ADBI Working Paper Series

THE GENDER GAP IN PEER-TO-PEER LENDING: EVIDENCE FROM THE PEOPLE'S REPUBLIC OF CHINA

Xiao Chen, Bihong Huang,
and Dezhu Ye

No. 977
July 2019

Asian Development Bank Institute

Xiao Chen is a postdoctoral fellow at the School of Management and Economics, The Chinese University of Hong Kong, Shenzhen, People's Republic of China (PRC). Bihong Huang is a research fellow at the Asian Development Bank Institute in Tokyo, Japan. Dezhu Ye is a professor at the School of Economics of Jinan University, Guangzhou, PRC.

The Working Paper series is a continuation of the formerly named Discussion Paper series; the numbering of the papers continued without interruption or change. ADBI's working papers reflect initial ideas on a topic and are posted online for discussion. Some working papers may develop into other forms of publication.

Suggested citation:

Chen, X., B. Huang, and D. Ye. 2019. The Gender Gap in Peer-to-Peer Lending: Evidence from the People's Republic of China. ADBI Working Paper 977. Tokyo: Asian Development Bank Institute. Available: https://www.adb.org/publications/gender-gap-peer-peer-lending-evidence-prc

Please contact the authors for information about this paper.

Email: bihuang@adbi.org

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

Abstract

This paper documents and analyzes the gender gap in the online credit market. Using data from Renrendai, a leading peer-to-peer lending platform in the People's Republic of China (PRC), we show that lending to female borrowers is associated with better loan performance, including a lower probability of default, a higher expected profit, and a lower expected loss than for their male peers. However, despite the higher creditworthiness, we don't find any measurable gender impact on funding success rate, meaning that female borrowers have to compensate lenders by providing higher profitability to achieve a similar funding probability to their male peers. This evidence indicates the existence of a gender gap that discriminates against female borrowers. Further analysis implies that this gender gap is independent of the amount of information disclosed by borrowers.

Keywords: P2P lending, gender gap, loan performance

JEL Classification: G21, J16, G20

Contents

1. INTRODUCTION ... 1

2. LITERATURE REVIEW ... 3

2.1 Gender and Financial Constraints ... 3
2.2 Gender and Default ... 4

3. DATA AND SUMMARY STATISTICS ... 5

3.1 Data Source ... 5
3.2 Key Variables ... 6
3.3 Summary Statistics ... 8

4. METHODOLOGY AND EMPIRICAL RESULTS ... 13

4.1 Baseline Estimation Results ... 13
4.2 Addressing Endogeneity Concerns ... 19
4.3 Other Robustness Checks ... 27

5. CONCLUSION ... 27

REFERENCES ... 28

1. INTRODUCTION

Financial inclusion for women has been embraced by policymakers as an important development priority. However, despite the fact that females have lower risk preferences and higher creditworthiness, a gender gap in access to finance still remains prevalent in the traditional credit market due to various factors such as employment opportunities, legal obstacles, cultural norms, and limited access to the guarantee mechanism, among others (Alesina, Giuliano, and Nunn 2013a; Bellucci, Borisov, and Zazzaro 2010; Buvinic and Berger 1990; Eckel and Fullbrunn 2015; Moro, Wisniewski, and Mantovani 2017; Muravyev, Talavera, and Schafer 2009; Paglia and Harjoto 2014). According to the latest Global Findex report, more than 1 billion women still do not use or have access to the financial system (Demirguc-Kunt et al. 2018). With the advance of digital technology, online peer-to-peer (P2P) lending has emerged as an alternative to traditional lending institutions around the world. Bypassing banks, both lenders and borrowers are anonymous. Borrowers can post loan requests without providing collateral while the investors make lending decisions according to the information disclosed by the borrowers. This might help to moderate females' concern about potential discrimination. However, little is known about the gender gap in this new but rapidly growing fintech market. In this paper, we attempt to shed new light on whether decentralized lenders treat female and male borrowers differently on the P2P platform.

Identifying the gender gap in the credit market is not an easy task. The controversy begins with how to define and measure discriminatory lending behavior towards different segments of borrowers. Becker's seminal book (1957) introduced the first economic model of discrimination. In this model, there is a disamenity value to employing minority workers who may have to "compensate" employers by being more productive at a given wage level or, equivalently, by accepting a lower wage for identical productivity. Similarly, in the credit market, disadvantaged borrowers have to compensate lenders by providing a higher rate of return or, equivalently, discriminator pay or forgo income or profit for the privilege of exercising prejudicial tastes (Ladd 1998). Other well-known economists like Arrow (1971) and Phelps (1972) have developed models to enable an understanding of the profit-driven discriminatory behavior that may occur when lenders find it cheaper to exploit the characteristics of an applicant's group, such as its race, color, national origin, neighborhood, or gender, rather than the applicant's own past history, to evaluate the applicant's creditworthiness. Assuming that biased lenders would require higher expected profitability from loans granted to disadvantaged groups, these theories have implied that studies of loan performance are a way of inferring discriminatory lending. A better loan performance observed for the differentially treated groups, including a higher rate of return, lower default rates, and/or lower losses in the event of default, could be regarded as evidence of discrimination in the credit market (Berkovec et al. 1998).

These theoretical predictions form the basis of our empirical analysis. We investigate whether P2P loans extended to female borrowers perform better than loans granted to male borrowers, controlling for as many credit-risk factors as data availability allows. In particular, we assess the incidence of default, expected profit, and expected losses incurred in the event of default. Our empirical analysis of a large sample of transaction data obtained from Renrendai, a leading P2P lending platform in the PRC, reveals a significantly negative relationship between females and probability of default. Specifically, the marginal effect estimated by our logit model suggests that the default rate for female borrowers is 19.2 percentage points lower than that for their male counterparts. We also analyze gender differences in expected profit and expected loss and find that all else being equal, loans to female borrowers would enable investors to earn 2.04% more expected profit than loans to male borrowers. At the same time, lending to female borrowers reduces the expected loss by 0.14%. Despite the better loan performance, the funding success rate of female borrowers is not significantly different from that of males. These findings imply that female borrowers have to provide a higher rate of return to lenders to achieve a funding success rate comparable to their male peers.

In further analysis, we investigate whether the information availability changes the relationship between gender and credit constraint. Whenever the cost of acquiring evaluation of individual applicants is excessive (Phelps 1972), a financial institution will make decisions based on its previous statistical experience or prevailing sociological beliefs to maximize its expected utility. As borrowers disclose more information about themselves, the information asymmetry between the borrowers and the lenders will be alleviated. However, our empirical evidence indicates that the increase in the amount of information available to investors doesn't affect the probability of female borrowers obtaining loans via the P2P lending market. To examine whether the gender gap varies across different cohorts, we include the interaction terms between gender and other individual characteristics in our regression for funding success. We find that females who are married or who have longer working experience are more likely to raise funds through P2P lending platforms.

This paper contributes to the research on P2P lending and the gender gap in accessing financial services in the following aspects. First, this research enhances our understanding of the chances and challenges of using new financial technology to financially empower women to reduce the gender gap. Various types of bias in lending to borrowers have been detected in the P2P lending market. Using data from Prosper.com, the leading P2P lending platform in the US, Pope and Sydnor (2011) find evidence of significant racial disparities. Loan listings by blacks are less likely to be funded than those by whites with similar credit profiles while the interest rate paid by blacks is higher than that paid by comparable whites. Employing similar data, Ravina (2012) finds that good-looking applicants have a higher probability of getting loans, and pay lower interest rates, but have similar default rates to average-looking borrowers, indicating that there is an appearance discrimination in P2P lending. Duarte, Siegel, and Young (2012) show that borrowers appearing to be more trustworthy are more likely to have their borrowing requests funded. The empirical evidence provided by Lin and Viswanathan (2016) suggests that home bias is a robust phenomenon even in the context of a large online crowdfunding marketplace. Closely related to our study, Barasinska and Schafer (2014) provide evidence of the success of female borrowers on a large German peer-to-peer lending platform and conclude that there is no effect of gender on an individual borrower's chance of receiving funds. However, evidence on the gender gap in the P2P lending market is still scarce. This paper aims to fill this gap in the Chinese institutional context. The PRC has developed the biggest and fastest-growing market for online P2P lending. Despite its rapid expansion, research on the Chinese online credit market is still very limited. At the same time, the gender gap is still a critical issue in the PRC. The World Economic Forum (2017) placed the PRC 100th out of 144 countries and territories in its Global Gender Gap Index.[1] Rapidly spreading digital technologies offer an opportunity to provide financial services at much lower cost to populations usually excluded by traditional financial institutions like women.

Second, we contribute to the debate on the gender gap in financial markets by using comprehensive loan performance indicators, including default rate, expected profit, expected loss, and repayment ratio. Although Becker's (1971) theory has implied that

[1] http://reports.weforum.org/global-gender-gap-report-2017/.

the existence of discriminatory lending in the credit market will be inferred from loan performance, the joint estimation of the gender impact on the indicators of funding success rate, default rate, expected profitability, and expected loss is infeasible in most of the existing research due to data constraints (Barasinska and Schafer 2014; Blanchflower, Levine, and Zimmerman 2003; Moro, Wisniewski, and Mantovani 2017). The rich and unique data obtained from Renrendai allow us to fill this gap. To the best of our knowledge, this is the first example of the use of these indicators to establish evidence of a gender gap in the P2P credit market.

Third, we infer the causal effect of gender on P2P lending by innovatively addressing different types of endogeneity concerns. In our data, the number of male borrowers is more than six times that of female borrowers, implying that females are in the minority in the P2P lending market. Assuming that selection depends on observables, we use the propensity score matching (PSM) approach to mitigate the potential bias that might arise from this sample gap. Given that some unobservable or omitted variables may contaminate our estimation results, we implement the probit instrumental variable (IV) estimation where the gender of borrowers is instrumented by the gender ratio of the city they are living in. Finally, the Heckman Selection Model is adopted to moderate the sample selection bias arising from the fact that we can only observe the defaults among borrowers who have successfully got their loan listings funded and cannot observe defaults by those who fail to raise funds.

The rest of the paper is organized as follows. Section 2 reviews the literature on the gender gap in gaining access to finance; Section 3 describes the data and summary statistics; Section 4 presents our empirical strategies and findings; and Section 5 concludes this paper.

2. LITERATURE REVIEW

In this section, we review the relevant literature on studying the gender gap in financial markets.

2.1 Gender and Financial Constraints

The glaring gap between women and men in gaining access to finance has been repeatedly investigated, especially for women entrepreneurs, although much more remains to be done. Buvinic and Berger (1990) find that women continue to be more credit-rationed than men by microfinance institutions. Based on a survey covering firms in both Western and Eastern Europe, Muravyev, Talavera, and Schafer (2009) conclude that female-managed firms are less likely to obtain a bank loan than their male-managed counterparts. In addition, female entrepreneurs are charged higher interest rates when loan applications are approved. Bellucci, Borisov, and Zazzaro (2010) show that female entrepreneurs in Italy face tighter credit availability and higher collateral requirements, although they do not pay higher interest rates. Examining the contracts between banks and microfirms and self-employed individuals, Alesina, Giuliano, and Nunn (2013b) provide robust evidence that women in Italy pay more for overdraft facilities than men. Exploiting data provided by a Brazilian microfinance institute, Agier and Szafarz (2013) detect no gender bias in loan denial, but uncover a “glass ceiling” effect for women on loan size. Using the data from the World Bank’s Enterprise Survey (WBES), Asiedu, Freeman, and Nti-Addae (2013) find that female-owned firms in Sub-Saharan Africa are more likely to be financially constrained than male-owned firms. Analyzing individual-level survey data from the FinMark Trust (FinScope) for nine countries in Sub-Saharan Africa, Aterido, Beck, and Iacovone (2013) conclude that the lower use of formal financial services by women in nine Sub-Saharan Africa countries can be explained by gender differences in education, income levels, formal employment, and being the head of the household. Investigating the effects of private equity (PE) and venture capital (VC) financing on small and medium-sized single-entity business establishments, Paglia and Harjoto (2014) conclude that females are less likely to receive PE and VC financing.

