Strategic Credit Risk Disclosure in Marketplace Lending

Yutong Hu* Shasha Li†

Latest draft here

June 16, 2023‡

ABSTRACT

We document that online marketplace lending platforms only selectively disclose their data, which significantly reduces market efficiency. Using peer-to-peer (P2P) lending data from a leading platform in China, we observe that the platform substantially under-reports actual loan default rates by cooperating with offline sister companies and using the risk control fund. Our baseline estimations suggest that the monthly default rates are under-reported by 5.26 percentage points on average. The Loss Given Default (LGD)-adjusted interest rate in the online lending market is not at a market-efficient level. The behavior of hiding default data drives the online market further away from information efficiency and lowers the quality of active borrowers. Moreover, liquidity plays an essential role in the deviations of the market prices from the information-efficient level.

Keywords: Marketplace Lending, Information Disclosure, Market Efficiency

JEL classification: G14, G23, G33, G50

*London Business School, yhu@london.edu.

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

‡This is a preliminary draft. The authors thank Tim Adam (discussant), Svetlana Bryzgalova, Julian Franks, Hira Ghaffar (discussant), Anna Pavlova, Darcy Pu, Stefano Rossi, Tarun Ramadorai, Nicolas Serrano-Velarde, Xu Zhang and Yueteng Zhu (discussant) for their helpful comments and suggestions. All errors remain ours.

1. Introduction

The rise of financial technology, so-called Fintech, is reshaping the landscape of financial services. More recently, COVID has accelerated the digital transformation. Banks are rapidly embracing Fintech collaboration, and the oversight of Fintech is crucial from a financial stability perspective. Barba Navaretti and Pozzolo (2021) argues that the Fintech oversight requires a good understanding of their complex business models. In Fintech lending, the transparency on delinquency and pricing has long been the sticking point[1]. There's lack of mechanism to ensure that innovative lenders do not take on excess risk becoming insolvent. The vital role of information design in Fintech lending has been widely discussed in the literature (e.g. Iyer et al., 2016; Vallee and Zeng, 2019; Franks et al., 2021). The type and the extent of information disclosed by the market organizers are crucial, because participants rely on the visible information on the platforms to make borrowing or investment decisions.

We detect the evidence of selective disclosures on loan defaults in the online marketplace lending market and investigate its impact on market outcomes, especially informational efficiency. Online marketplace lending platforms work as information intermediaries to facilitate transactions between borrowers and lenders (investors). We use a dataset of Peer-to-Peer (P2P) loan applications and repayments manually collected from the Renrendai platform, one of China's biggest online lending platforms. The data contains all loan applications between January 2012 to December 2016. We observe the repayment performance of successfully funded loans until September 2018.

The loan-level data of Renrendai shows a puzzling feature: the annual interest rate on average is more than 10%, about twice the bank rates. In contrast, the annual default rate on average is surprisingly lower than 3%. In some periods, default rates are even close to zero. However, according to recent studies on the Fintech credit market, online borrowers are usually under-served by the banks and at the lower end of the credit quality spectrum (e.g., Tang, 2019) and thus are supposed to have higher default risk. Furthermore, in contrast to the extremely low credit risk disclosed, Renrendai faced a massive risk of debt overdue in 2021, shocking investors and the public. In 2021, a group of investors sued Renrendai for its non-transparent practices and a lack of

[1] In 2016, the U.S. Treasury Department released a white paper to call for greater transparency in online lending, including transparent loan performance metrics, standardized loan-level data, and clear pricing terms. In July 2022, the Committee on Small Business held a hearing on "Fintech and Transparency in Small Business Lending" and pointed out the need to improve transparency in small business lending, where Fintech plays an important role.

disclosure of operating status and financial condition. [2] All the evidence implies that Renrendai is hiding its actual performance.

On the one hand, online lending platforms have strong incentives to selectively disclose information to attract more customers, because their service fees are proportional to the value of successful transactions. The lower the default risk disclosed, the more investors are willing to put money into the platform. On the other hand, the platforms have the ability to garble and obfuscate the data. Furthermore, information manipulation has long been a topic in traditional financial markets, including in the lending market. For example, banks strategically understate the risk, especially when systematic risk is high (Begley et al., 2017) or manipulate credit ratings before sharing with competitors to keep informational rent (Giannetti et al., 2017). Fintech lenders are no exception, and this study provides novel empirical evidence for the online lending platforms' information hiding behavior.

We find that Renrendai's credit risk data manipulation started in November 2012 by cooperating with an offline lending service company, Ucredit (Youxin in Chinese). Ucredit offline teams contact potential borrowers, obtain new loan applications and post the loan requests on the Renrendai online platform. Renrendai marks the loan applications from offline networks as "Field" type, different from the "Credit" type loans directly requested online. From November 2012, default rates disclosed on Renrendai suddenly went down. Strikingly, the default rates of "Field" type loans coming from offline UCredit offices were reported to be zero. We find that, in an event where a "Field" type borrower cannot repay, Ucredit uses its risk control fund, also called safeguard fund, to repay the online lenders. Renrendai does not disclose such borrower defaults. The risk control fund comes from borrowers. Upon loan approval, the borrower pays a credit-rating-based fee, proportional to the loan amount, to the risk control fund.

First, with the observed facts in mind, we use empirical methods to find statistical evidence of default rates under-reporting. To detect default rate manipulation, we use borrower characteristics to estimate what the actual default rates should be if there is no data manipulation. Renrendai borrowers, offline and online, post detailed personal information, including income, age, marriage, etc., online when they request loans, and lenders make lending decisions based on the observed borrower characteristics. We specify a regression model using these borrower characteristics to

[2] https://min.news/en/tech/6bbb31c76f39767719e5c46fb0fb769a.html

estimate default rates. We focus on short windows before and after Renrendai's cooperation with offline Ucredit. We use the regression model estimated in the pre-window as a benchmark to predict the post-window default rates. We begin with 2-month pre-window and 2-month post-window. Within a short 2-month window, the loan default prediction model is not likely to change quickly. We then expand the window to 6-month before and 6-month after to have more observations, which can help us better evaluate the consequences. To rule out the influence of changes in overall borrower characteristics, we also use Propensity Score Matching (PSM) to match the borrowers in the post window to the borrowers in the pre window. The matching criteria are based on the borrower characteristics. Using PSM, we can proxy whether a borrower is likely to default without the impact of Renrendai's merger with Ucredit.

We start from the nationwide platform data and take a 2-month (6-month) window before and after the first introduction of "Field" type loans from Ucredit to the Renrendai platform in November 2012. Our estimation results suggest that Renrendai largely under-reports the default risk of loans. Based on the regression results estimated in the pre-window, the predicted monthly default rates in the post-window are, on average, 9.1 (21.7) percentage points higher than the reported default rates. Even after matching the borrower characteristics, we find that the estimated monthly default rate is, on average, 12.6 (35.4) percentage points higher than the published default rates. Results stay robust with different PSM matching criteria, and the Rosenbaum Bound test shows that our results are insensitive to hidden bias.

Further investigating how Renrendai processes the loans for "Field" type borrowers, we find that Renrendai's sister company Ucredit has established offline branches all over China at different times since the end of 2012. Renrendai and Ucredit use these offline branches to receive loan applications and manage loans for "Field" type borrowers, whose disclosed default rates are most significantly under-reported. We collect the city-level locations and opening dates of the Ucredit offline branches from 2012 to 2017. We construct a pooled sample comprised of sub-samples based on each treatment city with an offline branch introduced in our sample period. Again, we take 6-month before and after windows for each offline branch opening date. We conduct a difference-in-difference analysis on the pooled sample and find that opening offline branches is associated with lower reported default rates. Our empirical evidence shows that, if we set the default rate reporting in the pre window as a benchmark, the average effect of each offline branch opening is associated with 5.26 percentage points under-reporting of monthly default rates.

