loans. The dependent variable, default dummy, equals one if loan i defaults in the performance month t. The regressions include credit rating fixed effects, the three-stage loan's life cycle fixed effects, loan issue year-month fixed effects, and borrower city fixed effects. Control variables include liquidity measures and active bidding shares. Standard errors are bootstrapped and clustered by borrower user id. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Market Efficiency Test, Offline Branches

	Efficiency Test, Online	
	(1)	(2)
	Loan Default	Loan Default
treat	0.0427	0.0721
	(0.0965)	(0.103)
postEstab	-0.0541	-0.0508
	(0.0513)	(0.0537)
treat × postEstab	0.285**	0.275**
	(0.128)	(0.134)
rstar	0.0295***	0.0303***
	(0.00196)	(0.00205)
treat × rstar	-0.000831	-0.00196
	(0.00473)	(0.00509)
postEstab × rstar	0.00356	0.00324
	(0.00266)	(0.00287)
treat × postEstab × rstar	-0.0127**	-0.0124**
	(0.00585)	(0.00605)
AA	0.0613	0.0693
	(0.0524)	(0.0524)
A	0.186***	0.192**
	(0.0714)	(0.0760)
B	-0.0844***	-0.0773***
	(0.0208)	(0.0230)
C	-0.0510	-0.0578*
	(0.0319)	(0.0319)
D	-0.212***	-0.209***
	(0.0197)	(0.0193)
E	-0.141***	-0.138***
	(0.0153)	(0.0150)
(sum) normall		0.00000771**
		(0.00000331)
weekly_borrow cyclical component from hp filter		1.27e-10
		(1.93e-10)
agg_weekly_autobid		0.0372
		(0.0447)
agg_weekly_bids cyclical component from hp filter		-1.37e-10
		(1.38e-10)
ddamount_sum cyclical component from hp filter		-1.01e-10*
		(5.69e-11)
EstabDate FE	Yes	Yes
City FE	Yes	Yes
CreditRating FE	Yes	Yes
N	26632	24806
R^2	0.0930	0.0954

This table reports the estimates of the market efficiency test equation (9) for the "Credit" type online P2P loans on the Renrendai platform originated in the window periods of the pooled sample. Columns (1) and (2) report the estimates for the online sample in the pooled sample. The regressions include credit rating fixed effects, treatment city fixed effects, and borrower city's offline branch establish date fixed effects. Control variables include liquidity measures and active bidding shares. Standard errors are robust and clustered by borrower's city. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Offline Branches: Influence on Investor Characteristics

Influence of Offline Branches Introduction on Investor Characteristics

VARIABLES	(1) Investment Frequency	(2) Average Monthly Investment	(3) Normal Bid Proportion
Treat	2.999**	0.136**	0.009
	(1.248)	(0.059)	(0.013)
Post	0.162	-0.046	0.003
	(0.901)	(0.055)	(0.010)
Relative Month	-0.013	0.010	0.001
	(0.115)	(0.007)	(0.001)
Treat $\times Post$	-4.301***	-0.275***	-0.025*
	(1.619)	(0.077)	(0.014)
Company Size	-0.043	-0.035	-0.009**
	(0.723)	(0.028)	(0.004)
Marital Status	0.089	0.035	-0.004
	(0.868)	(0.048)	(0.005)
Education	-0.749	-0.026	0.004
	(0.469)	(0.026)	(0.005)
Work Experience	-0.203	-0.021	0.001
	(0.488)	(0.014)	(0.003)
Age	-0.058	0.001	0.000
	(0.052)	(0.003)	(0.001)
Gender	-1.690*	-0.063*	-0.005
	(0.905)	(0.034)	(0.009)
Property	0.935	0.126**	0.015
	(0.747)	(0.049)	(0.010)
Housing Mortgage	-0.094	-0.028	-0.008
	(0.849)	(0.055)	(0.009)
Income	-0.308	0.020	0.001
	(0.343)	(0.022)	(0.003)
Car	0.030	-0.024	0.002
	(0.970)	(0.051)	(0.009)
Car Mortgage	2.235	0.083	-0.012
	(1.922)	(0.087)	(0.014)
Loan Amount	0.479	0.085***	-0.001
	(0.392)	(0.018)	(0.004)
Maturity	0.038	0.017***	0.001*
	(0.032)	(0.002)	(0.000)
Interest Rate	0.033	0.073***	0.031***
	(0.216)	(0.012)	(0.003)
Constant	24.270***	8.313***	0.306***
	(5.212)	(0.273)	(0.049)
Observations	4,898	4,898	4,898
R-squared	0.101	0.169	0.111