Cultrual and social norms might explain the gender gap in economic activities. They may dictate whether activities like participating in the labor market, raising credit, and setting up an enterprise are seen as appropriate for women. If women are delineated as caregivers, they are more likely to suffer financial constraints. In addition, women are more averse to competition than men (Croson and Gneezy 2009). As the competition intensifies, the performance and participation of men improve relative to that of women (Gneezy, Leonard, and List 2009). The competition for loans is very tough on the P2P lending platforms as the overall funding success rate is very low. In such a highly competitive environment, it might become even more difficult for female applicants to raise funds.

However, studies on the gender-related prejudice in the credit market haven't reached consensus yet. Some literature suggests females may not suffer more often from financial constraints than their male peers. Analyzing credit applications and denial rates, loans outstanding, and interest rates across demographic groups, Cavalluzzo and Cavalluzzo (1998) conclude that white men and women can expect similar treatment in credit markets. Moreover, women with businesses located in the highly concentrated banking market are more successful in getting credit. Employing data from the 1998 and 2003 Survey of Small Business Finances (SSBF), Asiedu, Freeman, and Nti-Addae (2012) claim that firms led by white women did not face discrimination in terms of access to loans. They even paid a lower interest rate than firms led by white males in the US. Using the survey on enterprises within the European Union, Moro, Wisniewski, and Mantovani (2017) find no evidence that financial institutions discriminate against female managers.

According to Becker (1971), prejudicial discrimination raises a firm's costs, thereby reducing its competitiveness. The highly competitive industries like online lending should be less tolerant towards gender discrimination. Moreover, both lenders and borrowers are anonymous in the P2P lending market (Chen, Huang, and Ye 2018; Michels 2012). These facts might help to alleviate gender discrimination. Pope and Sydnor (2011) discover a higher funding success rate for female applicants in Prosper. Barasinska and Schafer (2014) argue that women would preferentially obtain loans from German P2P lending platforms where credit is highly marketable.

2.2 Gender and Default

A large amount of evidence has shown that women tend to be more risk-averse than men, especially when they are making financial decisions and investments (Eckel and Fullbrunn 2015; Huang and Kisgen 2013; Sundén and Surette 1998). For example, Powell and Ansic (1997) show that males and females adopt different strategies in making financial decisions, but overall females are less risk-seeking than males irrespective of familiarity and framing, costs, or ambiguity. Barber and Odean (2001) find that men trade 45% more excessively than women while men's net returns are lower than those of women, indicating that men tend to be more confident than women in terms of financial investment. Arch (1993) suggests that these differences can be explained by variations in the responses of males and females to situations perceived as risky because "males are more likely to see a challenge that calls forth participation while females tend to respond as to a threat in ways that encourage avoidance of the risk." Coates and Herbert (2008) illustrate that the endocrine system may account for financial risk taking since higher testosterone contributes to economic return whereas cortisol is increased by risk.

In addition to risk aversion, women are found to have higher moral standards. Investigating the data on 494 managers and seniors, Bernardi and Arnold (1997) conclude that female managers are at a significantly higher level of moral development than male managers. Cumming, Leung, and Rui (2015) suggest that women on boards effectively mitigate securities fraud because women are more ethically sensitive and less likely to risk committing fraud. Using a unique data set for a microbank in Albania over the period 1996 to 2006, Beck, Behr, and Guettler (2013) show that female loan officers experience significantly lower default rates than male loan officers due to their greater efforts in screening and monitoring loan quality and better skills in dealing with borrowers. The gender difference in default is attributed to the higher "moral responsibility" of females (Croson and Gneezy 2009; Hertzberg, Liberti, and Paravisini 2010). Moreover, females are more willing to plan for the future (Wiswall and Zafar 2018), because their loans are more relevant to children and families.

Taking all these arguments into consideration, we believe that female borrowers are better clients than their male peers in the credit market.

3. DATA AND SUMMARY STATISTICS

3.1 Data Source

The data used for this study come from all loan listings posted on Renrendai, a leading Chinese P2P lending platform, in the years between 2012 and 2014. Renrendai is one of the most popular peer-to-peer lending platforms in the PRC. Founded in 2010, it now has over 1 million members located in more than 200 cities. Moreover, the reputation of Renrendai has been well recognized in the PRC. In 2014 and 2015, it was awarded the status of an AAA (highest level) online lending platform by the Internet Society of China and the China Academy of Social Science. It ranked no. 53 in a list of the PRC's top 100 internet companies in 2015 released by the Internet Society of China and the Ministry of Industry and Information.

At Renrendai, borrowers post loan listings with the required information, including loan title, borrowing amount, interest rate, description of loan usage, and monthly installment. Renrendai provides verification services on national identification cards, credit reports, and addresses provided by borrowers. It assigns a credit score to each borrower according to his or her borrowing/lending history and the amount of verified information. Like Prosper.com, Renrendai's profit mainly comes from closing fees for borrowers and servicing fees for lenders. Since the verification and credit rating provided by Renrendai are limited, it is of critical importance for lenders to identify the trustworthiness of borrowers from the information disclosed on the platform (Iyer et al. 2016; Michels 2012). Renrendai requires all borrowers to provide an ID card that indicates gender, meaning that the disclosure of gender is compulsory. This requirement eliminates the self-selection bias and provides us with a compelling setting to study the gender gap. When creating loan listings, borrowers are encouraged to disclose additional information regarding the purpose of the loan and other personal information in a free-form text field called the "loan description." Once a loan listing is posted online, lenders may place bids by stating the amount they want to fund. With a minimum investment amount of RMB 50 yuan, a listing typically requires dozens of bids to become fully funded. A listing that achieves 100% funding status is a "successful" listing; otherwise, the borrower receives zero funding.

This study uses loan requests posted by borrowers from the PRC between 1 January 2012 and 31 December 2014. The original sample includes 371,508 listings. We eliminate 84,004 listings guaranteed by the platform because they are not typical P2P lending. We winsorize the loan listings whose AMOUNT and AGE are in the top or bottom 1 percentile to eliminate outliers. As a result, our sample includes 287,504 loan listings, of which 14,423 were successfully funded while the rest were not funded. We track the repayment performances of all successful loan listings. By the end of February 2017, there were 2,061 defaulted listings and 656 samples in progress of repayment.

3.2 Key Variables

To analyze the gender gap in P2P lending, we collect two categories of data from Renrendai. The first category is loan characteristics, including the term, interest rate, amount, and purpose of borrowing. The second category is the borrowers' characteristics, including the credit score, ownership of mortgage and car loans, age, education, income, marriage status, length of working experience, the industries that a borrower is working for, etc. The definition of each variable is summarized in Table 1.

In addition to the variables directly obtained from Renrendai, we construct the following variables to test the gender impact on loan performance.

Expected Profit

Assume that each loan is for $1, and if the borrower repays the loan, the lender receives $(1 + \mathrm{r})$, where r is the interest rate. This means that the lender earns a net profit of r if the borrower repays the loan, and loses the entire dollar if the borrower fails to repay the loan. If the default probability (DP) is δ, a lender's expected profit (PROFIT) on a loan listing is $E[\pi] = (1-\delta)r - \delta$. To get the DP, the likelihood that a borrower defaults, we first estimate the following equation using the probit model:

$$\mathrm{Pr}(DEFAULT_i) = \alpha_0 + \alpha_1 F\mathrm{emale}_i + \alpha_2 X_i + u_i \tag{1}$$

where the dependent variable indicates whether a loan listing *i* defaults after it is successfully funded. It equals 1 if the borrower defaults, and 0 otherwise. $F\mathrm{emale}_i$, the variable of main interest, represents the gender of a borrower, which equals 1 if a borrower is female, and 0 otherwise. X_i is a vector of control variables including loan characteristics, borrower characteristics, year effect, and regional effect. u_i refers to the error term. The coefficients estimated from Equation (1) are then used to predict the default probability of each loan listing. With the default probability and interest rate, we are able to measure the expected profit for each loan listing.

Expected Loss

Following the literature on credit risk management (Bessis 2015), we define the expected loss (EL) of a loan listing as the product of loss given default (LGD) and DP, i.e.

$$EL = LGD \times DP.$$

Table 1: Variables and Definitions

Variable	Name	Definition
Probability of Funding	*SUCCESS*	1 if a loan listing is fully funded, and 0 otherwise
Probability of Default	*DEFAULT*	1 if the funded loan has been defaulted, and 0 otherwise
Gender of Borrower	*Female*	1 if a borrower is female, and 0 otherwise
Expected Profit	*PROFIT*	Expected profit of a loan listing
Repayment Ratio	*RR*	Repayment ratio of a loan listing
Expected Loss	*EL*	Expected loss of a loan listing
Loan Amount (RMB)	*AMOUNT*	Loan amount requested by the borrower
Interest Rate (%)	*INTEREST*	The annual interest rate that a borrower pays on the loan
Loan Term (months)	*MONTHS*	Loan term requested by a borrower
Credit Score on Platform	*CREDIT*	Credit score of borrowers, taking values between 1 (high risk) and 7 (AA)
High Risk	*HighRisk*	1 if a borrower's *CREDIT* = HR, and 0 otherwise
Mortgage	*House_L*	1 if a borrower has a mortgage, and 0 otherwise
Car Loans	*Car_L*	1 if a borrower has a car loan, and 0 otherwise
Age	*AGE*	Age of a borrower (in years)
Education Level	*EDUCATION*	Education achievement of a borrower: 1=middle/high school; 2=college graduate; 3=university graduate; 4=postgraduate
Income (in RMB)	*INCOME*	Monthly income of a borrower: 1=less than 1,000; 2=1,001–2,000; 3=2,001–5,000; 4=5,001–10,000; 5=10,001–20,000; 6=20,001–50,000; 7=more than 50,000
Working Experience	*WORKTIME*	Borrowers' working experience: 1=less than 1 year; 2=1–3 years; 3=3–5 years; 4=more than 5 years
Marital Status	*Married*	1 if a borrower is married, and 0 otherwise
The Length of Title	*T_Length*	The number of Chinese characters in a loan title
The Length of Text	*D_Length*	The number of Chinese characters in a loan description
Number of loans Applied	*APPTIME*	The number of times that a borrower has posted a loan listing
Loan Purpose	*Purpose*	The purpose of borrowing described by the borrowers, including short-term turnover, personal consumption, car loans, housing loans, wedding planning, education and training, investment, medical expenditure, decorate, and others
Borrower's Industry	*Industry*	The industry that a borrower is working for, including IT, restaurants/hotels, real estate, public utilities, public welfare organizations, computer systems, construction, transportation, education/training, finance, law, retail/wholesale, media/advertising, energy, agriculture, other, sports/arts, entertainment, medical/sanitation/health care, government agencies and manufacturing
Borrower's Location	*Region*	The area in which a borrower is located
Education Disclosure	*Edu_Disclosure*	1 if education level is disclosed, and 0 otherwise
Work Experience Disclosure	*Worktime_Disclosure*	1 if the length of working experience is disclosed, and 0 otherwise
Income Disclosure	*Income_Disclosure*	1 if income is disclosed, and 0 otherwise
Marital Status Disclosure	*Marry_Disclosure*	1 if marital status is disclosed, and 0 otherwise
Location Disclosure	*Addr_Disclosure*	1 if location is disclosed, and 0 otherwise
Industry Disclosure	*Ind_Disclosure*	1 if the working industry is disclosed, and 0 otherwise
Information Disclosure	*DSCORE*	The total amount of information disclosed by a borrower

We define LGD as the fraction of the principal amount remaining if the borrower defaults at time *t*. According to the common practice applied at Renrendai, we assume that all loan listings are fully amortized. The borrower pays off the debt with a fixed monthly repayment schedule in equal installments so that the loan will be fully paid off at maturity. Hence, according to Hayre and Mohebbi (1992), LGD can be computed as follows:

$$LGD = 1 - \frac{(1 + r^m)^t - 1}{(1 + r^m)^n - 1}$$

where r^m is the monthly rate (i.e. the note rate divided by 12) and the loan term *n* is quoted in months. For the loan listings fully repaid at maturity, $t = n$, and hence $LGD = 0$. After computing LGD, we can get the repayment ratio (RR) for problematic loans as $RR = 1 - LGD$.