Second, with clear evidence showing that the platform is indeed hiding actual default rates, we further explore the consequences of the manipulation. Following Franks et al. (2021), we test the market efficiency by regressing the default dummy on LGD-adjusted interest rate. The regression model is derived from the investor's participation constraint. If the market is efficient, we expect a unit coefficient. Starting from the first window period of 2012 June to 2013 May, we use the monthly repayment performance data and find that the online lending market is far from market efficiency. The coefficients of LGD-adjusted interest rates are close to zero both in the periods with and without default data hiding, and credit ratings retain predictive power, which contradicts the market efficiency predictions. However, we find similar results with Franks et al. (2021) and Iyer et al. (2016) that the coefficient of LGD-adjusted interest rate significantly predicts default probability over and above credit ratings. More interestingly, our results suggest that the selective disclosure of default data leads to further deviation from market efficiency as the coefficient becomes lower with the introduction of offline loans. In addition, liquidity variation drives the market prices away from the information-efficient level.

In the next step, we use the pooled sample to test the efficient market hypothesis and again find that introducing "Field" type loans leads to further deviation from market efficiency. With the pooled sample, we consider the staggered introduction of offline branches in different cities and construct a comparable control group for each treated city. The pooled sample has more observations to help us understand the channels and consequences better, and can also control time-invariant city-specific and macroeconomic fluctuations. We analyze the sample data of successfully funded loans and all applications separately. We find that the credit quality of online borrowers in a city with offline branches is poorer than that of online borrowers in a city without an offline branch.

To the best of our knowledge, this paper is the first to study data manipulation problems in Fintech credit. Selective information disclosures cause inefficient pricing of default risks. Although we detect and examine the impact of default data manipulation in a Chinese online lending platform, we believe our results and stories behind are informative about data transparency and financial stability in general. The behavior of hiding bad debts is not rare in Fintech lending models. It may become more of a problem if banks adopt Fintech models more widely during the digital transformation process. Most P2P lending companies in China, including the top companies such as Yirendai and Lufax, take the online-offline approach and can selectively hide loan defaults. The safeguard fund policy is also a popular approach worldwide in the Fintech credit industry that can help hide actual loan performance. For instance, Zopa, the world's first-ever P2P lending company in the U.K., had a Safeguard plan, and it claims the default rate since it launched in 2005 is 0.6%, which is extremely low. Hence, understanding how the Fintech credit companies operate and disclose information is essential. Effective regulation and oversight in information disclosure, especially credit risk disclosure, is in need.

2. Literature Review

This paper contributes to several strands of literature. First, our work relates to the literature on information efficiency in the Fintech industry. Iyer et al. (2016) find that online lending markets that rely on nonstandard information to screen the peer borrowers' creditworthiness can predict the likelihood of default 45% more accurately. Their study suggests that aggregating the views of peers and leveraging nonstandard information can enhance the efficiency of online lending. Vallee and Zeng (2019) investigate the importance of information distribution on marketplace lending platforms. They find that sophisticated investors systematically outperform, and the outperformance shrinks when the lending platform reduces information provided to investors. The closest work to our study is Franks et al. (2021). In their study, a leading British peer-to-business platform retreat from auctions, and the platform sets prices and allocates credit on its own instead. The study shows that the change makes the platform vulnerable to liquidity shocks, which leads to deviations from information efficiency. Liao et al. (2021) use the same dataset from Renrendai and find that investors appear to primarily focus on interest rates to make their investment decisions and largely ignore the credit ratings.

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 the Fintech industry.

3. Data and the Platform

3.1. 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%, 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.

3.1.1. 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 safeguard 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

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

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 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.

3.2. 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.

3.3. Data

Our data contains 862,232 loan applications on the Renrendai P2P lending platform, applied between January 2012 and January 2017. The data is hand collected and is accessible to all users on the official site of Renrendai. For each borrower, we can observe the repayment flows until September 2018, the loan status (normally repaid or default) for each loan, and the characteristics of the borrowers.

Among these 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 loan granted was 55.8 billion RMB (about $8.79 billion). 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 house 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).

3.4. Sudden drop in default rate

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 from 2012 Nov. We find that the sudden drop in default rates is 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)[4]. 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

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

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 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 will also explore the impact of the default rate "manipulation" on the platform's market efficiency.

4. Empirical Analysis

In this section, we describe the method and results of empirical analysis. First, we detect the default data manipulation on Renrendai. Second, we examine the impact of default data manipulation on online lending market outcomes, especially market efficiency.

4.1. Detect Default Data Manipulation

4.1.1. 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 the end of November 2012 as the breaking point when Renrendai integrated 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 the end of November 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 meantime, we then expand the window to 6-month 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:

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

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 the borrower has. $WorkExperience_{it}$ 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. 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 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. Nevertheless, in the meantime, the borrowers in the post window have 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.

4.1.2. Default Rate Estimations

Table 2 shows the OLS regression results using the 2-month pre window and the post window. Column (1) shows the regression results on the loans from October 2012 to November 2012. Column (2) shows the regression results on the loans from December 2012 to January 2013. The table shows that the significance of the coefficients is different before and after the event. The loan amount loses its prediction power in the post window. In the post window, the education level, income, and whether the borrower owns a car 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.

In Figure 6, we plot the weekly default rates in the data together with the predicted weekly default rates. The left figure are 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 only for the 2-month window results, because there are only two months in the sample. For other analyses in the paper, we consistently use monthly default rates. The predicted default rates are based on the regression results of the pre-window data. The red vertical line indicates the event time, November 2012. The plot shows that after November 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, 9.1 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 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 2012. The results of the difference test show that, in the post window, the predicted default rates are, on average, 12.6 percentage points significantly higher than the published default rates.

Table 3 shows the OLS regression results using samples in the 6-month pre and post windows. Work experience, gender, and whether the borrower owns a car lose their prediction power in the post window. In the post window, whether the borrower owns a property becomes a significant predictor of default rates. In Figure 7, we plot the monthly default rates in the data together with the predicted monthly default rates. The plot shows that after November 2012, the predicted default rates are consistently higher than the published default rates. The results using 6-month windows validate our conclusions using 2-month windows. The difference test indicates that the predicted monthly default rates are, on average, 21.7 percentage points significantly higher than the published default rates.

Column (3) of Table 3 reports the regression results on the propensity-score-matched sample using all borrower characteristics included in Equation (1). The second graph of Figure 7 plots the predicted default rates based on the PSM results versus the published default rates. The predicted default rates are still consistently higher than the reported default rates after November 2012. The difference test results show that, in the post window, the predicted default rates are, on average, 35.4 percentage points significantly higher than the published default rates.

In general, both the baseline regression results and the PSM regression results support that the default rates reported by Renrendai after November 2012 are substantially lower than they should be if they reported the default rates as truthfully as before the company consolidation. Because PSM results largely depend on the variables used to implement the match, we change the PSM matching criteria in the following subsection to test the robustness of the PSM regression results.

4.1.3. Robustness of PSM Results

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 A2 and Table A3 in Appendix show the regression results on samples matched using different criteria. For example, for column (1) of Table A2, which is labeled as "No Company Size", the matched sample is matched on all the other nine borrower characteristics except for *Company Size*. Similarly, "No Marital Status" means that the matched sample is matched on all the other nine borrower characteristics except for *Marital Status*. 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.

In addition to testing the PSM results using different matching criteria, we apply the Rosenbaum Bound Test to the difference between the reported default rates and predicted default rates using estimation in Column (3) of Table 3. Rosenbaum Bounds examines whether the average treatment effects on the treated are sensitive to hidden bias. Table A4 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 and insensitive to different levels of hidden bias measured by gamma.

These results support the robustness of our conclusion that, if Renrendai has reported the default rates as truthfully as before the platform consolidated with UCredit, the default rates after November 2012 should be significantly higher than the default rates currently reported in their data.