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

This table reports the estimates of equation (13). The column names represent the dependent variables in each regression. The regressions include borrowers' city fixed effects, offline branch city fixed effects, and time fixed effects of the loan. Control variables include borrowers characteristics, loan characteristics, the number of months relative to the offline branch establishment date, and borrowers' transportation distance to the corresponding offline branch. Standard errors are robust and clustered by borrower's city. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix A. Additional Results

A1. Additional Tables

Table A1: Balance Tests

	Before			Post			
	n	mean	sd	n	mean	sd	Diff
companysize	14665	2.28	1.03	25273	2.26	0.98	-0.017*
marriage	14747	1.56	0.52	25428	1.68	0.57	0.119***
education	14743	1.86	0.77	25426	1.83	0.78	-0.035***
workexperience	14739	2.46	1.02	25417	2.46	0.99	-0.006
age	14747	35.41	7.47	25429	36.83	10.41	1.428***
property	14747	0.43	0.50	25429	0.33	0.47	-0.102***
housingmortgage	14747	0.12	0.32	25429	0.09	0.29	-0.026***
income	14741	3.95	1.32	25422	4.21	1.36	0.260***
car	14747	0.28	0.45	25429	0.30	0.46	0.022***
carmortgage	14747	0.05	0.21	25429	0.04	0.19	-0.009***

	Online			Offline			
	n	mean	sd	n	mean	sd	Diff
companysize	6432	2.84	0.99	636663	1.36	0.71	-1.483***
marriage	7011	1.68	0.57	636663	1.79	0.61	0.109***
education	7009	2.02	0.94	636663	2.09	0.96	0.065***
workexperience	7011	2.75	1.02	628560	2.43	1.14	-0.325***
age	7011	35.05	6.82	655179	36.36	11.28	1.302***
property	7011	0.51	0.50	636841	0.52	0.50	0.006
housingmortgage	7011	0.21	0.41	636841	0.25	0.43	0.033***
income	7011	4.09	1.15	636663	4.69	1.18	0.605***
car	7011	0.28	0.45	636841	0.34	0.48	0.064***
carmortgage	7011	0.07	0.25	636841	0.09	0.29	0.024***
group(guaranteefee)	7011	7.80	0.53	655179	2.80	1.02	-5.003***

The first panel of the table reports results of balance tests in the window period (2012 May to 2013 June). The second panel reports results of balance tests for the period after 2012 November. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Regression Results: 6-month window

Regression Results Before and After the Event			
	(1)	(2)	(3)
VARIABLES	Before	After	PSM
Company Size	0.017*	0.000	-0.003
	(0.008)	(0.001)	(0.019)
Marriage	0.006	-0.003	0.030
	(0.027)	(0.002)	(0.047)
Education	-0.024*	-0.004**	-0.006
	(0.012)	(0.002)	(0.016)
Work Experience	-0.029***	-0.000	-0.020
	(0.006)	(0.001)	(0.024)
Age	0.003*	-0.000	-0.004
	(0.001)	(0.000)	(0.004)
Gender	-0.018	-0.002	0.009
	(0.025)	(0.002)	(0.037)
Property	-0.004	0.024	-0.024
	(0.016)	(0.015)	(0.024)
House Mortgage	-0.014	0.004	0.012
	(0.021)	(0.021)	(0.047)
Income	0.009	0.003**	-0.002
	(0.009)	(0.001)	(0.008)
Car	-0.006	-0.008**	-0.029
	(0.017)	(0.004)	(0.022)
Car Mortgage	0.000	0.019	-0.030
	(0.032)	(0.022)	(0.060)
Log Loan Amount	0.008	-0.003	-0.007
	(0.011)	(0.003)	(0.015)
Loan Maturity	0.011***	-0.001***	0.023***
	(0.002)	(0.000)	(0.007)
Interest Rate	0.008**	0.017***	0.009
	(0.004)	(0.004)	(0.008)
Constant	-0.211**	-0.148***	0.082
	(0.095)	(0.048)	(0.188)
Difference Test	0.173***	N/A	0.298***
	(0.006)	N/A	(0.035)
Week FE	Yes	Yes	No
Province FE	Yes	Yes	Yes
Observations	886	7,061	7,060
R-squared	0.196	0.097	0.334