3.3 Summary Statistics

Table 2 presents the summary statistics (Panel A), the distribution of loan listings by funding status (Panel B), and the credit rating between male and female borrowers (Panel C). Panel A shows that only 5% of loan listings are successfully funded, among which 14.3% default. In terms of gender distribution, males account for 86% of all borrowers, playing a dominant role at Renrendai. This is true in most P2P online lending platforms in the PRC. For example, according to the annual report released by Paipaidai, another major P2P lending platform in the PRC, females accounted for only 18.63% of all its borrowers in 2015.[2] This is low compared to other countries. For example, the female participation rate is 33% at Prosper, one of the biggest P2P platforms in the United States (Duarte, Siegel, and Young 2012), and 27% at Smava, one of the biggest P2P platforms in Germany (Barasinska and Schafer 2014). The average annual funding cost is as high as 14.45%, while the average amount of loan requested is about RMB 66,000 (around USD 10,000). Among all borrowers, only 0.3% get a credit rating of "A" or above. Given the low credit rating of most borrowers, it is of critical importance to identify the trustworthy borrowers using observable signals. Among all the loan applicants, 10.6% have a mortgage and 4.3% have car loans.

For the borrowers' characteristics, we find that participants of the Chinese online financing market are young, well educated, with a medium income and limited working experience. The average age of borrowers is less than 32, and around 60% of them have a college education or above, much higher than the national average. In addition, 25.5% of borrowers earn a monthly income of more than 10,000 yuan. In line with their young age, the working experience of most borrowers is limited. Around 63.3% of borrowers have a working experience of less than three years and 44% of borrowers are married. In addition, 58.2% of borrowers are located in the eastern provinces whose GDP per capita is higher than that of the rest of the country. In terms of loan usage, 53% of loan listings are for "short-term turnover," followed by "startup business," which accounts for 12.4% of all loan listings. The borrowers work in diversified industries, with 19.1% employed in the manufacturing sector, 15.6% in retail/wholesale, and the rest in the other 18 industries. Following Michels' (2012) approach, we create a variable named DSCORE to measure the amount of information voluntarily disclosed by borrowers. We award a point for each piece of information disclosed. DSCORE is the sum of the points that a loan listing is awarded. Panel A shows that 98.3% of the borrowers choose to disclose their marital status, while 91% disclose their education level, and the mean value of

---

2 http://www.ppdaicdn.com/download/doc/ppdai2015.pdf.

DSCORE is 5.03, indicating that borrowers try to disclose as much information as possible to obtain loans.

Table 2: Summary Statistics, Distribution of Loan Listings and Risk Scores

Panel A	Summary Statistics				
Variable	**Mean**	**SD**	**Min**	**Max**	**N**
SUCCESS	0.05	0.218	0	1	287,504
DEFAULT	0.143	0.350	0	1	14,423
PROFIT	24.55	11.38	–26.03	169.76	208,957
RR	0.429	0.274	0	1	2,061
EL	–0.585	0.692	–14.399	2.001	208,957
Female	0.140	0.347	0	1	287,504
AMOUNT	66,000	99,000	3,000	500,000	287,504
INTEREST	14.45	3.350	6.1	24.40	287,504
MONTHS	15.72	10.17	1	36	287,504
CREDIT	1.091	0.520	1	7	287,504
Credit=HR	0.961	0.195	0	1	287,504
Credit=E	0.01	0.101	0	1	287,504
Credit=D	0.018	0.132	0	1	287,504
Credit=C	0.005	0.073	0	1	287,504
Credit=B	0.003	0.054	0	1	287,504
Credit=A	0.001	0.028	0	1	287,504
Credit=AA	0.002	0.048	0	1	287,504
AGE	31.98	6.55	23	53	282,724
Married	0.440	0.496	0	1	261,658
EDUCATION	1.818	0.792	1	4	261,658
Edu=High School	0.406	0.491	0	1	261,658
Edu=College graduate	0.383	0.486	0	1	261,658
Edu=University graduate	0.196	0.397	0	1	261,658
Edu=Postgraduate	0.014	0.119	0	1	223,601
WORKTIME	2.395	0.998	1	4	223,601
Worktime<=1year	0.175	0.380	0	1	223,601
Worktime=1~3year	0.458	0.498	0	1	223,601
Worktime=3~5year	0.164	0.370	0	1	223,601
Worktime>=5year	0.203	0.402	0	1	234,993
INCOME	3.970	1.206	1	7	234,993
Income<=1000	0.004	0.065	0	1	234,993
Income=1001~2000	0.024	0.152	0	1	234,993
Income=2001~5000	0.408	0.492	0	1	234,993
Income=5001~10000	0.309	0.462	0	1	234,993
Income=10001~20000	0.125	0.330	0	1	234,993
Income=20001~50000	0.076	0.265	0	1	234,993
Income>50000	0.054	0.226	0	1	287,504
HOUSE_L	0.106	0.308	0	1	287,504
CAR_L	0.043	0.202	0	1	287,504
T_Length	6.537	3.579	1	43	287,504
D_Length	46.20	38.45	1	500	223,584

continued on next page

Table 2 *continued*

Panel A	Summary Statistics				
Variable	**Mean**	**SD**	**Min**	**Max**	**N**
Region=East	0.582	0.493	0	1	223,584
Region=West	0.133	0.339	0	1	223,584
Region=Northeast	0.061	0.240	0	1	223,584
Region=Middle	0.224	0.417	0	1	223,584
Purpose=Short turnover	0.531	0.499	0	1	287,504
Purpose=Consumption	0.079	0.270	0	1	287,504
Purpose=Car purchase	0.053	0.223	0	1	287,504
Purpose=House purchase	0.027	0.162	0	1	287,504
Purpose=Wedding	0.017	0.129	0	1	287,504
Purpose=Education	0.009	0.094	0	1	287,504
Purpose=Other	0.037	0.188	0	1	287,504
Purpose=Startup business	0.124	0.330	0	1	287,504
Purpose=Medical	0.005	0.073	0	1	287,504
Purpose=Decoration	0.118	0.323	0	1	287,504
Ind=IT	0.076	0.265	0	1	219,519
Ind=Restaurant, hotel	0.034	0.182	0	1	219,519
Ind=Real estate	0.033	0.178	0	1	219,519
Ind=Public utilities	0.019	0.135	0	1	219,519
Ind=Nonprofit organization	0.002	0.041	0	1	219,519
Ind=Computer	0	0.012	0	1	219,519
Ind=Construction	0.056	0.230	0	1	219,519
Ind=Transportation	0.043	0.203	0	1	219,519
Ind=Education/training	0.036	0.186	0	1	219,519
Ind=Finance/Legal	0.043	0.202	0	1	219,519
Ind=Retail/wholesale	0.156	0.362	0	1	219,519
Ind=Media/advertising	0.031	0.173	0	1	219,519
Ind=Energy	0.039	0.195	0	1	219,519
Ind=Agriculture	0.021	0.144	0	1	219,519
Ind=Other	0.104	0.305	0	1	219,519
Ind=Sport/art	0.005	0.071	0	1	219,519
Ind=Medical	0.031	0.172	0	1	219,519
Ind=Entertainment	0.028	0.165	0	1	219,519
Ind=Government officer	0.053	0.224	0	1	219,519
Ind=Manufacturing	0.191	0.393	0	1	219,519
Edu_Disclosure	0.910	0.286	0	1	287,504
Worktime_Disclosure	0.778	0.416	0	1	287,504
Income_Disclosure	0.817	0.386	0	1	287,504
Ind_Disclosure	0.764	0.425	0	1	287,504
Addr_Disclosure	0.778	0.416	0	1	287,504
Marry_Disclosure	0.983	0.128	0	1	287,504
DSCORE	5.030	1.783	0	6	287,504
Year=2012	0.098	0.298	0	1	287,504
Year=2013	0.206	0.404	0	1	287,504
Year=2014	0.695	0.460	0	1	287,504

continued on next page

Table 2 *continued*

	Summary Statistics of Loan Status					
Panel B	**Loan Status**	**Obs.**	**Female**	**Proportion** (%)	**Male**	**Proportion** (%)
1	All	287,504	40,350	14.03	247,154	85.97
2	Funded	14,423	3,128	13.16	12,525	86.84
3	Not Funded	273,081	38,452	14.08	234,629	85.92
4	Loan Default	2,061	236	11.45	1,825	88.55
	Summary Statistics of Loan Status					
Panel C	**Credit**	**All Sample**	**Female**	**Ratio**	**Male**	**Ratio**
1	AA	657	74	0.18	583	0.23
2	A	222	9	0.02	213	0.08
3	B	840	116	0.28	724	0.29
4	C	1,553	178	0.44	1,375	0.55
5	D	5,114	661	1.63	4,453	1.8
6	E	2,960	401	0.99	2,559	1.03
7	HR	276,158	38,911	96.43	237,247	95.99
8	Total	287,504	40,350	100	247,154	100

Note: In this table, Panel A presents the summary statistics, Panel B reports the distribution of loan listings by funding status, and Panel C compares the credit rating between male and female borrowers.

Panel B indicates that female borrowers account for 14.03% of all loan listings, 13.16% of 14,423 successfully funded loan requests, 14.08% of failed requests, and 11.45% of defaults. Panel C shows the gender differences in the credit ratings. The overall credit ratings of borrowers are very low, with less than 3% of the borrowers receiving the top three ratings of AA, A, or B. The average credit rating of the female borrowers is slightly higher than that of the male borrowers.

Table 3 compares the descriptive statistics between male and female loan applicants and reports the mean difference test results. The funding success rate of male borrowers is 5%, 0.4 percentage points higher than that of female borrowers. In contrast, the default rate of female borrowers is 12.4%, 2.2 percentage points lower than that of male borrowers. In addition, except for working experience, other indicators also show significant differences between male and female borrowers. For example, the female loan applicants are slightly older and more educated but earn a lower income than their male peers. They tend to borrow more and for a longer term. Another point worth noting is that the interest rate for female borrowers at Renrendai is slightly lower than that for male borrowers (14.27% vs. 14.48%). In addition, we also find that the loan purposes for female borrowers are concentrated in the areas of "decoration," "house purchase," "education," and "medical expenditure," indicating that females borrow mainly for the needs of their families. In terms of employment, female borrowers are more likely to work in restaurants/hotels, retail/wholesale, education/training, finance/law, medical/sanitation/health care, and entertainment than their male peers. In terms of information disclosure, female borrowers are more likely to disclose their working experience and living places, although no gender difference is found in the disclosure of income level.