4.1.4. 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 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. Column (1) in Table 4 shows the regression results. The difference test in the table shows the mean difference between the predicted default rates and the reported default rates for the offline borrowers after November 2012. The difference test shows that, assuming the default rates for online borrowers are truthfully reported, the default rates for the offline borrowers are under-reported by 44.9 percentage points on average. Figure 8 plots the predicted monthly default rates for the offline borrowers in the red line. The blue horizontal line plots the reported zero default rates.

In addition to directly estimating the default prediction model on the online borrowers, we also use the Propensity Score Matching method to proxy the default rate for offline borrowers. In the second half of Table A1, we compare the characteristics of online and offline borrowers after November 2012. The difference test shows that the average characteristics for online and offline borrowers are significantly different.

To rule out the effect of changes in borrower characteristics, we 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. We then run the regression specified in equation (1) on the matched samples, and compare the predicted default rates with the reported default rates. The results are reported in column (2) of Table 4. The difference test shows that the predicted default rates are 54.9 percentage points higher than the reported default rates on average. The second graph in Figure 8 plots the predicted monthly default rates. Table A5 shows the Rosenbaum bounds for the difference between the predicted and reported default rates for the offline borrowers. Both the lower bound and higher bound are significantly higher than zero and insensitive to unobserved bias.

To test the robustness of the PSM results, we change the matching criteria for PSM using the same methods as in Section 4.1.3. Table A6 and Table A7 show the regression results. Figure **??** shows the predicted monthly default rates. The difference tests in Table A6 and Table A7 show that the predicted default rates are consistently higher than zero.

The evidence above shows that assuming the default rates for online borrowers are truthfully reported, Renrendai significantly under-reports the default rates for offline borrowers after November 2012.

4.1.5. 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 they have access to them. The offline loan officers create loan request listings on the Renrendai platform on behalf of verified borrowers. Of course, borrowers can still directly requests P2P loans online by themselves. In the previous subsections, we have found evidence that the default rates for offline borrowers are largely manipulated downward. In this subsection, we further explore the impact of each offline branch establishment on the default rates and quantify how much are the default rates 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. The first offline branch in our sample was established in Chengdu city on November 26th of 2012, and the last offline branch was established in Nanning city on April 12th of 2016.

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 six 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 is originated after the corresponding offline branch j opens and equals 0 if it is 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 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 5. The coefficient for the interaction term is negative and highly significant. The magnitude and sign of the coefficient indicates that the opening of an offline branch is on average associated with a 10.1 percentage points 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 6-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 5. 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 5.3 percentage points monthly. The magnitude of default rate manipulation is much smaller than the 21.7 percentage points under-reporting of monthly default rates in Table 3, and the lower under-reporting of default rate is reasonable. In Table 3, the loans analyzed were applied in the early stage of Renrendai's operation. A large fraction of online borrowers defaulted before November 2012 due to poor risk management strategies. The loans used to estimate the effects of opening offline branches were applied between 2012 and 2017. The management strategies of the Renrendai platform develop over time to allow them to choose borrowers with better quality and collect bad debt in more effective ways. Hence, the monthly default rates decrease over time, and the average under-reported default rates between 2012 to 2017 also decrease to a lower level.

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

4.2. Market Efficiency Test

As shown above, the Renrendai platform hides the 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 November 2012 window period. We use 205,278 monthly repayment performance 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.

4.2.1. 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 condition,

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

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 $\alpha^\star = -\dfrac{\rho}{LGD_i^e}$ and $r_i^\star = \dfrac{r_i}{LGD_i^e}$. 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.

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

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,

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

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.

4.2.2. 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[5].

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 9 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 10.

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,

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

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

where $Default_{it}$ is loan i's default dummy at performance month t. As before, FE_Rating and 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 6 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 8. 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 8. 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 8 shows the results of pre-default recovery rates for all credit ratings.

$\langle$*insert Table 6 and 8 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,

$$RRecoveryPost_i = \alpha + \theta \times FE\_Rating_i + \psi \times MRecovery_i + \gamma \times FE\_MIssue_i + \epsilon_i \qquad (8)$$

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 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[6].

Table 7 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 7 here*⟩

Results of Market Efficiency Test Table 9 shows the results of market efficiency test (equation (6)). Table 9 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. Column (2), (4) and (5) of Table 9 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.

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

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 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.

4.2.3. 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-month before and 6-month 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-month before and 6-month 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 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,

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

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-sample's 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 11 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 integral 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 10 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.

4.2.4. Other results

Investor Confidence Figure 9 depicts that interest rates of offline P2P loans decrease over time, suggesting increasing investor confidence in the platform. Because Renrendai and its offline sister companies will use the safeguard fund to recover all offline P2P loans fully, the only risk investors face is the platform run. Thus, interest rates of offline P2P loans capture online investors' confidence in the Renrendai platform. The higher the interest rate, the less confident the investor is on the platform. The lowering interest rate suggests that investor confidence, on average, increases over time.

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.

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

We begin with the sample of successful applications to test model (11). Figure 11 plots the coefficients of each $FE\_RelativeMonth \times Treat$ level in the regression model (refeqn:out1) 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 12 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 13 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 14). In addition, Figure 15 and 16 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. Conclusion

Puzzled by online lending's extremely low default rate and relatively high-interest rate, we detect default data manipulation and study its impact on marketplace lending outcomes. In this paper, we find statistical evidence showing that the online platform largely under-reports monthly default rates by 14% on average. The data manipulation on loan repayment performance drives the interest rate of online P2P loans away from market efficiency. Liquidity shock also makes the prices deviate from information-efficient levels.

The latest wave of Fintech innovation brings modern and convenient business models never observed before, and it is essential to keep up with the innovation. The most recent global financial crisis is a lesson about information transparency. If this time we ignore again what is really going on in the new Fintech sector, we may put our financial system in danger again. This paper points out the existence of manipulative practices of Fintech platforms in terms of information disclosure. It raises important policy questions concerning consistent and transparent information disclosure of Fintech companies.

References

Barba Navaretti, G, G. C., Pozzolo, A. F., 2021. Regulating fintech in europe: Lessons from the collapse of wirecard. VoxEU.org.22 November .

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

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

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.

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.