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table reports the regression results for the specification in Equation (1). Column (1) reports the regression results on the loans in the 6-month pre-window (June 2012 to November 2012). Column (2) reports the regression results on the loans in the 6-month post window (December 2012 to May 2013). Column (3) reports the regression results on the propensity score matched sample. Observations are matched on the 10 borrower characteristics included in Equation (1). *Difference Test* reports the mean difference between the predicted default rates and the published default rates in the post window (December 2012 to May 2013). Standard errors are clustered at the month level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A3: PSM Regressions with Different Matching Criteria, Part 1

VARIABLES	(1) No Company Size	(2) No Education	(3) No Marriage	(4) No Work Experience	(5) No Age
comp_size	0.020	0.022	-0.024	0.016	0.012
	(0.022)	(0.029)	(0.025)	(0.023)	(0.021)
marriage	-0.088**	-0.051	-0.019	-0.090**	-0.148**
	(0.037)	(0.040)	(0.040)	(0.035)	(0.063)
education	-0.007	0.009	-0.015	0.008	-0.002
	(0.031)	(0.034)	(0.032)	(0.033)	(0.041)
work_experience	0.010	0.075*	0.038	0.108**	0.070**
	(0.027)	(0.038)	(0.028)	(0.044)	(0.033)
age	0.003	0.002	-0.003	-0.000	0.003
	(0.003)	(0.004)	(0.002)	(0.004)	(0.004)
Gender	0.000	0.041	-0.006	0.016	-0.000
	(0.044)	(0.069)	(0.068)	(0.100)	(0.085)
property	-0.048	0.037	0.012	0.003	0.091
	(0.055)	(0.114)	(0.055)	(0.065)	(0.085)
housing_mortgage	0.036	0.059	0.144*	-0.038	0.026
	(0.057)	(0.086)	(0.083)	(0.075)	(0.092)
income	-0.005	-0.001	0.014	-0.011	-0.002
	(0.010)	(0.015)	(0.013)	(0.022)	(0.011)
car	0.016	-0.048	-0.024	-0.053	-0.035
	(0.028)	(0.032)	(0.058)	(0.040)	(0.050)
car_mortgage	-0.010	-0.029	-0.131	-0.035	-0.132**
	(0.033)	(0.046)	(0.105)	(0.050)	(0.060)
l_amount	-0.058**	-0.011	-0.060*	-0.014	-0.066***
	(0.021)	(0.020)	(0.034)	(0.041)	(0.020)
monthsloan	0.028***	0.018*	0.027*	0.032***	0.029***
	(0.009)	(0.009)	(0.013)	(0.008)	(0.009)
interestloan	0.009**	0.010	0.003	0.008	0.017**
	(0.004)	(0.008)	(0.005)	(0.007)	(0.008)
Constant	0.161	-0.451	0.448	-0.329	0.010
	(0.240)	(0.449)	(0.371)	(0.532)	(0.315)
Difference Test	0.290***	0.201***	0.277***	0.386***	0.283***
	(0.030)	(0.019)	(0.025)	(0.040)	(0.022)
Province FE	Yes	Yes	Yes	Yes	Yes
Observations	2,187	2,186	2,187	2,187	2,187
R-squared	0.652	0.520	0.467	0.695	0.605

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table reports the regression results on the samples matched using Propensity Score Matching (PSM). The matching criteria for different columns are different. The column header indicates which borrower characteristic is dropped out from the the ten borrower characteristics when conducting PSM. For example, *No Company Size* means that when matching the observations, we only match the other nine borrower characteristics except for *Company Size*. *Difference Test* reports the mean difference between the predicted default rates and the published default rates in the post window (December 2012 to January 2013). Standard errors are clustered at the month level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A4: PSM Regressions with Different Matching Criteria, Part 2