Table 3: Difference Test (Male vs. Female: Full Sample)

Variables	Male	Female	Mean Diff	Variables	Male	Female	Mean Diff
SUCCESS	0.051	0.047	0.004***	Region=West	0.130	0.147	–0.017***
DEFAULT	0.146	0.124	0.022**	Region=Northeast	0.058	0.082	–0.023***
PROFIT	24.254	26.44	–2.187***	Region=Middle	0.227	0.208	0.019***
RR	0.426	0.458	–0.033*	Ind=IT	0.08	0.051	0.028***
EL	–0.562	–0.734	0.172***	Ind=Restaurant/hotel	0.034	0.039	–0.006***
AMOUNT	65,885	68,413	–2,528.03***	Ind=Real estate	0.033	0.031	0.002*
INTEREST	14.48	14.27	0.214***	Ind=Public utilities	0.019	0.018	0.001
MONTHS	15.57	16.61	–1.041***	Ind=Commonweal	0.002	0.001	0.000*
CREDIT	1.093	1.080	0.014***	Ind=Computer	0	0	0
AGE	31.93	32.34	–0.413***	Ind=Construction	0.059	0.038	0.021***
Married	0.437	0.461	–0.025***	Ind=Transportation	0.046	0.027	0.018***
EDUCATION	1.805	1.899	–0.093***	Ind=Education/training	0.032	0.063	–0.031***
WORKTIME	2.395	2.394	0.001	Ind=Finance/Legal	0.041	0.053	–0.012***
INCOME	3.986	3.872	0.114***	Ind=Retail/wholesale	0.149	0.194	–0.045***
HOUSE_L	0.106	0.110	–0.004***	Ind=Media/advertising	0.031	0.032	–0.001
CAR_L	0.044	0.038	0.005***	Ind=Energy	0.042	0.023	0.018***
APPTIME	3.094	3.093	0.00100	Ind=Agriculture	0.022	0.013	0.010***
T_Length	6.550	6.456	0.094***	Ind=Other	0.098	0.140	–0.043***
D_Length	46.04	47.22	–1.178***	Ind=Sport/art	0.005	0.005	0
Purpose=Short turnover	0.531	0.529	0.002	Ind=Medical	0.027	0.054	–0.027***
Purpose=Consumption	0.081	0.067	0.014***	Ind=Entertainment	0.028	0.032	–0.004***
Purpose=Car purchase	0.054	0.045	0.009***	Ind=Government officer	0.054	0.046	0.008***
Purpose=House purchase	0.026	0.033	–0.007***	Ind=Manufacturing	0.200	0.139	0.061***
Purpose=Wedding	0.018	0.011	0.007***	Edu_Disclosure	0.911	0.904	0.007***
Purpose=Education	0.008	0.014	–0.006***	Worktime_Disclosure	0.777	0.784	–0.007***
Purpose=Other	0.036	0.041	–0.005***	Income_Disclosure	0.818	0.816	0.001
Purpose=Startup business	0.124	0.125	–0.001	Ind_Disclosure	0.766	0.746	0.020***
Purpose=Medical	0.005	0.007	–0.001***	Addr_Disclosure	0.777	0.784	–0.007***
Purpose=Decoration	0.117	0.129	–0.012***	Marry_Disclosure	0.984	0.980	0.004***
Region=East	0.585	0.564	0.021***	DSCORE	5.032	5.014	0.018*

Note: This table compares the descriptive statistics between male and female loan applicants and reports the means difference test results in the column of "Mean Diff." *, **, and *** indicate the statistical significance at the 10%, 5%, and 1% levels, respectively.

4. METHODOLOGY AND EMPIRICAL RESULTS

4.1 Baseline Estimation Results

4.1.1 Gender and Loan Performance

According to the implication of Becker's theory, the presence of differentiated performance of loans granted to different segments of borrowers is an important indicator of biased lending. If discriminatory lending exists, loans to minority borrowers should have lower expected rates of default, higher expected rates of return, or lower expected losses in the event of default than those to majorities with similar exogenous characteristics observed by lenders at the time of loan origination (Ladd 1998).

Following this theoretical implication, we first examine whether gender affects the default risk using Equation (1). If female borrowers are less risky than male borrowers, the coefficient on the gender (α_1) will be negative. Columns (1) and (2) of Table 4 present the baseline logit estimation results on the relationship between probability of default and the determinants of that probability. The analysis controls for the characteristics of loans and borrowers. To control for regional differences in economic conditions and time variations, region-level and time dummy variables are included in all estimations. We find that the coefficients of *Female* are negative at the 1% level of statistical significance, indicating that female borrowers are less likely to default. The corresponding marginal effect implies that female borrowers' probability of default is 19.2% (0.0269/0.149) lower than that of males.[3] In the interest of brevity, we briefly note only the role of control variables that have the most significant effects on the default likelihood. Higher likelihoods of default are associated with loans that have a larger amount of borrowing, a higher interest rate, and a longer term of maturity, as well as loans granted to people with a higher risk, less education, or less working experience.

In addition to the widely used indicator of default, for each loan listing we compute the expected profit, expected loss (EL), and repayment ratio (RR) in the event of default to illustrate the gender difference in loan performance from different perspectives. The OLS estimation results presented in Column (3) of Table 4 show that, all else being equal, lending to female borrowers is associated with an expected profit 2.042% higher than when lending to males. The difference test shown in Table 3 indicates that the interest rate offered by male borrowers is 0.214 percentage points higher than that by female borrowers. However, when the characteristics of loans and borrowers as well as time and regional effects are accounted for, lending to females is more profitable than lending to males. Columns (4) and (5) of Table 4 present the OLS estimation results on the two performance indicators of expected loss and repayment ratio. In the regression for expected loss, the coefficient on *Female* is negative at the 1% level of statistical significance, suggesting that lending to female borrowers is associated with an expected loss 0.140% lower than that for males. The coefficient on *Female* is positive but insignificant in the regression for repayment ratio. These findings suggest that lending to females is more profitable and less risky. The coefficients on other borrowers' characteristics indicate that loans to borrowers who are well educated would bring higher expected profit and incur lower expected loss for the lenders. In terms of loan characteristics, we find that loan listings with a larger requested amount, a longer term, and a higher risk are associated with lower expected profit and higher expected loss.

[3] 0.0269 is the marginal effect of Female in column (3). We don't show the marginal effects of all variables for the sake of brevity. They are available upon request.

We further assess the gender impact on the funding probability with the following equation:

$$\Pr(SUCCESS_i) = \alpha_0 + \alpha_1 Female_i + \alpha_2 X_i + \varepsilon_i \quad (2)$$

where the dependent variable is whether a loan listing *i* is successfully funded, equaling 1 if funded and 0 otherwise; X_i is a vector of control variables including loan characteristics, borrower characteristics, year effect, and regional effect; and ε_i is the error item. Columns (6) and (7) of Table 4 present the baseline logit estimation results. In Column (6), the coefficient on Female is -0.056, significant at the 10% statistical level. We further add personal indicators of marriage, education, working experience, and income into the regression and find that the coefficient of Female becomes insignificant in Column (7), indicating that gender has no measurable impact on funding success rates. This result is consistent with the findings of previous papers. For example, Agier and Szafarz (2013) demonstrate that no significant difference in the rejection rate between females and males is found in Brazilian microfinance. Barasinska and Schafer (2014) also show that gender cannot exert any influence on the funding success rate in the German P2P lending market. However, without investigating the gender disparity in loan performance, these researches cannot identify any potential gender prejudice against female borrowers. Our findings on gender differences in funding probability and loan performances together imply that females have to offer a higher profitability than males to achieve a similar funding rate on the P2P lending platform, indicating that female borrowers are treated differently from male borrowers by lenders.

Table 4: Gender and Loan Performance

	(1) DEFAULT	(2) DEFAULT	(3) Expected Profit	(4) Expected Loss	(5) Repayment Ratio	(6) SUCCESS	(7) SUCCESS
Female	–0.325***	–0.277***	2.042***	–0.140***	0.026	–0.056*	0.046
	(–3.99)	(–3.13)	(118.81)	(–52.99)	(1.27)	(–1.93)	(1.44)
lnAMOUNT	0.273***	0.072	–0.614***	0.008***	0.004	–0.528***	–0.810***
	(7.80)	(1.56)	(–101.36)	(9.07)	(0.43)	(–56.05)	(–69.48)
INTEREST	0.149***	0.159***	0.114***	0.071***	–0.005	–0.179***	–0.179***
	(11.00)	(10.85)	(60.56)	(247.12)	(–1.55)	(–42.23)	(–39.88)
MONTHS	0.030***	0.034***	–0.277***	0.021***	–0.003***	0.013***	0.015***
	(10.01)	(10.26)	(–383.71)	(184.74)	(–3.57)	(10.16)	(10.45)
HighRisk	3.309***	3.326***	–23.906***	0.344***	–0.121***	–3.269***	–2.831***
	(23.84)	(23.12)	(–831.08)	(77.73)	(–3.05)	(–128.01)	(–102.21)
AGE	0.029***	0.021***	–0.188***	0.013***	–0.002	0.064***	0.034***
	(7.24)	(4.24)	(–175.54)	(81.28)	(–1.46)	(45.21)	(18.56)
HOUSE_L	–0.495***	–0.469***	3.711***	–0.214***	0.054***	0.535***	0.137***
	(–6.91)	(–6.03)	(207.99)	(–78.01)	(3.02)	(20.06)	(4.85)
CAR_L	0.212**	0.026	–0.212***	0.024***	0.002	0.529***	0.174***
	(2.13)	(0.24)	(–8.05)	(5.94)	(0.07)	(13.15)	(4.12)
APPTIME	–0.138***	–0.178***	0.966***	–0.042***	–0.001	–0.001	–0.007***
	(–10.65)	(–12.08)	(754.30)	(–211.16)	(–0.59)	(–0.46)	(–4.52)
T_Length	0.013*	0.020**	–0.143***	0.012***	–0.000	0.038***	0.017***
	(1.75)	(2.52)	(–86.83)	(46.09)	(–0.05)	(14.16)	(5.85)
D_Length	0.001	0.001	–0.010***	0.001***	–0.000	0.004***	0.003***
	(1.54)	(1.15)	(–62.53)	(24.44)	(–1.43)	(16.20)	(11.62)
Married		–0.023	0.113***	0.015***	0.003		0.182***
		(–0.36)	(8.60)	(7.44)	(0.20)		(7.40)
EDUCATION		–0.464***	3.880***	–0.231***	0.027***		0.238***
		(–11.73)	(482.39)	(–187.03)	(3.01)		(16.86)
WORKTIME		–0.067**	0.602***	–0.024***	0.015**		0.301***
		(–2.03)	(89.77)	(–22.81)	(2.02)		(24.97)
INCOME		0.273***	–2.346***	0.161***	–0.005		0.468***
		(9.01)	(–384.10)	(171.02)	(–0.70)		(45.06)
_cons	–9.995***	–8.071***	59.272***	–2.838***	0.564***	4.636***	5.271***
	(–22.98)	(–15.47)	(764.81)	(–238.24)	(4.76)	(43.81)	(37.95)
Year	YES	YES	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES	YES	YES
N	14,423	13,148	208,957	208,957	1,963	287,504	208,957
r2_p/a	0.242	0.280	0.945	0.647	0.037	0.289	0.323

Note: (1) This table reports logit and OLS regression results on loan performance and funding success. The dependent variables are (i) Default, taking a value of 1 if the funded loan has been defaulted, and 0 otherwise; (ii) Success, taking a value of 1 if a loan listing is fully funded, and 0 otherwise; (iii) Expected Profit, Expected Loss, and Repayment Ratio are defined in Section 3.2 and calculated by the authors. The explanatory variables include: Female – the gender of a borrower, taking the value of 1 if a borrower is female, and 0 otherwise; lnAMOUNT – natural log of loan amount (in RMB) requested by the borrowers; INTEREST – the interest rate that the borrower pays on the loan; MONTHS – loan term (in months) requested by the borrower; HighRisk – a dummy variable taking a value of 1 if a borrower's credit score is HR; AGE – the age of a borrower expressed in years; HOUSE_L – a dummy variable taking a value of 1 if a borrower has a mortgage, and 0 otherwise; CAR_L – a dummy variable taking a value of 1 if a borrower has any car loan, and 0 otherwise; APPTIME – the number of times that a borrower has applied for a loan; T_Length – the number of characters in a loan title; D_Length – the number of characters in a loan description; Married – a dummy variable taking a value of 1 if a borrower is married, and 0 otherwise; EDUCATION – measuring the education level of a borrower; WORKTIME – a borrower's working experience measured in years; INCOME – the monthly income of a borrower; Year – a dummy controlling year effect; Region – a dummy variable reflecting the area in which a borrower is located; Industry – a dummy variable reflecting the industry that a borrower is working in; and Purpose – a dummy describing different purposes of borrowing.