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

Figures

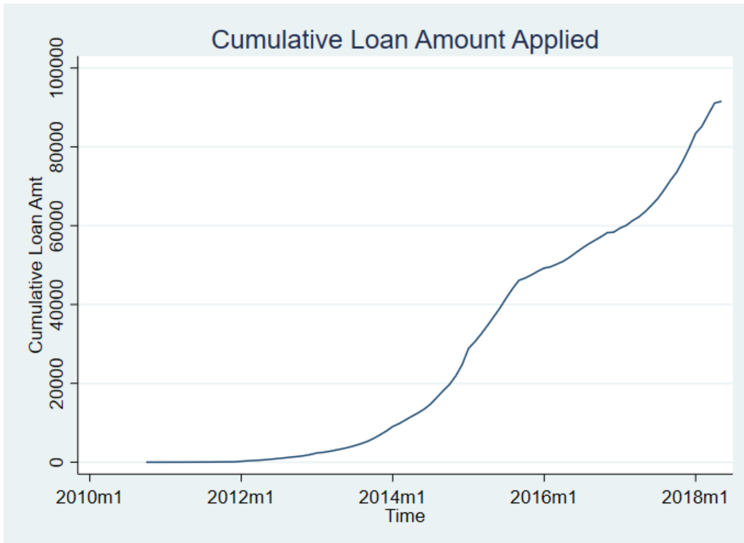

Fig. 1. Cumulative Loan Amount

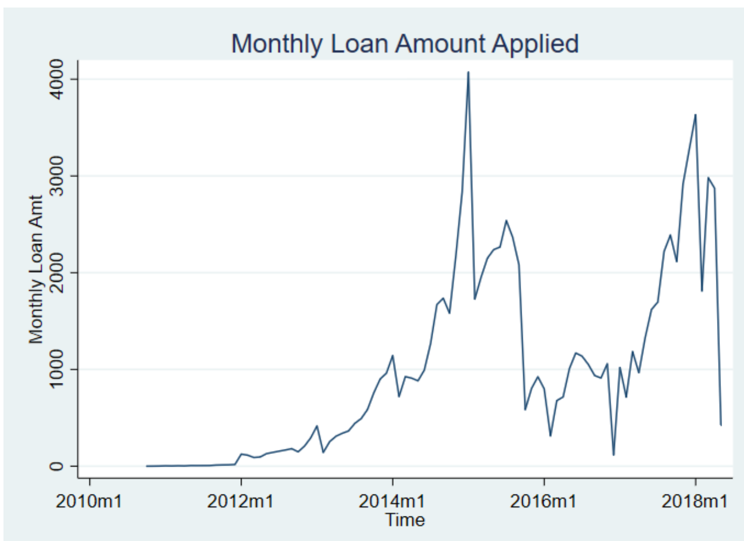

Fig. 2. Monthly Loan Amount

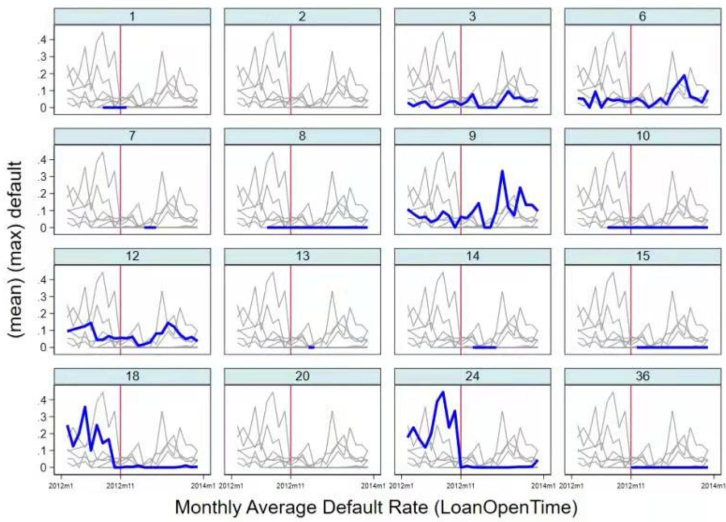

Fig. 3. Monthly Default Rates for Loans with Different Maturities

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.

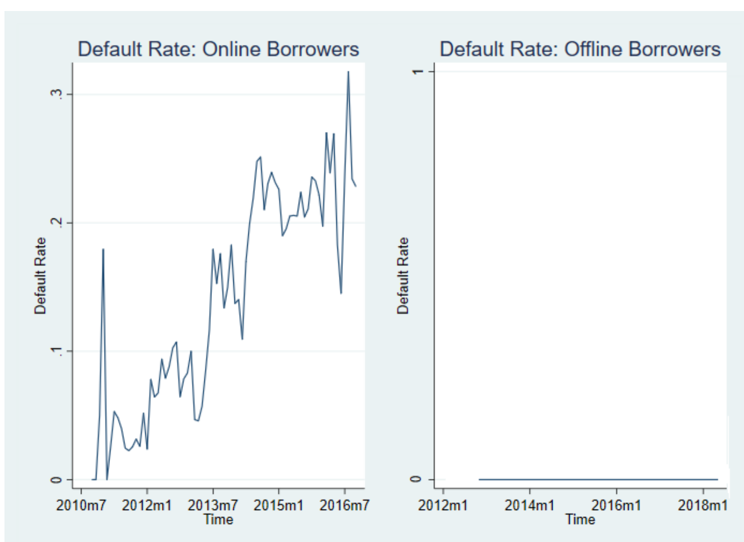

Fig. 4. Monthly Default Rates

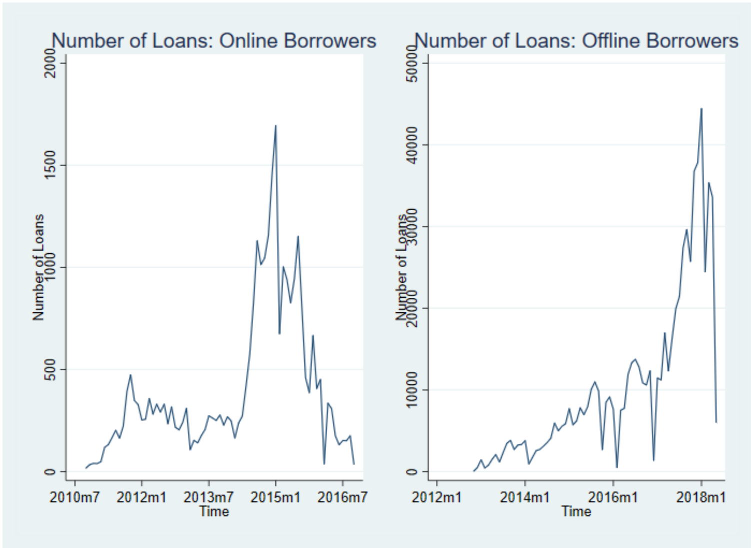

Fig. 5. Monthly Number of Loans

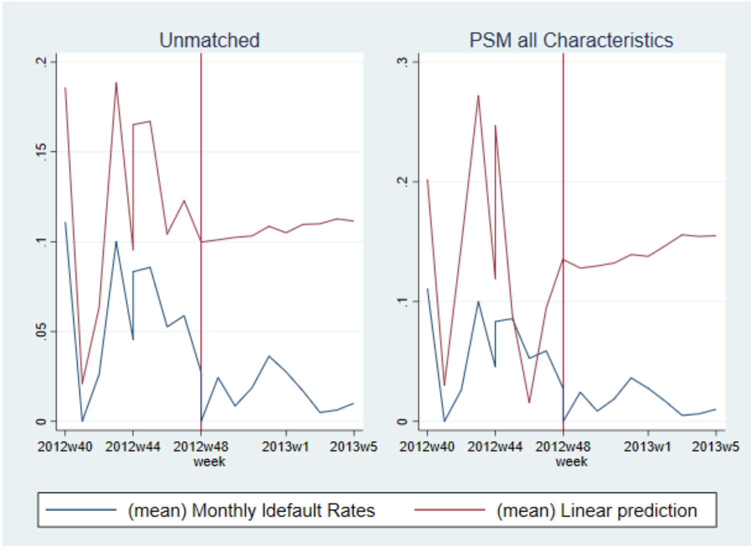

Fig. 6. Predicted v.s Published Default Rates, 2-month Window

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.

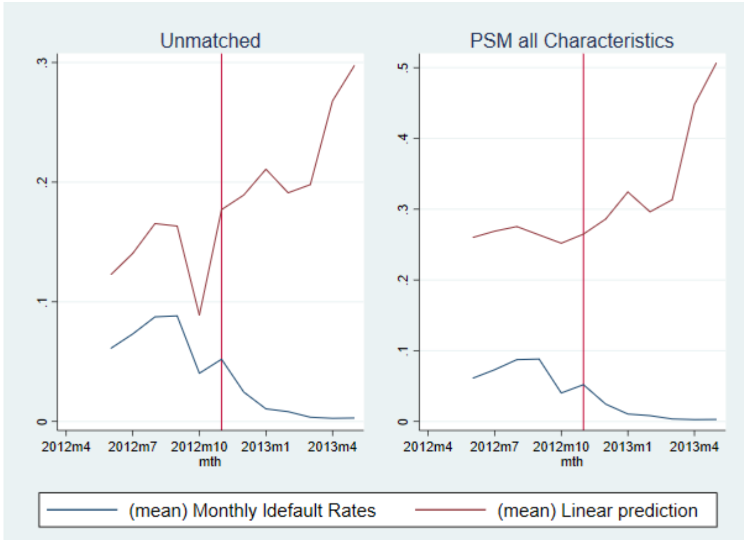

Fig. 7. Predicted v.s Published Default Rates, 6-month Window

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.