VARIABLES	(1) No Property	(2) No House Mortgage	(3) No Income	(4) No Car	(5) No Car Mortgage
comp_size	-0.025	-0.003	0.006	0.018	0.004
	(0.018)	(0.023)	(0.023)	(0.026)	(0.034)
marriage	0.009	-0.067	-0.072**	-0.220***	-0.098**
	(0.038)	(0.043)	(0.031)	(0.058)	(0.046)
education	0.001	0.012	0.005	0.034	0.010
	(0.017)	(0.026)	(0.032)	(0.039)	(0.033)
work_experience	0.026*	0.057**	0.039	0.048	0.078**
	(0.015)	(0.021)	(0.031)	(0.034)	(0.035)
age	-0.003	-0.001	0.000	0.010**	0.002
	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)
Gender	0.009	-0.012	-0.008	0.095	0.022
	(0.029)	(0.066)	(0.074)	(0.092)	(0.080)
property	0.027	-0.013	0.058	-0.016	0.090
	(0.078)	(0.062)	(0.085)	(0.090)	(0.108)
housing_mortgage	0.092	0.043	0.014	0.229*	0.093
	(0.068)	(0.034)	(0.097)	(0.112)	(0.086)
income	0.026**	0.010	-0.002	0.014	-0.010
	(0.012)	(0.011)	(0.012)	(0.014)	(0.010)
car	-0.013	-0.015	-0.031	-0.036	-0.064*
	(0.058)	(0.056)	(0.032)	(0.054)	(0.032)
car_mortgage	-0.029	-0.085	-0.045	-0.042	-0.035
	(0.046)	(0.057)	(0.044)	(0.057)	(0.049)
l_amount	-0.053**	-0.049*	-0.041	-0.091**	-0.052*
	(0.023)	(0.025)	(0.024)	(0.036)	(0.027)
monthsloan	0.020*	0.028**	0.023**	0.021**	0.011
	(0.010)	(0.011)	(0.010)	(0.009)	(0.009)
interestloan	0.007	0.005	0.011**	0.027**	0.025*
	(0.009)	(0.005)	(0.005)	(0.012)	(0.014)
Constant	0.287	0.132	0.055	-0.232	-0.082
	(0.206)	(0.313)	(0.407)	(0.533)	(0.429)
Difference Test	0.204***	0.292***	0.234***	0.222***	0.121***
	(0.021)	(0.030)	(0.021)	(0.008)	(0.009)
Province FE	Yes	Yes	Yes	Yes	Yes
Observations	2,186	2,187	2,187	2,186	2,186
R-squared	0.664	0.485	0.385	0.711	0.387

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table reports the second half of the regression results on the samples matched using Propensity Score Matching (PSM). The matching criteria for different columns are different. The column header indicates which borrower characteristic is dropped out from the ten borrower characteristics when conducting PSM. For example, *No Company Size* means that when matching the observations, we only match the other nine borrower characteristics except for *Company Size*. *Difference Test* reports the mean difference between the predicted default rates and the published default rates in the post window (December 2012 to January 2013). Standard errors are clustered at the month level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A5: Rosenbaum Bound Test

Gamma	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	0	0	.3385	.3385	.3285	.3485
1.5	0	0	.2958	.3817	.2862	.3920
2	0	0	.2671	.4121	.2569	.4227
2.5	0	0	.2450	.4354	.2354	.4465
3	0	0	.2279	.4546	.2176	.4667

The table reports the Rosenbaum Bound Test on the difference between default rates predicted by the PSM regression results in column (3) of Table A2 and the default rate reported by Renrendai. Gamma is the log odds of differential assignment due to unobserved factors. Sig+ is the upper bound significance level. Sig- is the lower bound significance level. T-hat+ is the upper bound Hodges-Lehmann point estimate. T-hat- is the lower bound Hodges-Lehmann point estimate. CI+ is the upper bound confidence interval (a= .95). CI- is the lower bound confidence interval (a= .95)