(2) *, **, and ***indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Ordinary standard errors are used and Z-statistics are reported in parentheses. N is number of observations. r2_p is pseudo R-square. r2_a is adjusted R-square.

4.1.2 Gender and Information Disclosure

Females tend to face tougher constraints in the traditional lending market, including higher interest rates and a higher denial rate (Bellucci, Borisov, and Zazzaro 2010;

Muravyev, Talavera, and Schafer 2009). With regard to potential prejudice, females might provide more information than males to mitigate the gender gap. We first estimate the linkage between gender and information disclosure with OLS regression. The results in Column (1) of Table 5 indicate that the coefficient on Female is -0.021 and significant at the 5% level, suggesting that the amount of information disclosed by female applicants is less than that of their male peers. Given that the dependent variables are nonnegative integers, we follow Michels (2012) by using Poisson regression. The result shown in Column (2) of Table 5 is qualitatively similar to that for OLS regression. This might be due to the reality that women are more cautious than men (Cronqvist et al. 2016; Huang and Kisgen 2013) and therefore are more conservative in terms of information disclosure for reasons of privacy.

We then follow Bellucci, Borisov, and Zazzaro (2010) to augment the model on funding success by including an interaction term between the gender of the borrower and the amount of information disclosed by the borrower (*DSCORE*). The estimation result reported in Column (3) of Table 5 shows that the coefficient on the interaction of *Female* and *DSCORE* is insignificant. At the same time, in line with the baseline estimation results, the coefficient on *Female* is insignificant. This implies that the investors' attitude toward female borrowers doesn't change as the information availability increases, suggesting that the credit restriction on females is possibly due to noneconomic discrimination.

4.1.3 Gender Gap Across Different Cohorts

Using data from the American P2P lending platform Prosper, Pope and Sydnor (2011) find that the loan success rate of single women tends to be higher than that of single men whereas they pay lower interest rates than single men. Sundén and Surette (1998) show that single women are more risk-seeking than single men. To account for the heterogeneity across different cohorts among female borrowers, we separately add the interaction term between Female and other individual characteristic variables into the regression on loan success rate.

Table 5: Gender and Information Disclosure

OLS and Poisson Regression Result on DSCORE; Logit Regression Result on Funding Success

	(1) DSCORE	(2) DSCORE	(3) SUCCESS
Female	–0.021**	–0.004*	0.183
	(–2.42)	(–1.72)	(0.57)
Female_DSCORE			–0.038
			(–0.70)
DSCORE			0.712***
			(32.31)
lnAMOUNT	–0.082***	–0.017***	–0.551***
	(–28.87)	(–22.22)	(–56.25)
INTEREST	0.014***	0.003***	–0.186***
	(14.00)	(12.71)	(–43.73)
MONTHS	0.015***	0.003***	0.001
	(39.87)	(29.81)	(0.50)
HighRisk	–0.077***	–0.018***	–3.183***
	(–4.63)	(–4.33)	(–123.95)
AGE	0.009***	0.002***	0.064***
	(18.01)	(13.31)	(43.94)
HOUSE_L	0.744***	0.140***	0.346***
	(72.90)	(54.17)	(13.08)
CAR_L	0.676***	0.127***	0.424***
	(43.86)	(32.71)	(10.73)
APPTIME	0.080***	0.010***	–0.005***
	(108.39)	(76.89)	(–3.77)
T_Length	0.046***	0.009***	0.032***
	(51.97)	(39.14)	(11.86)
D_Length	0.004***	0.001***	0.003***
	(46.27)	(33.97)	(12.79)
_cons	4.120***	1.451***	1.419***
	(105.53)	(139.16)	(8.13)
Year	YES	YES	YES
Purpose	YES	YES	YES
N	287,504	287,504	287,504
r2_p		0.020	0.322
r2_a	0.146		

Note: (1) This table reports OLS and Poisson regression results on information disclosure and funding success. The dependent variables are (i) DSCORE, measuring the amount of information disclosed by the borrower; (ii) Success, taking a value of 1 if a loan listing is fully funded, and 0 otherwise. The explanatory variables include: Female – the gender of a borrower, taking the value of 1 if a borrower is female, and 0 otherwise; lnAMOUNT – natural log of loan amount (in RMB) requested by the borrowers; INTEREST – the interest rate that the borrower pays on the loan; MONTHS – loan term (in months) requested by the borrower; HighRisk – a dummy variable taking a value of 1 if a borrower's credit score is HR; AGE – the age of a borrower expressed in years; HOUSE_L – a dummy variable taking a value of 1 if a borrower has a mortgage, and 0 otherwise; CAR_L – a dummy variable taking a value of 1 if a borrower has any car loan, and 0 otherwise; APPTIME – the number of times that a borrower has applied for a loan; T_Length – the number of characters in a loan title; D_Length – the number of characters in a loan description; Year – a dummy controlling year effect; Region – a dummy variable reflecting the area in which a borrower is located; and Purpose – a dummy describing different purposes of borrowing.

(2) *,**, and ***indicate the statistical significance at the 10%, 5%, and 1% levels, respectively. Ordinary standard errors are used and T/Z statistics are reported in parentheses. N is number of observations. r2_p is pseudo R-square. r2_a is adjusted R-square.

Table 6: Logit Regression Result on Funding Success Probability Across Different Cohorts

.	(1) SUCCESS	(2) SUCCESS	(3) SUCCESS	(4) SUCCESS	(5) SUCCESS
Female	0.111	–0.075	0.130	–0.049	–0.175*
	(0.74)	(–1.40)	(1.53)	(–0.48)	(–1.91)
Female_AGE	–0.002				
	(–0.45)				
Female_Married		0.190***			
		(2.87)			
Female_EDUCATION			–0.041		
			(–1.07)		
Female_INCOME				0.023	
				(0.98)	
Female_WORKTIME					0.079***
					(2.59)
lnAMOUNT	–0.810***	–0.810***	–0.810***	–0.810***	–0.810***
	(–69.48)	(–69.48)	(–69.47)	(–69.48)	(–69.48)
INTEREST	–0.179***	–0.180***	–0.179***	–0.180***	–0.180***
	(–39.88)	(–39.92)	(–39.89)	(–39.89)	(–39.91)
MONTHS	0.015***	0.015***	0.015***	0.015***	0.015***
	(10.45)	(10.45)	(10.44)	(10.46)	(10.44)
HighRisk	–2.831***	–2.831***	–2.831***	–2.831***	–2.831***
	(–102.20)	(–102.18)	(–102.21)	(–102.21)	(–102.19)
AGE	0.034***	0.034***	0.034***	0.034***	0.034***
	(17.62)	(18.61)	(18.53)	(18.55)	(18.49)
HOUSE_L	0.138***	0.137***	0.137***	0.137***	0.138***
	(4.85)	(4.84)	(4.83)	(4.83)	(4.86)
CAR_L	0.174***	0.174***	0.174***	0.174***	0.173***
	(4.12)	(4.12)	(4.12)	(4.12)	(4.09)
APPTIME	–0.007***	–0.007***	–0.007***	–0.007***	–0.007***
	(–4.52)	(–4.45)	(–4.51)	(–4.49)	(–4.52)
T_Length	0.017***	0.017***	0.017***	0.017***	0.017***
	(5.86)	(5.84)	(5.85)	(5.86)	(5.85)
D_Length	0.003***	0.003***	0.003***	0.003***	0.003***
	(11.62)	(11.64)	(11.62)	(11.63)	(11.65)
Married	0.182***	0.157***	0.182***	0.183***	0.183***
	(7.37)	(6.00)	(7.38)	(7.42)	(7.44)
EDUCATION	0.238***	0.239***	0.244***	0.238***	0.239***
	(16.84)	(16.92)	(16.23)	(16.88)	(16.90)
WORKTIME	0.301***	0.301***	0.301***	0.301***	0.290***
	(24.97)	(24.96)	(24.97)	(24.97)	(22.85)
INCOME	0.468***	0.469***	0.468***	0.465***	0.468***
	(45.06)	(45.09)	(45.05)	(42.64)	(45.04)
_cons	5.261***	5.284***	5.262***	5.286***	5.306***
	(37.41)	(38.01)	(37.81)	(37.82)	(38.01)
Year	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES
N	208957	208957	208957	208957	208957
r2_p	0.323	0.323	0.323	0.323	0.323

Note: (1) This table reports logit regression results on the funding success probability across different cohorts. The dependent variable is Success, taking a value of 1 if a loan listing is fully funded, and 0 otherwise. The explanatory variables include: Female – the gender of a borrower, taking a value of 1 if a borrower is female, and 0 otherwise; lnAMOUNT – natural log of loan amount (in RMB) requested by the borrower; INTEREST – the interest rate that the borrower pays on the loan; MONTHS – loan term (in months) requested by the borrower; HighRisk – a dummy variable taking a value of 1 if a borrower's credit score is HR; AGE – the age of a borrower expressed in years; HOUSE_L – a dummy variable taking a value of 1 if a borrower has a mortgage, and 0 otherwise; CAR_L – a dummy variable taking a value of 1 if a borrower has any car loan, and 0 otherwise; APPTIME – the number of times that a borrower has applied for a loan; T_Length – the number of characters in a loan title; D_Length – the number of characters in a loan description; Married – a dummy variable taking a value of 1 if a borrower is married, and 0 otherwise; EDUCATION – measuring the education level of a borrower; WORKTIME – a borrower's working experience measured by years; INCOME – the monthly income of a borrower; Year – a dummy controlling year effect; Region – a dummy variable reflecting the area in which a borrower is located; Industry – a dummy variable reflecting the industry that a borrower is working in; and Purpose – a dummy describing different purposes of borrowing.

(2) *,**, and *** indicate the statistical significance at the 10%, 5%, and 1% levels, respectively. Ordinary standard errors are used and Z-statistics are reported in parentheses. N is number of observations. r2_p is pseudo R-square.