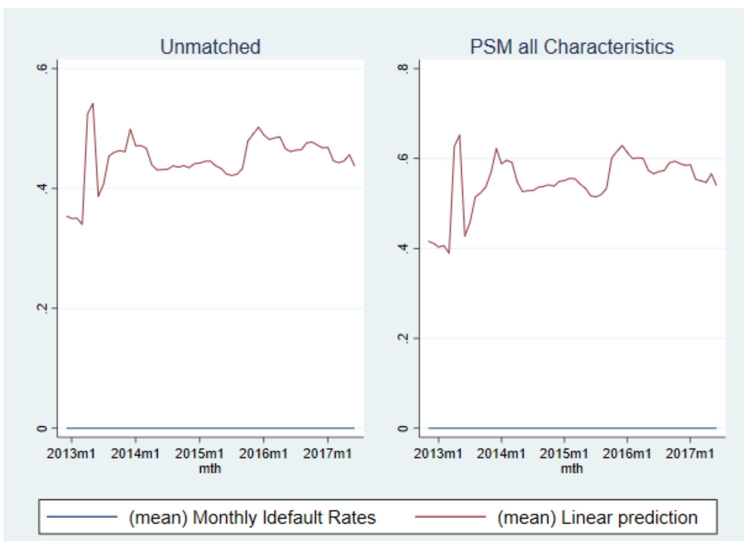

Fig. 8. Default Rates Predicted: Online v.s. Offline Borrowers

This figure reports the default rates predicted for offline borrowers using regression results on online borrowers after November 2012. The red line plots the predicted default rates for the offline borrowers using the regression results on online borrowers' data. The regression model is specified in 1. The blue line plots the reported zero default rates for offline borrowers.

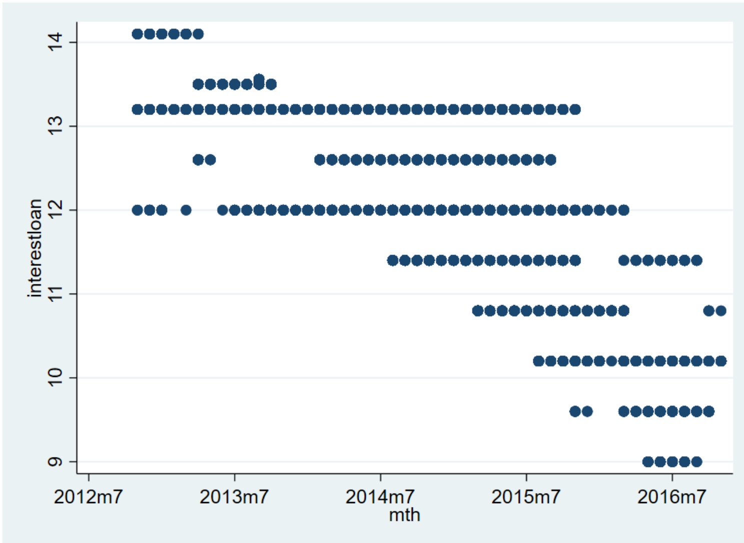

Fig. 9. Offline P2P Loans and Investor Confidence

This figure reports interest rates of P2P loans acquired offline. Since these offline P2P loans are 100% guaranteed, the only risk for investors is the platform run. Thus interest rates of offline P2P loans represent investor confidence.

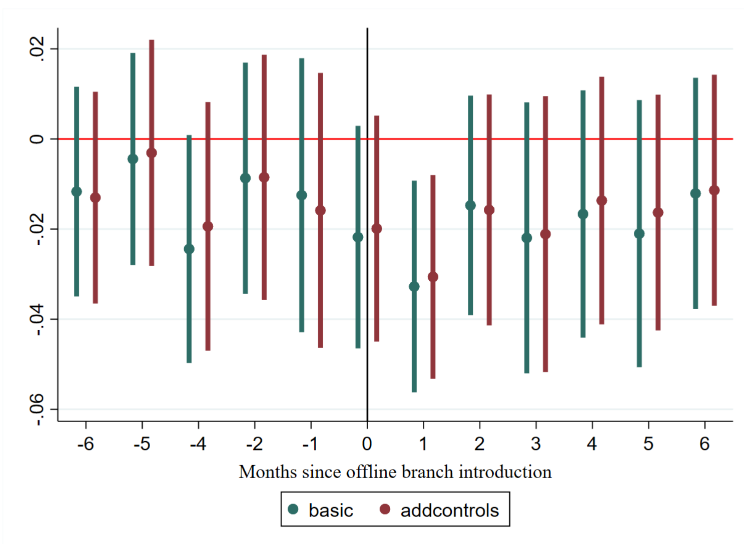

Fig. 10. Market Efficiency Test: Offline Branches' Establishment

Figure 10 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%.

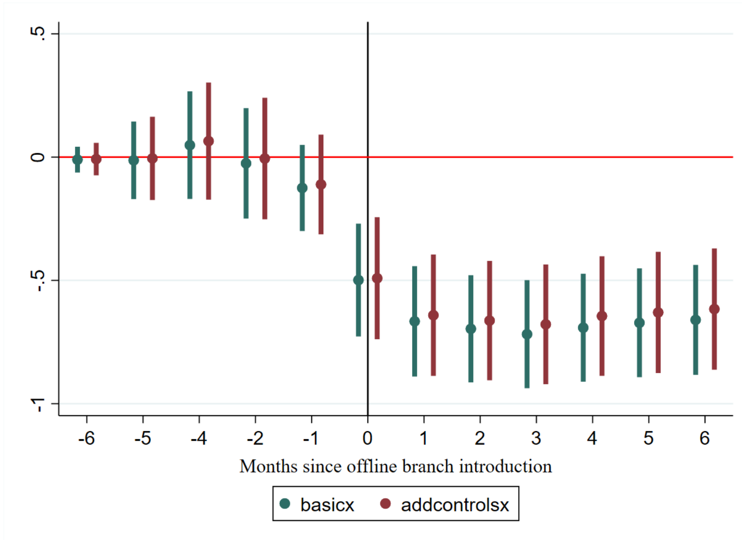

Fig. 11. “Credit” Type Loan Fraction: Offline Branches’ Establishment

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

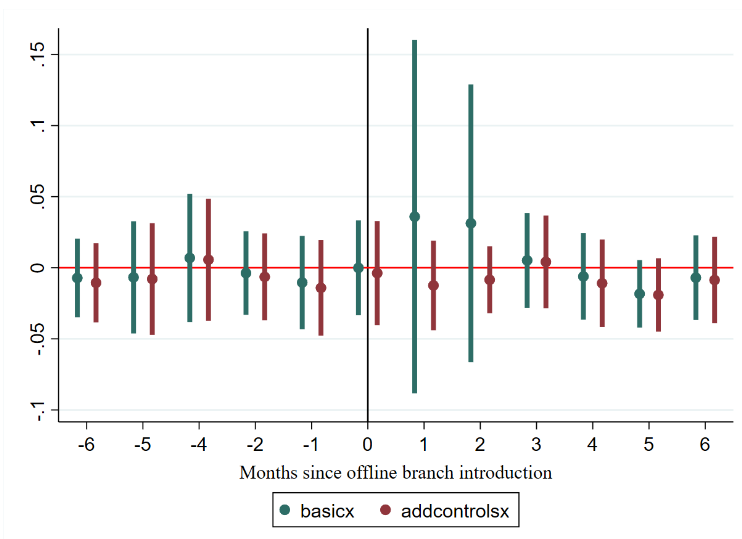

Fig. 12. The Number of “Credit” Type Loan: Offline Branches’ Establishment

Figure 12 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%.