Table A6: Regression for Offline Branches: 6-month Window

	Offline Establishment	
	(1)	(2)
VARIABLES	DID	Pre-Window
Treat	0.088***	
	(0.028)	
Post	0.005**	
	(0.002)	
Relative Month	0.001	-0.000
	(0.000)	(0.001)
Treat * Post	-0.087***	
	(0.028)	
Company Size	0.023***	0.015***
	(0.004)	(0.003)
Marriage	-0.004***	-0.004***
	(0.001)	(0.001)
Education	-0.006**	-0.007***
	(0.002)	(0.003)
Work Experience	0.024***	0.036***
	(0.004)	(0.005)
Age	-0.000***	-0.000***
	(0.000)	(0.000)
Gender	-0.001	-0.001
	(0.001)	(0.001)
Property	0.006*	0.003
	(0.003)	(0.004)
Housing Mortgage	-0.008**	-0.008**
	(0.003)	(0.003)
Income	0.004***	0.005***
	(0.001)	(0.001)
Car	-0.002**	-0.003**
	(0.001)	(0.001)
Car Mortgage	-0.002	-0.001
	(0.001)	(0.001)
Log Loan Amount	-0.029***	-0.026***
	(0.003)	(0.004)
Loan Maturity	0.001*	0.001**
	(0.000)	(0.000)
Log Distance	0.003**	-0.005*
	(0.001)	(0.002)
Constant	0.232***	0.287***
	(0.026)	(0.052)
Loan Time FE	Yes	Yes
Branch Open Time FE	Yes	Yes
Borrower Living City FE	Yes	Yes
Observations	201,832	76,404
R-squared	0.151	0.174

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Column (1) reports the difference-in-difference regression results based on Equation (2). Column (2) reports the regression results on all observations in the pre-window. Standard errors are clustered at month level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix B. Notes

B1. Renrendai and Ucredit

Renrendai Renrendai provides an online platform to serve borrowers and lenders (investors). We collect P2P loan data from this platform.

Ucredit The full Chinese company name is Youzhong Xinye Company. The company provides services to prospective borrowers, including: (1) assessing borrowers' creditworthiness and likelihood of obtaining loans; (2) helping borrowers register and facilitate P2P loan matching at the cooperating P2P lending platform Renrendai; (3) helping borrowers manage their repayment accounts and repay for borrowers in case of defaults. It charges loan matching service fees from borrowers. In case of defaults, it repays to current creditors (investors or lenders) for the borrower, charges repayment service fees from borrowers and gains creditorship. As the creditor, it also charges defaulted borrowers the overdue penalty fees. Ucredit started to cooperate with Renrendai in November 2012, helping borrowers post online applications (marked as "Field" type loans).

Ucredit City Branches In some Chinese cities, there are several Ucredit branches which are offline offices, specializing in consulting services and marketing. The services provided include: (1) explaining the lending process to prospective borrowers; (2) assisting them in gathering, organizing, and submitting loan application documents; (3) verifying documents; (4) facilitating P2P loan transactions. The company charges consulting fees from borrowers.

Example. On March 16, 2017, Customer Wu signed a Credit Consultation and Management Services Agreement with Ucredit and Ucredit Chongqing branch. Ucredit assisted Customer Wu in obtaining a P2P loan successfully through the Renrendai platform (Loan ID 1157588) with an amount of 15,600 RMB, an annual interest rate of 10.20%, and a maturity period of 36 months. A digital loan agreement was signed among Customer Wu, Renrendai, creditors (investors participating in the loans through Renrendai), Ucredit and Uredit Chongqing. Customer Wu agreed to make monthly repayments of 504.83 RMB to the creditors from April 16 2027 to March 16, 2020.

According to agreements, Customer Wu pays a monthly service fee of 23.33 RMB to Ucredit Chongqing branch (totaling 840 RMB for 36 months) and a monthly service fee of 132.22 RMB to Ucredit (totaling 4,760 RMB for 36 months). Customer Wu started to default on January 16, 2018. As previously agreed, on January 16, 2018, February 16, 2018, and March 16, 2018, Ucredit repaid 505.44 RMB for Customer Wu to creditors. On April 16, 2018, Ucredit repaid 505.44 RMB, together with the remaining loan principal of 10,489.44 RMB. Ucredit became the creditor of Loan ID 1157588. [8]

[8] The example details come from a public civil judgment document in 2020.