The estimation results are shown in Table 6 where the interaction terms of “Female and AGE,” “Female and Married,” “Female and EDUCATION,” “Female and INCOME,” and “Female and WORKTIME” are added into regressions, respectively. For all these interaction terms, we find that only the coefficients of Column (2) and Column (5) are significantly positive, while those of the other three interaction terms are not significant. Interestingly, the results of Column (2) show that married women have significantly higher funding success rates than single women, indicating that investors prefer to invest in married women. This might imply that marriage represents a signal of higher creditworthiness. By entitling a woman to a share of the aggregate household income, marriage has been found to decrease the overall risk of a woman’s asset position in her portfolio (Bertocchi, Brunetti, and Torricelli 2011). Moreover, marriage may also have a wealth growth mechanism (Fernandez and Wong 2014). Unlike a single person with only one source of income, married people have access to the family’s total income, which reduces the risk from income fluctuations. In line with our expectation, the results in Column (5) show that women who have longer working experience have significantly higher loan success rates than women who have just embarked on their career. By participating in social and economic activities for years, women will be able to accumulate rich and beneficial social relationships (Fang and Huang 2017). Therefore, a long work history is a positive signal for investors, resulting in higher loan success rates for women who are senior in their career.

4.2 Addressing Endogeneity Concerns

In evaluating the impact of gender on loan performance, there are a number of important methodological challenges that need to be addressed. First, the number of male borrowers is more than six times that of female borrowers in our sample, implying that females are in the minority in the P2P lending market. We implement the propensity score matching approach to mitigate the potential estimation bias that might arise from this gap. Second, some unobservable or omitted variables may contaminate our estimation results. For example, social networks and investor sentiment may change the funding success rate (Berger et al. 2013; Carnabuci and Dioszegi 2015; Lin and Viswanathan 2016) while the risk preference of borrowers and unexpected emergencies may affect the default rate and other loan performance indicators (Grinblatt, Keloharju, and Linnainmaa 2011; Shahriar 2016; Steinbuks 2015). We employ the probit instrumental variable (IV) model to address this concern. Third, as default depends on success, we can only observe the defaults among the borrowers who have successfully got their loan requests funded but cannot observe defaults by those who fail to raise funds. Hence our estimation on default might be susceptible to sample selection bias. The Heckman Selection Model is adopted to moderate this bias.

4.2.1 Propensity Score Matching Estimation

In our sample, the number of female borrowers is far outweighed by that of male borrowers. Table 2 shows that the percentage of female applicants in the whole sample is only 14%, that is, the proportion of male applicants is almost six times that of females. To solve this problem, ideally we would like to run an experiment with pairs of loan listings that are identical in all respects except gender. The observed difference in loan performance and funding success across all pairs would then be a robust estimate of the gender effect. While such an experiment is not feasible, we employ the propensity score matching (PSM) developed by Rosenbaum and Rubin (1983) and Heckman et al. (1998) to construct good matched samples based on observed characteristics. For each loan listing, we use a number of loan and borrower characteristics except gender to generate a propensity score. We then match each loan listing requested by a female with a set of loan listings requested by males based on the similarity of propensity scores. For each loan borrowed by a female, we adopt nearest neighbor matching to choose the n (n=1, 2, 3, 4, or 5) loans with the closest propensity scores, and then compare their arithmetical average of loan performance and funding success rate. Table 7 reports the results of mean difference in loan performance and funding success rate between females and males. We find that the default rates of female borrowers are significantly lower than those of males. At the same time, loans to female borrowers have a higher expected profit but lower expected loss than those to males. However, similarly to the logit regression result, female borrowers' loan success rates are not significantly different from those of males.

Table 7: PSM Estimation Results

Variables	Sample	Treated: Female=1	Control: Female=0	ATT
Panel A		**One-to-one matching**		
Success Rate	After match	0.0585	0.0578	0.0006
Default Rate	After match	0.1298	0.1582	–0.0283**
PROFIT	After match	26.4402	24.3053	2.1348***
EL	After match	–0.7341	–0.5872	–.1468***
Panel B		**One-to-two matching**		
Success Rate	After match	0.0585	0.0581	0.0004
Default Rate	After match	0.1298	0.157	–0.0268**
PROFIT	After match	26.4402	24.3567	2.0835***
EL	After match	–0.7341	–0.5891	–.1449***
Panel C		**One-to-three matching**		
Success Rate	After match	0.0585	0.0581	0.0004
Default Rate	After match	0.1298	0.0581	–0.0271***
PROFIT	After match	26.4402	24.3579	2.0823***
EL	After match	–0.7341	–0.5876	–.1464***
Panel D		**One-to-four matching**		
Success Rate	After match	0.0585	0.0577	0.0007
Default Rate	After match	0.1298	0.1525	–0.0227**
PROFIT	After match	26.4402	24.3311	2.1091***
EL	After match	–0.7341	–0.5885	–.1455***
Panel E		**One-to-five matching**		
Success Rate	After match	0.0585	0.0577	0.0007
Default Rate	After match	0.1298	0.153	–0.0231**
PROFIT	After match	26.4402	24.3323	2.1079***
EL	After match	–0.7341	–0.5891	–0.1449***

Note: (1) We use the nearest neighbor matching of 1:1, 1:2, 1:3, 1:4, and 1:5.

(2) The variables used for matching include: lnAMOUNT – natural log of loan amount (in RMB) requested by the borrowers; INTEREST – the interest rate that the borrower pays on the loan; MONTHS – loan term (in months) requested by the borrower; HighRisk – a dummy variable taking a value of 1 if a borrower's credit score is HR; AGE – the age of a borrower expressed in years; HOUSE_L – a dummy variable taking a value of 1 if a borrower has a mortgage, and 0 otherwise; CAR_L – a dummy variable taking a value of 1 if a borrower has any car loan, and 0 otherwise; APPTIME – the number of times that a borrower has applied for a loan; T_Length – the number of characters in a loan title; D_Length – the number of characters in a loan description; Married – a dummy variable taking a value of 1 if a borrower is married, and 0 otherwise; EDUCATION – measuring the education level of a borrower; WORKTIME – a borrower's working experience measured by years; INCOME – the monthly income of a borrower; Year – a dummy controlling year effect; Region – a dummy variable reflecting the area in which a borrower is located; Industry – a dummy variable reflecting the industry that a borrower is working in; and Purpose – a dummy describing different purposes of borrowing.

(3) The treatment group is female borrowers.

Table 8: Hidden Bias in Propensity Score Matching Results

Panel A	Success Rate: One-to-one Matching			
	Wilcoxon Statistics		Mantel-Haenszel Statistics	
Γ	Upper Bound	Lower Bound	Upper Bound	Lower Bound
1.0000	0.3745	0.3745	0.4682	0.4682
1.0500	0.8537	0.0453	0.0803	0.1139
1.1000	0.9909	0.0013	0.0039	0.0068
1.1500	0.9998	<0.0001	<0.0001	0.0001
1.2000	1.0000	<0.0001	<0.0001	<0.0001
1.2500	1.0000	<0.0001	<0.0001	<0.0001
1.3000	1.0000	<0.0001	<0.0001	<0.0001
1.3500	1.0000	0.0000	<0.0001	<0.0001
1.4000	1.0000	0.0000	0.0000	0.0000
1.4500	1.0000	0.0000	0.0000	0.0000
1.5000	1.0000	0.0000	0.0000	0.0000
1.5500	1.0000	0.0000	0.0000	0.0000
1.6000	1.0000	0.0000	0.0000	0.0000
1.6500	1.0000	0.0000	0.0000	0.0000
1.7000	1.0000	0.0000	0.0000	0.0000
1.7500	1.0000	0.0000	0.0000	0.0000
1.8000	1.0000	0.0000	0.0000	0.0000
1.8500	1.0000	0.0000	0.0000	0.0000
1.9000	1.0000	0.0000	0.0000	0.0000
1.9500	1.0000	0.0000	0.0000	0.0000
2.0000	1.0000	0.0000	0.0000	0.0000
Panel B	**Default Rate: One-to-one Matching**			
	Wilcoxon Statistics		**Mantel-Haenszel Statistics**	
Γ	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**
1.0000	0.0096	0.0096	0.0087	0.0087
1.0500	0.0022	0.0327	0.0021	0.0292
1.1000	0.0004	0.0857	0.0004	0.0762
1.1500	<0.0001	0.1802	<0.0001	0.1608
1.2000	<0.0001	0.3152	<0.0001	0.2843
1.2500	<0.0001	0.4739	<0.0001	0.4339
1.3000	<0.0001	0.6307	<0.0001	0.4521
1.3500	<0.0001	0.7637	<0.0001	0.3108
1.4000	<0.0001	0.8619	<0.0001	0.1968
1.4500	<0.0001	0.9260	<0.0001	0.1150
1.5000	<0.0001	0.9635	<0.0001	0.0622
1.5500	<0.0001	0.9833	<0.0001	0.0314
1.6000	<0.0001	0.9929	<0.0001	0.0148
1.6500	<0.0001	0.9972	<0.0001	0.0065
1.7000	<0.0001	0.9990	<0.0001	0.0027
1.7500	<0.0001	0.9996	<0.0001	0.0011
1.8000	0.0000	0.9999	0.0000	0.0004
1.8500	0.0000	0.9999	0.0000	0.0001
1.9000	0.0000	0.9999	0.0000	<0.0001
1.9500	0.0000	0.9999	0.0000	<0.0001
2.0000	0.0000	0.9999	0.0000	<0.0001

The identification of PSM estimators requires that the selection of treatment and nontreatment can be considered random after matching, meaning that selection bias is caused by observables rather than unobservables (Heckman et al. 1998; Zhao 2004). Any unobservable variables that simultaneously affect assignment to the treatment and outcome variables would result in a hidden bias. We implement two sensitivity analyses outlined by Mantel and Haenszel (1959) and Rosenbaum (2002) to estimate the extent to which the selection of unobservables may bias our inference on the gender gap in funding rate and default rate. Suppose we have a matched pair of loan listings *i* and *j*. If there is any hidden bias, two listings with the same observed covariates will have different chances of being funded or defaulted. In contrast, without hidden bias, *i* and *j* will have the same chances of being funded or defaulted, meaning their odds ratio (Γ) is equal to 1. Table 8 reports the Wilcoxon statistics and Mantel-Haenszel statistics. For the funding success rate, these two statistics are 0.3745 and 0.4682, respectively, for the upper bounds. For the default rate, these two statistics are 0.0096 and 0.0087, respectively. These results indicate that the PSM estimation on default is free of hidden bias while the funding success is not free of such bias.

4.2.2 Instrumental Variable Estimation

The second challenge of this study is that our estimation might be susceptible to the bias arising from unobservable variables. For example, the social networks of borrowers and investment sentiment are likely to impact the loan success rate (Berger et al. 2013; Carnabuci and Dioszegi 2015; Lin and Viswanathan 2016). The risk preference of borrowers and unexpected emergencies are important determinants of defaults (Grinblatt, Keloharju, and Linnainmaa 2011; Shahriar 2016; Steinbuks 2015). But all of them are unobservable. To address this concern, one potential solution is to find an instrumental variable (IV) that is correlated with the gender of the borrowers but does not directly affect the loan success rate or loan performance except through the gender.