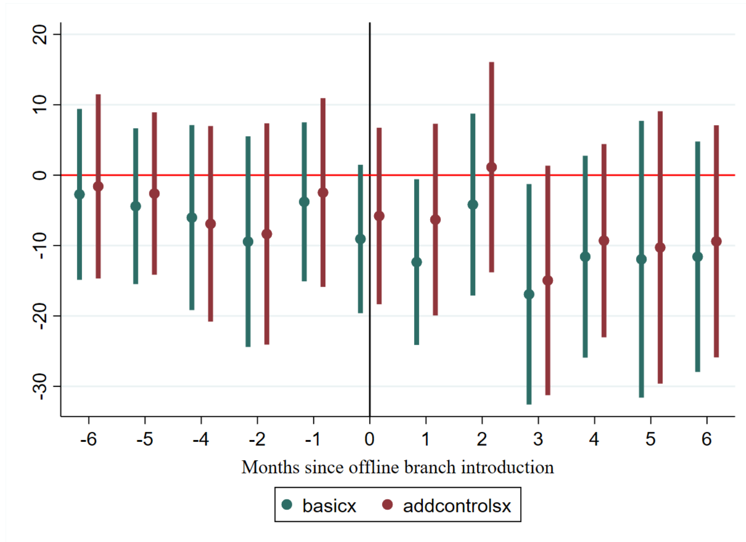

Fig. 13. Latest Credit Score: Offline Branches' Establishment

Figure 13 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%.

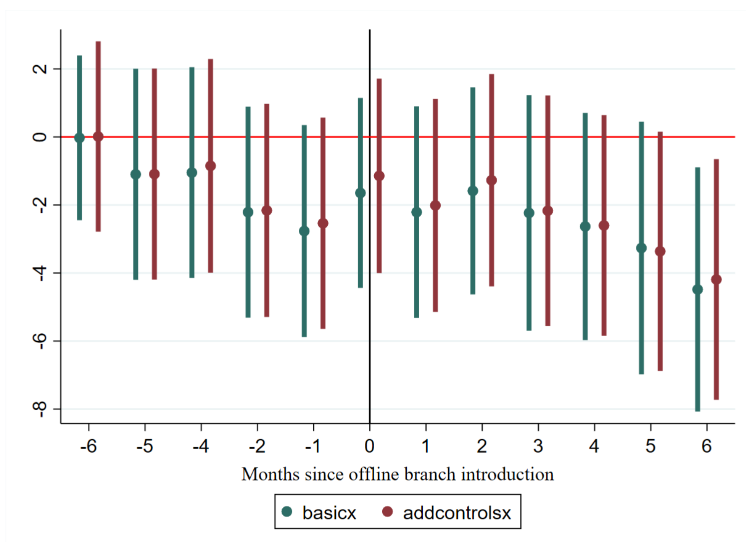

Fig. 14. Latest Credit Score: Offline Branches' Establishment, all applications

Figure 14 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 5%.

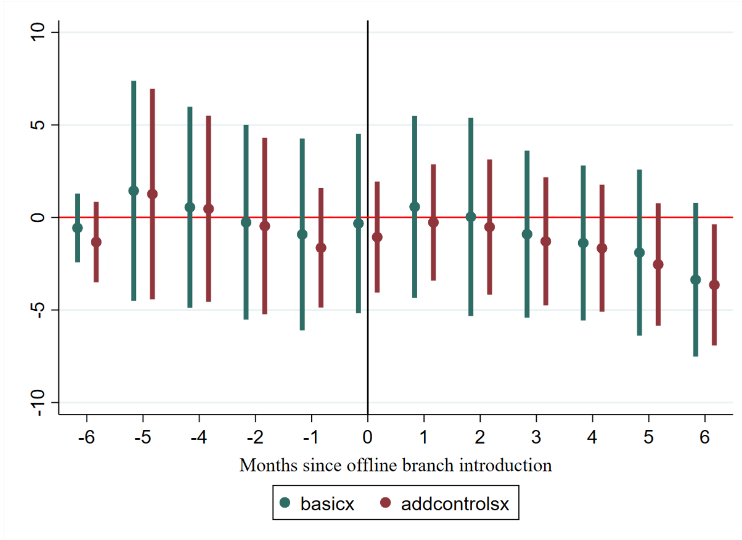

Fig. 15. Finishing Ratio: Offline Branches' Establishment, all applications

Figure 15 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 5%.

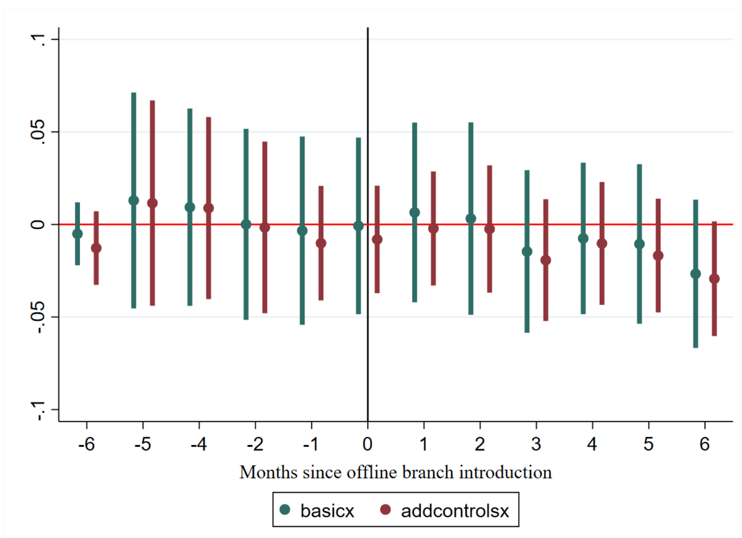

Fig. 16. Success Rate: Offline Branches' Establishment, all applications

Figure 15 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 5%.

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			
	(1)	(2)	(3)
VARIABLES	Before	After	PSM
Company Size	0.004	-0.001	0.039
	(0.019)	(0.003)	(0.023)
Marriage	-0.031	0.001	-0.085*
	(0.043)	(0.004)	(0.040)
Education	-0.039	-0.006	-0.028
	(0.026)	(0.005)	(0.029)
Work Experience	-0.017	0.001	0.046*
	(0.015)	(0.003)	(0.023)
Age	0.003	-0.000	0.003
	(0.002)	(0.000)	(0.004)
Gender	-0.020	-0.001	-0.003
	(0.032)	(0.003)	(0.065)
Property	0.019	0.038	0.039
	(0.029)	(0.034)	(0.066)
House Mortgage	0.026	-0.012	-0.007
	(0.045)	(0.042)	(0.053)
Income	0.011	0.009***	-0.016
	(0.017)	(0.002)	(0.016)
Car	-0.043	-0.023**	-0.024
	(0.035)	(0.008)	(0.023)
Car Mortgage	-0.041	0.008	0.033
	(0.022)	(0.042)	(0.031)
Log Loan Amount	-0.019	-0.014**	-0.037*
	(0.018)	(0.006)	(0.019)
Loan Maturity	0.007	-0.001	0.015
	(0.006)	(0.001)	(0.009)
Constant	0.209	0.173***	0.154
	(0.160)	(0.042)	(0.231)
Difference Test	0.091***	N/A	0.126***
	(0.004)	N/A	(0.006)
Week FE	Yes	Yes	Yes
Observations	245	2,179	2,179
R-squared	0.090	0.053	0.253

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). Standard errors are clustered at the month level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3: Regression Results: 6-month window

Regression Results Before and After the Event

VARIABLES	(1) Before	(2) After	(3) PSM
Company Size	0.016	-0.001	0.033
	(0.008)	(0.001)	(0.023)
Marriage	-0.004	-0.003	-0.006
	(0.018)	(0.002)	(0.047)
Education	-0.035***	-0.004***	-0.011
	(0.007)	(0.001)	(0.021)
Work Experience	-0.029**	0.001	-0.003
	(0.007)	(0.001)	(0.020)
Age	0.002	-0.000	-0.009
	(0.001)	(0.000)	(0.005)
Gender	-0.022*	-0.002	-0.021
	(0.009)	(0.001)	(0.065)
Property	-0.012	0.035*	-0.020
	(0.021)	(0.015)	(0.032)
House Mortgage	-0.012	-0.008	0.006
	(0.027)	(0.025)	(0.040)
Income	0.004	0.004*	0.022
	(0.009)	(0.002)	(0.014)
Car	-0.030**	-0.011	-0.094**
	(0.010)	(0.007)	(0.037)
Car Mortgage	0.012	0.031	-0.023
	(0.036)	(0.021)	(0.078)
Log Loan Amount	0.010	-0.004	0.017
	(0.010)	(0.004)	(0.019)
Loan Maturity	0.013***	-0.001*	0.025***
	(0.002)	(0.000)	(0.003)
Constant	-0.032	0.069	0.013
	(0.093)	(0.038)	(0.192)
Difference Test	0.217***	N/A	0.354***
	(0.021)	N/A	(0.040)
Month FE	Yes	Yes	Yes
Observations	983	7,059	7,059
R-squared	0.116	0.040	0.246

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 4: Online v.s. Offline Borrowers: Baseline Results

Regression Results for Online Versus Offline Borrowers

	(1)	(2)
VARIABLES	Online	PSM
comp_size	-0.013***	-0.023
	(0.003)	(0.021)
marriage	-0.013**	0.017
	(0.006)	(0.035)
education	-0.086***	-0.064***
	(0.003)	(0.017)
work_experience	-0.005	-0.009
	(0.003)	(0.012)
age	0.004***	0.004
	(0.001)	(0.004)
Gender	-0.029***	-0.022
	(0.008)	(0.027)
property	0.025**	0.000
	(0.010)	(0.041)
housing_mortgage	-0.067***	-0.022
	(0.009)	(0.034)
income	0.027***	0.044**
	(0.005)	(0.017)
car	-0.064***	-0.049
	(0.008)	(0.033)
car_mortgage	0.006	0.047
	(0.012)	(0.051)
l_amount	-0.018***	-0.041
	(0.005)	(0.036)
monthsloan	0.012***	0.017***
	(0.001)	(0.003)
Constant	0.288***	0.341
	(0.044)	(0.276)
Difference Test	0.449***	0.549***
	(0.005)	(0.008)
Week FE	Yes	Yes
Observations	19,978	280,498
R-squared	0.117	0.168

Robust standard errors in parentheses

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

The table reports the regression results for the specification in Equation (1). Column (1) reports the regression results on the online borrowers after November 2012. Column (2) shows the regression results on offline borrowers, and the default rates are matched from similar online borrowers using PSM. *Difference Test* reports the mean difference between the predicted default rates and the published default rates. Standard errors are clustered at the month level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: Regression for Offline Branches

	Offline Establishment	
	(1)	(2)
VARIABLES	DID	Pre-Window
Treat	0.097***	
	(0.026)	
Post	0.006***	
	(0.002)	
Relative Month	0.000	-0.001**
	(0.000)	(0.000)
Treat * Post	-0.101***	
	(0.025)	
Company Size	0.023***	0.016***
	(0.004)	(0.002)
Marriage	-0.004***	-0.004***
	(0.001)	(0.001)
Education	-0.006**	-0.007***
	(0.002)	(0.003)
Work Experience	0.022***	0.033***
	(0.003)	(0.005)
Age	-0.000***	-0.000***
	(0.000)	(0.000)
Gender	-0.001	-0.001
	(0.001)	(0.001)
Property	0.006*	0.002
	(0.003)	(0.004)
House Mmortgage	-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	-0.028***	-0.025***
	(0.002)	(0.003)
Loan Maturity	0.001*	0.001**
	(0.000)	(0.000)
Distance	0.002**	-0.006***
	(0.001)	(0.002)
Constant	0.229***	0.291***
	(0.024)	(0.048)
Loan Time FE	Yes	Yes
Branch Open Time FE	Yes	Yes
Borrower Living City FE	Yes	Yes
Observations	200,557	76,058
R-squared	0.152	0.179

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 6: 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 7: 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 8: 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 6. Panel 2 reports the pre-default recovery rate, post-default recovery rate, and estimated Loss Given Default (LGD) conditional on credit ratings.

Table 9: Market Efficiency Test

	Efficiency Test, All		Efficiency Test, Online		
	(1)	(2)	(3)	(4)	(5)
	default	default	default	default	default
interestar	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)
interestar × 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 10: 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 11: Market Efficiency Test, Offline Branches

	Efficiency Test, Online	
	(1) Loan Default	(2) 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

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: 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	0.068	0.024	0.055*	0.000	0.026
	(0.035)	(0.018)	(0.023)	(0.030)	(0.020)
marriage	0.020	0.030	0.000	0.036	-0.058
	(0.040)	(0.061)	(0.051)	(0.053)	(0.054)
education	-0.028	-0.007	-0.044	-0.025	-0.011
	(0.032)	(0.026)	(0.029)	(0.020)	(0.032)
work_experience	-0.002	-0.032	-0.019	0.009	-0.010
	(0.030)	(0.020)	(0.024)	(0.020)	(0.019)
age	-0.003	-0.007	-0.008	-0.008*	-0.005
	(0.005)	(0.005)	(0.006)	(0.004)	(0.005)
Gender	-0.048	-0.034	-0.064	-0.039	0.062
	(0.085)	(0.067)	(0.050)	(0.058)	(0.057)
property	-0.065	0.003	-0.044	-0.021	0.022
	(0.057)	(0.052)	(0.033)	(0.027)	(0.031)
housing_mortgage	-0.012	-0.006	0.024	0.022	-0.039
	(0.053)	(0.040)	(0.023)	(0.034)	(0.032)
income	0.016	0.018**	0.019	0.002	0.026*
	(0.020)	(0.006)	(0.012)	(0.013)	(0.011)
car	-0.106*	-0.144**	-0.160***	-0.134**	-0.015
	(0.046)	(0.043)	(0.038)	(0.043)	(0.051)
car_mortgage	-0.133	0.011	0.003	0.013	-0.140
	(0.087)	(0.096)	(0.091)	(0.053)	(0.077)
l_amount	0.012	0.033	0.017	0.021	-0.012
	(0.015)	(0.021)	(0.017)	(0.029)	(0.021)
monthsloan	0.024***	0.025***	0.027***	0.019**	0.023**
	(0.004)	(0.004)	(0.003)	(0.006)	(0.007)
Constant	-0.118	-0.108	0.105	0.132	0.075
	(0.157)	(0.217)	(0.209)	(0.118)	(0.091)
Difference Test	0.331***	0.367***	0.380***	0.284***	0.276***
	(0.040)	(0.042)	(0.045)	(0.033)	(0.034)
Month FE	Yes	Yes	Yes	Yes	Yes
Observations	6,878	7,059	7,059	7,059	7,059
R-squared	0.274	0.337	0.289	0.233	0.255

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 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 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 2