We use the gender ratio (females to males) of the city where a borrower is living as an instrument to account for the probability of this borrower being female. The gender-related indicator of a city or state where a firm is located has been widely employed as an instrument to identify the impact of female presence in the top management on corporate performance. Based on the data from 1500 companies listed in the US S&P, Chen, Woon, and Marc (2017) adopt the ratio of the female labor force participation rate to the male labor force participation rate in a given state as an instrument for the number of corporate female directors to study the relationship between female directors and corporate dividend payment. The rationale behind this instrument is that firms in states where the female-to-male labor force participation ratio is higher are more likely to find good female candidates for their directorships. Jurkus, Park, and Woodard (2011) use the percentage increase in the female resident population between 1990 and 2000 in the state where a firm operates as the instrument for the number of female members on the board and investigate its impact on agency cost. Conyon and He (2017) choose the percentage of female residents in the US state where the given company has its headquarters as an instrument for the presence of women on the board to identify the impact of boardroom gender diversity on firm performance. In line with these researches, we construct an IV named LMFR, i.e. the gender ratio of each city and town. The data come from “China’s 2010 Census County Information.” The rationale for using this instrument is that the higher the female ratio in the population, the higher the probability of females participating in the P2P lending market. It is well-known in econometrics that the condition of exclusion restriction is vital for implementing IV regression. This implies that the gender ratio in a city will not directly affect the loan success rate and loan performance except through the gender of the borrowers at Renrendai. However, this assumption cannot be verified empirically because our model is exactly identified. Yet little evidence, if any, suggests that the gender ratio in a city would affect an individual borrower's funding success rate or loan performance.

Table 9 shows the estimation results. The first-stage regression result presented in Column (1) shows a positively significant coefficient on the instrumental variable (*LMFR*), meaning that the likelihood of a woman applying for a loan through P2P lending platforms is higher in places with a higher ratio of female residents. Moreover, the F-statistic shown at the bottom is 131.386. According to Staiger and Stock (1994), the suggested critical F-value is 8.96 when the number of instruments is one. With the F-statistic much greater than 10, we can reject the null hypothesis that the coefficient on the instrument is insignificantly different from zero at the 1% level, excluding the concern of a weak instrument. The second-stage regression results shown in Columns (2) to (5) are in line with the baseline estimations. The probability of default for female borrowers is significantly lower than that for male peers while no measurable gender effect is found for the probability of funding success. At the same time, lenders to female borrowers are expected to receive more profit and less loss if default happens. These findings further confirm that female borrowers are treated differently than male borrowers in the P2P lending market.

4.2.3 Heckman Selection Model

Another methodological challenge of this study is that default is dependent on success. We can only observe the defaults by borrowers who have successfully got their loan requests funded, but not defaults by those who fail to raise funds. Our estimation on the default might be contaminated by the sample selection bias. We employ Heckman's (1979) Selection Model to address this concern. We first estimate the choice model on the probability of funding success as follows:

$$Pr(SUCCESS_i) = \gamma_0 + \gamma_1 Z_i + \gamma_2 X_i + v_i \tag{5}$$

where X is a vector of control variables, and Z is an exogenous variable that could be excluded from the estimation on the default. The error terms in Equations (1) and (5), i.e. u and v respectively, are assumed to have a bivariate normal distribution with a mean zero and covariance matrix:[4]

$$\begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}.$$

[4] The bivariate normal distribution of the errors is an important assumption for the selection model. But again, this assumption cannot be empirically tested, as we cannot verify the exclusion restriction for the exactly identified IV model.

Table 9: IV Probit and 2SLS Estimation Results

	(1) Female	(2) DEFAULT	(3) Expected Profit	(4) Expected Loss	(5) SUCCESS
LMFR	0.036***				
	(6.51)				
Female		–1.790***	5.147***	–0.726***	–0.627
		(–4.33)	(3.96)	(–3.52)	(–0.79)
lnAMOUNT	0.008***	0.068***	–0.639***	0.013***	–0.416***
	(10.81)	(2.98)	(–50.99)	(6.66)	(–12.58)
INTEREST	–0.003***	0.065***	0.122***	0.070***	–0.088***
	(–11.05)	(5.36)	(30.60)	(110.20)	(–22.67)
MONTHS	0.001***	0.020***	–0.281***	0.021***	0.009***
	(12.39)	(8.17)	(–167.24)	(79.42)	(10.91)
HighRisk	0.010***	1.397***	–23.939***	0.350***	–1.519***
	(2.87)	(8.63)	(–707.96)	(65.12)	(–14.63)
AGE	0.001***	0.010***	–0.192***	0.014***	0.018***
	(9.97)	(3.60)	(–91.43)	(42.43)	(19.28)
HOUSE_L	–0.004*	–0.234***	3.723***	–0.216***	0.076***
	(–1.69)	(–5.57)	(187.63)	(–68.58)	(4.71)
CAR_L	–0.011***	0.034	–0.179***	0.018***	0.080***
	(–3.14)	(0.62)	(–5.68)	(3.56)	(3.19)
APPTIME	0.001***	–0.063***	0.964***	–0.041***	–0.003***
	(3.94)	(–7.52)	(595.96)	(–160.24)	(–2.92)
T_Length	–0.001***	0.007	–0.140***	0.011***	0.007***
	(–4.59)	(1.62)	(–64.41)	(32.14)	(3.87)
D_Length	0.000***	0.001*	–0.010***	0.001***	0.001***
	(6.14)	(1.87)	(–43.76)	(17.88)	(11.84)
Married	0.012***	0.019	0.078***	0.022***	0.095***
	(6.95)	(0.56)	(3.78)	(6.67)	(7.60)
EDUCATION	0.019***	–0.203***	3.820***	–0.220***	0.141***
	(18.98)	(–6.09)	(143.54)	(–52.00)	(13.90)
WORKTIME	–0.004***	–0.019	0.614***	–0.026***	0.147***
	(–4.52)	(–1.06)	(69.97)	(–18.49)	(10.79)
INCOME	–0.019***	0.098***	–2.287***	0.150***	0.227***
	(–24.18)	(3.74)	(–90.66)	(37.30)	(7.55)
_cons	–0.058**	–3.613***	58.933***	–2.774***	2.807***
	(–2.12)	(–7.05)	(358.99)	(–106.35)	(26.06)
Year	YES	YES	YES	YES	YES
Region	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Purpose	YES	YES	YES	YES	YES
N	208,957	13,148	208,957	208,957	208,957
F	131.386				

Note: (1) This table reports IV probit regression results on expected profit and expected loss, and 2SLS regression results on default and funding success probability. The dependent variables are (i) Female, taking a value of 1 if a borrower is female, and 0 otherwise; (ii) Default, taking a value of 1 if the funded loan has been defaulted, and 0 otherwise; (iii) Expected Profit, Expected Loss, and Repayment Ratio are defined in Section 3.2 and calculated by the authors. The explanatory variables include: Female – the gender of a borrower, taking a value of 1 if a borrower is female, and 0 otherwise; lnAMOUNT – natural log of loan amount (in RMB) requested by the borrower; INTEREST – the interest rate that the borrower pays on the loan; MONTHS – loan term (in months) requested by the borrower; HighRisk – a dummy variable taking a value of 1 if a borrower's credit score is HR; AGE – the age of a borrower expressed in years; HOUSE_L – a dummy variable taking a value of 1 if a borrower has a mortgage, and 0 otherwise; CAR_L – a dummy variable taking a value of 1 if a borrower has any car loan, and 0 otherwise; APPTIME – the number of times that a borrower has applied for a loan; T_Length – the number of characters in a loan title; D_Length – the number of characters in a loan description; Married – a dummy variable taking a value of 1 if a borrower is married, and 0 otherwise; EDUCATION – measuring the education level of a borrower; WORKTIME – a borrower's working experience measured by years; INCOME – the monthly income of a borrower; Year – a dummy controlling year effect; Region – a dummy variable reflecting the area in which a borrower is located; Industry – a dummy variable reflecting the industry that a borrower is working in; and Purpose – a dummy describing different purposes of borrowing.

(2) *,**, and ***indicate the statistical significance at the 10%, 5%, and 1% levels, respectively. Ordinary standard errors are used and T/Z-statistics are reported in parentheses. N is number of observations.

If the error terms u and v are correlated, i.e. $\rho \neq 0$, then the estimation without controlling sample selection will be biased. Heckman proposes controlling for this bias by estimating from Equation (5) the ratio of the probability density function to the cumulative distribution function of a distribution, or the inverse Mills ratio (IMR):

$$IMR = \varphi(\hat{\alpha}_0' Z + \hat{\alpha}_1' X)/\Phi(\hat{\alpha}_0' Z + \hat{\alpha}_1' X) \quad (6)$$

where $\varphi(.)$ and $\Phi(.)$ are the normal density and cumulative distribution functions, respectively. Adding the IMR to Equation (1) can control the selection bias. The estimation of default hence becomes:

$$Pr(DEFAULT_i) = \theta_0 + \theta_1 Female_i + \theta_2 IMR_i + \theta_3 Control_i + \epsilon_i. \quad (7)$$

A convincing implementation of the Heckman Selection Model is to identify from the first-stage choice model at least one exogenous independent variable (Z) that can be validly excluded from the vector of explanatory variables in the second-stage regression (Lennox, Francis, and Wang 2012; Little 1985; Tourani-Rad, Gilbert, and Chen 2016). We leverage the peer effect for identification.[5] The important role of peers in forming financial decisions has been well recognized in finance literature. For example, Leary and Roberts (2014) acknowledge that firms' financing decisions are in large part responses to the financing decisions of peer firms. To identify the managerial traits in corporate choices, Faccio, Marchica, and Mura (2016) use the fraction of firms with a female CEO as an instrument for a firm to have a female CEO. Moreover, Grinblatt and Keloharju (2001), Huberman (2001), and Seasholes and Zhu (2010), among others, document that familiarity appears to be important to investors in an investment setting. We borrow from these studies and develop an instrument named Me_SUCCESS for the model identification. It is the average loan success rate of borrowers with a similar educational level, monthly income, and length of working experience. We believe that the loan success rate of peers with similar characteristics will affect the funding probability of an individual borrower, but not this borrower's probability of default.

The estimation results are shown in Table 10. Column (1) reports the first-step estimation on SUCCESS. The coefficient on Me_SUCCESS is positively significant, implying that the higher the funding success rate of peers, the higher the likelihood of a borrower getting a loan application funded. Column (2) presents the endogeneity-adjusted estimate on default where the inverse Mills ratio (IMR) estimated by the first stage is added. The coefficient on the IMR is significant at the 1% level, indicating the existence of sample selection bias and the need to use the Heckman Selection Model. The coefficient on Female is –0.261, which is significant and similar in size to the baseline estimation, meaning that our conclusions are robust after controlling for sample selection bias. The variance inflation factor (VIF) shown at the bottom of Column (1) is 5.68, indicating that our estimation is free of the risks arising from multicollinearity (Belsley, Kuh, and Welsch 1980; Greene 2003).

[5] According to the definition given by Wikipedia, peer effect or peer pressure is the direct influence on people by peers, or the effect on an individual who gets encouraged to follow their peers by changing their attitudes, values, or behaviors to conform to those of the influencing group or individual. See the link: https://en.wikipedia.org/wiki/Peer_pressure.