VARIABLES	(1) No Property	(2) No House Mortgage	(3) No Income	(4) No Car	(5) No Car Mortgage
comp_size	0.029	0.046*	0.050	0.060*	0.049*
	(0.018)	(0.020)	(0.029)	(0.024)	(0.022)
marriage	-0.030	0.030	-0.006	-0.032	-0.003
	(0.024)	(0.030)	(0.067)	(0.062)	(0.054)
education	-0.017	0.016	0.000	-0.028	-0.024**
	(0.024)	(0.036)	(0.026)	(0.030)	(0.009)
work_experience	-0.018	-0.012	-0.021	-0.010	0.000
	(0.023)	(0.029)	(0.031)	(0.034)	(0.016)
age	-0.005*	-0.009**	-0.004	-0.007	-0.008
	(0.002)	(0.004)	(0.006)	(0.005)	(0.004)
Gender	-0.019	0.031	0.043	0.082	0.023
	(0.060)	(0.104)	(0.055)	(0.063)	(0.042)
property	-0.033	-0.085*	-0.031	-0.031	0.016
	(0.040)	(0.038)	(0.052)	(0.051)	(0.041)
housing_mortgage	0.033	0.017	-0.032	0.080	0.021
	(0.036)	(0.048)	(0.036)	(0.048)	(0.034)
income	0.012	0.019	-0.004	-0.001	0.019
	(0.012)	(0.030)	(0.017)	(0.013)	(0.011)
car	-0.037	-0.016	-0.052	-0.052	-0.086**
	(0.051)	(0.073)	(0.036)	(0.049)	(0.026)
car_mortgage	0.043	-0.060	-0.052	-0.022	-0.072
	(0.062)	(0.053)	(0.078)	(0.033)	(0.064)
l_amount	0.024	0.005	0.042	0.014	0.012
	(0.032)	(0.031)	(0.027)	(0.024)	(0.020)
monthsloan	0.015***	0.016**	0.018**	0.021***	0.024***
	(0.003)	(0.005)	(0.005)	(0.003)	(0.005)
Constant	-0.018	0.087	-0.321	-0.027	-0.065
	(0.159)	(0.101)	(0.312)	(0.311)	(0.183)
Difference Test	0.284***	0.252***	0.308***	0.313***	0.314***
	(0.024)	(0.025)	(0.030)	(0.032)	(0.038)
Month FE	Yes	Yes	Yes	Yes	Yes
Observations	7,059	7,059	7,059	7,059	7,059
R-squared	0.138	0.144	0.192	0.251	0.288

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 May 2013). Standard errors are clustered at the month level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A4: Rosenbaum Bound Test

Gamma	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	0	0	0.1975	0.1975	0.1973	0.1976
1.5	0	0	0.1856	0.2085	0.1855	0.2087
2	0	0	0.1769	0.2159	0.1767	0.2160
2.5	0	0	0.1700	0.2213	0.1698	0.2214
3	0	0	0.1643	0.2255	0.1641	0.2257

The table reports the Rosenbaum Bound Test on the difference between default rates predicted by the PSM regression results in column (3) of Table 3 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 A5: Rosenbaum Bound Test

Gamma	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	0	0	0.406344	0.406344	0.406128	0.406561
1.5	0	0	0.389887	0.422732	0.389663	0.422953
2	0	0	0.378021	0.434387	0.377784	0.434618
2.5	0	0	0.368607	0.443431	0.368354	0.443671
3	0	0	0.360732	0.450803	0.360461	0.451053

The table reports the Rosenbaum Bound Test on the differences between default rates predicted by the PSM regression results in column (2) of Table 4 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: Online v.s. Offline Borrowers: PSM Results, Part 1

PSM Regression Results for Online Versus Offline Borrowers, Part 1

VARIABLES	(1) No Company Size	(2) No Education	(3) No Marriage	(4) No Work Experience	(5) No Age
comp_size	-0.033***	-0.027	-0.019	-0.024	-0.044
	(0.010)	(0.027)	(0.024)	(0.025)	(0.033)
marriage	-0.032	0.006	-0.024	-0.024	0.077
	(0.026)	(0.043)	(0.040)	(0.033)	(0.073)
education	-0.089***	-0.076***	-0.048**	-0.065***	-0.063**
	(0.012)	(0.021)	(0.018)	(0.016)	(0.028)
work_experience	0.011	0.015	0.001	0.010	-0.001
	(0.007)	(0.019)	(0.012)	(0.013)	(0.030)
age	0.005**	0.005*	0.006**	0.006**	0.006
	(0.002)	(0.003)	(0.002)	(0.002)	(0.005)
Gender	-0.012	-0.033	0.014	-0.004	-0.025
	(0.030)	(0.031)	(0.034)	(0.029)	(0.048)
property	0.030	-0.008	-0.015	-0.032	0.024
	(0.028)	(0.049)	(0.057)	(0.023)	(0.050)
housing_mortgage	-0.089***	-0.051	-0.002	-0.013	-0.112*
	(0.027)	(0.056)	(0.046)	(0.027)	(0.061)
income	0.034***	0.022*	0.053***	0.055***	0.068***
	(0.011)	(0.012)	(0.014)	(0.011)	(0.021)
car	-0.016	-0.081*	-0.005	0.005	-0.123**
	(0.029)	(0.043)	(0.027)	(0.031)	(0.048)
car_mortgage	-0.016	-0.033	0.020	-0.063	-0.042
	(0.026)	(0.066)	(0.041)	(0.046)	(0.079)
l_amount	-0.040**	-0.030	-0.057*	-0.063***	-0.045
	(0.018)	(0.022)	(0.031)	(0.018)	(0.038)
monthsloan	0.014***	0.015***	0.018***	0.018***	0.014***
	(0.001)	(0.003)	(0.002)	(0.002)	(0.004)
Constant	0.425***	0.338	0.264	0.373**	0.190
	(0.129)	(0.246)	(0.242)	(0.179)	(0.287)
Difference Test	0.486***	0.512***	0.525***	0.505***	0.457***
	(0.006)	(0.008)	(0.007)	(0.007)	(0.008)
Month FE	Yes	Yes	Yes	Yes	Yes
Observations	254,259	280,483	280,498	280,977	310,391
R-squared	0.162	0.192	0.186	0.184	0.292

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

This table reports the first half of the regression results on the offline and online borrowers 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 and published default rates after November 2012. Standard errors are clustered at the month level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A7: Online v.s. Offline Borrowers: PSM Results, Part 2

PSM Regression Results for Online Versus Offline Borrowers, Part 2					
	(1)	(2)	(3)	(4)	(5)
VARIABLES	No Property	No House Mortgage	No Income	No Car	No Car Mortgage
comp_size	-0.008	-0.028	-0.013	-0.010	-0.001
	(0.023)	(0.022)	(0.027)	(0.020)	(0.023)
marriage	0.021	0.017	-0.037	-0.017	-0.009
	(0.044)	(0.037)	(0.048)	(0.034)	(0.029)
education	-0.078***	-0.047***	-0.033*	-0.065***	-0.044**
	(0.018)	(0.013)	(0.018)	(0.013)	(0.018)
work_experience	0.013	-0.016	0.015	-0.002	-0.002
	(0.016)	(0.013)	(0.020)	(0.010)	(0.008)
age	0.001	0.006**	0.005	0.003	0.005
	(0.002)	(0.003)	(0.005)	(0.003)	(0.004)
Gender	-0.005	-0.001	-0.030	-0.023	-0.000
	(0.042)	(0.027)	(0.039)	(0.027)	(0.032)
property	0.031	-0.022	0.063	0.011	0.049
	(0.027)	(0.035)	(0.065)	(0.044)	(0.049)
housing_mortgage	-0.062*	-0.045	-0.060	-0.011	-0.062
	(0.036)	(0.034)	(0.059)	(0.046)	(0.042)
income	0.027	0.041***	0.036***	0.038**	0.060***
	(0.018)	(0.012)	(0.012)	(0.015)	(0.014)
car	-0.035	-0.046	-0.014	-0.004	-0.023
	(0.032)	(0.032)	(0.035)	(0.033)	(0.033)
car_mortgage	-0.004	-0.004	-0.063	-0.000	0.020
	(0.042)	(0.062)	(0.054)	(0.053)	(0.062)
l_amount	-0.014	-0.027	-0.059**	-0.056*	-0.063**
	(0.036)	(0.030)	(0.027)	(0.031)	(0.027)
monthsloan	0.015***	0.015***	0.015***	0.020***	0.018***
	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)
Constant	0.183	0.135	0.392	0.451	0.308
	(0.251)	(0.253)	(0.236)	(0.270)	(0.236)
Difference Test	0.526***	0.529***	0.478***	0.554***	0.527***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
Month FE	Yes	Yes	Yes	Yes	Yes
Observations	315,353	318,371	280,498	280,498	280,498
R-squared	0.178	0.186	0.184	0.180	0.186

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 offline and online borrowers 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 and published default rates after November 2012. Standard errors are clustered at the month level and displayed in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.