Table 10: Heckman Two-Step Regression on the Probability of Default

	(1) SUCCESS	(2) DEFAULT
Female	0.009	–0.261***
	(0.58)	(–2.95)
Me_SUCCESS	3.814***	
	(16.90)	
IMR		1.144***
		(3.31)
lnAMOUNT	–0.441***	–0.318**
	(–73.50)	(–2.52)
INTEREST	–0.088***	0.074**
	(–40.48)	(2.53)
MONTHS	0.008***	0.042***
	(12.28)	(10.30)
HighRisk	–1.549***	2.036***
	(–99.70)	(4.93)
AGE	0.017***	0.037***
	(17.70)	(5.34)
HOUSE_L	0.088***	–0.395***
	(6.12)	(–4.89)
CAR_L	0.097***	0.115
	(4.47)	(1.02)
APPTIME	–0.003***	–0.182***
	(–3.87)	(–12.33)
T_Length	0.008***	0.028***
	(5.80)	(3.38)
D_Length	0.001***	0.002***
	(11.11)	(2.74)
Married	0.095***	0.060
	(7.81)	(0.85)
EDUCATION	0.060***	–0.345***
	(7.29)	(–6.45)
WORKTIME	0.048***	0.084
	(5.58)	(1.49)
INCOME	0.157***	0.509***
	(21.04)	(6.58)
_cons	3.345***	–6.604***
	(43.20)	(–9.86)
Year	YES	YES
Region	YES	YES
Industry	YES	YES
Purpose	YES	YES
VIF (Me_SUCCESS)	5.68	
N	208,957	208,957

Note: (1) This table reports Heckman two-step regression results on the probability of default. In Column (1), the dependent variable is SUCCESS, taking a value of 1 if a loan listing is fully funded, and 0 otherwise. In Column (2), the dependent variable is DEFAULT dummy, taking a value of 1 if the funded loan has been defaulted, and 0 otherwise. IMR is the inverse Mills ratio. Me_SUCCESS is the average funding success rate of a borrower's peers. Other explanatory variables include: Female – the gender of a borrower, taking a value of 1 if a borrower is female, and 0 otherwise; lnAMOUNT – natural log of loan amount (in RMB) requested by the borrower; INTEREST – the interest rate that the borrower pays on the loan; MONTHS – loan term (in months) requested by the borrower; HighRisk – a dummy variable taking a value of 1 if a borrower's credit score is HR; AGE – the age of a borrower expressed in years; HOUSE_L – a dummy variable taking a value of 1 if a borrower has a mortgage, and 0 otherwise; CAR_L – a dummy variable taking a value of 1 if a borrower has any car loan, and 0 otherwise; APPTIME – the number of times that a borrower has applied for a loan; T_Length – the number of characters in a loan title; D_Length – the number of characters in a loan description; Married – a dummy variable taking a value of 1 if a borrower is married, and 0 otherwise; EDUCATION – measuring the education level of a borrower; WORKTIME – a borrower's working experience measured by years; INCOME – the monthly income of a borrower; Year – a dummy controlling year effect; Region – a dummy variable reflecting the area in which a borrower is located; Industry – a dummy variable reflecting the industry that a borrower is working in; and Purpose – a dummy describing different purposes of borrowing.

(2) *,**, and ***indicate the statistical significance at the 10%, 5%, and 1% levels, respectively. Ordinary standard errors are used and Z-statistics are reported in parentheses. N is number of observations.

4.3 Other Robustness Checks

In addition to addressing the endogeneity concerns, we provide several robustness checks for our findings. First, the data we have used so far are a full sample without excluding the samples with extreme loan amount and interest rate values. To control the potential bias, we eliminate the loan listings whose amounts are larger than 200,000 or whose interest rates are higher than 24%. Second, given that a probit model is also suitable for a binary selection model, we apply the probit model to re-estimate the impact of gender on the probability of funding success and default. Third, in order to control the influence of possible heteroskedasticity, we redo the regressions by using the robust standard error and the bootstrap (100 times) standard error. Finally, we exclude loans whose repayment is in progress and re-estimate the logit regression on the probability of default. For the sake of brevity, these results are not presented but are available upon request. All of them are consistent with the baseline estimation results.

5. CONCLUSION

Gender equality, embedded in the behavior of family, market, and society, affects the prospects of poverty reduction and economic growth by stimulating productivity and earnings. The importance of financial access for females motivates us to investigate the role of gender in the new and rapidly growing fintech market in the Chinese context. On the one hand, the PRC now has the world's biggest P2P lending market; on the other hand, it remains a low-ranked country globally with regard to the gender gap. Recent estimates by the World Economic Forum (2017) suggest that the PRC could see a US$2.5 trillion GDP increase from gender parity by 2025 by closing the gender gap in economic participation by 25% over the same period.[6]

Using data from Renrendai, a leading P2P lending platform in the PRC, we show that loan listings requested by female borrowers are associated with a lower default probability, higher expected profit and lower expected loss. However, despite the higher creditworthiness, we don't find any significant impact of gender on funding probability. All our findings are robust to various checks, suggesting that female borrowers have to provide lenders with a higher rate of return to obtain a funding success rate comparable to their male peers. This in turn implies that female borrowers have been treated unfavorably by lenders in the P2P lending market.

To moderate biased lending, platforms should analyze the loan performance of different groups of borrowers, for instance male versus female, and incorporate such information into their credit rating system. At the same time, these platforms should educate lenders on how to judge the creditworthiness of borrowers by using unbiased information. Since the history of P2P lending is still very short, most investors on the platform are new and not sophisticated enough to evaluate the risks of loan listings properly. They may not be able to interpret the signals of quality correctly. Addressing the lack of financial literacy is of particular importance for the fintech market where there are no financial intermediaries and all decisions are decentralized. In addition to providing financing tools, fintech companies like P2P lending platforms should be encouraged to improve the average financial literacy of the public.

[6] http://reports.weforum.org/global-gender-gap-report-2017/.

REFERENCES

Agier, I., Szafarz, A., 2013. Microfinance and gender: Is there a glass ceiling on loan size? World Development 42, 165–181.

Alesina, A., Giuliano, P., Nunn, N., 2013a. On the origins of gender roles: Women and the plough. Quarterly Journal of Economics 128, 469–530.

Alesina, A.F., Lotti, F., Mistrulli, P.E., 2013b. Do women pay more for credit? Evidence from Italy. Journal of the European Economic Association 11, 45–66.

Arch, E.C., 1993. Risk-taking: A motivational basis for sex-differences. Psychological Reports 73, 3–11.

Arrow, K., 1971. The theory of discrimination. Princeton University, Department of Economics, Industrial Relations Section.

Asiedu, E., Freeman, J.A., Nti-Addae, A., 2012. Access to credit by small businesses: How relevant are race, ethnicity, and gender? American Economic Review 102, 532–537.

Asiedu, E., Kalonda-Kanyama, I., Ndikumana, L., Nti-Addae, A., 2013. Access to credit by firms in sub-Saharan Africa: How relevant is gender? American Economic Review 103, 293–297.

Aterido, R., Beck, T., Iacovone, L., 2013. Access to finance in sub-Saharan Africa: Is there a gender gap? World Development 47, 102–120.

Barasinska, N., Schafer, D., 2014. Is crowdfunding different? Evidence on the relation between gender and funding success from a German peer-to-peer lending platform. German Economic Review 15, 436–452.

Barber, B.M., Odean, T., 2001. Boys will be boys: Gender, overconfidence, and common stock investment. Quarterly Journal of Economics 116, 261–292.

Beck, T., Behr, P., Guettler, A., 2013. Gender and banking: Are women better loan officers? Review of Finance 17, 1279–1321.

Becker, G.S., 1957. The economics of discrimination. University of Chicago Press.

———., 1971. The economics of discrimination. University of Chicago Press.

Bellucci, A., Borisov, A., Zazzaro, A., 2010. Does gender matter in bank−firm relationships? Evidence from small business lending. Journal of Banking & Finance 34, 2968–2984.

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression diagnostics: Identifying influential data and sources of collinearity. John Wiley & Sons.

Berger, A.N., Kick, T., Koetter, M., Schaeck, K., 2013. Does it pay to have friends? Social ties and executive appointments in banking. Journal of Banking & Finance 37, 2087–2105.

Berkovec, J.A., Canner, G.B., Gabriel, S.A., Hannan, T.H., 1998. Discrimination, competition, and loan performance in FHA mortgage lending. Review of Economics and Statistics 80, 241–250.

Bernardi, R.A., Arnold, D.F., 1997. An examination of moral development within public accounting by gender, staff level, and firm. Contemporary Accounting Research 14, 653–668.

Bertocchi, G., Brunetti, M., Torricelli, C., 2011. Marriage and other risky assets: A portfolio approach. Journal of Banking & Finance 35, 2902–2915.

Bessis, J., 2015. Risk management in banking. John Wiley & Sons.

Blanchflower, D.G., Levine, P.B., Zimmerman, D.J., 2003. Discrimination in the small-business credit market. Review of Economics and Statistics 85, 930–943.

Buvinic, M., Berger, M., 1990. Sex differences in access to a small enterprise development fund in Peru. World Development 18, 695–705.

Carnabuci, G., Dioszegi, B., 2015. Social networks, cognitive style, and innovative performance: A contingency perspective. Academy of Management Journal 58, 881–905.

Cavalluzzo, K.S., Cavalluzzo, L.C., 1998. Market structure and discrimination: The case of small businesses. Journal of Money, Credit and Banking 30, 771–792.

Chen, J., Woon, S.L., Marc, G., 2017. The impact of board gender composition on dividend payouts. Journal of Corporate Finance 43, 86–105.

Chen, X., Huang, B., Ye, D., 2018. The role of punctuation in P2P lending: Evidence from China. Economic Modelling 68, 634–643.

Coates, J.M., Herbert, J., 2008. Endogenous steroids and financial risk taking on a London, trading floor. Proceedings of the National Academy of Sciences of the United States of America 105, 6167–6172.

Conyon, M.J., He, L., 2017. Firm performance and boardroom gender diversity: A quantile regression approach. Journal of Business Research 79, 198–211.

Cronqvist, H., Previtero, A., Siegel, S., White, R.E., 2016. The fetal origins hypothesis in finance: Prenatal environment, the gender gap, and investor behavior. Review of Financial Studies 29, 739–786.

Croson, R., Gneezy, U., 2009. Gender differences in preferences. Journal of Economic Literature 47, 448–474.

Cumming, D., Leung, T.Y., Rui, O., 2015. Gender diversity and securities fraud. Academy of Management Journal 58, 1572–1593.

Demirguc-Kunt, A., Klapper, L., Singer, D., Ansar, S., Hess, J., 2018. The Global Findex Database 2017: Measuring financial inclusion and the Fintech revolution. The World Bank.

Duarte, J., Siegel, S., Young, L., 2012. Trust and credit: The role of appearance in peer-to-peer lending. Review of Financial Studies 25, 2455–2483.

Eckel, C.C., Fullbrunn, S.C., 2015. Thar SHE blows? Gender, competition, and bubbles in experimental asset markets. American Economic Review 105, 906–920.

Faccio, M., Marchica, M.T., Mura, R., 2016. CEO gender, corporate risk-taking, and the efficiency of capital allocation. Journal of Corporate Finance 39, 193–209.

Fang, L.H., Huang, S., 2017. Gender and connections among Wall Street analysts. Review of Financial Studies 30, 3305–3335.

Fernandez, R., Wong, J., 2014. Unilateral divorce, the decreasing gender gap, and married women's labor force participation. American Economic Review 104, 342–347.

Gneezy, U., Leonard, K.L., List, J.A., 2009. Gender differences in competition: Evidence from a matrilineal and a patriarchal society. Econometrica 77, 1637–1664.

Greene, W.H., 2003. Econometric analysis. Pearson Education India.

Grinblatt, M., Keloharju, M., 2001. How distance, language, and culture influence stockholdings and trades. Journal of Finance 56, 1053–1073.

Grinblatt, M., Keloharju, M., Linnainmaa, J., 2011. IQ and stock market participation. Journal of Finance 66, 2121–2164.

Hayre, L., Mohebbi, C., 1992. Mortgage mathematics. published in Handbook of Mortgage-Backed Securities, Frank Fabozzi (ed.), Chicago, Illinois: Probus Publishing Company.

Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.

Heckman, J., Ichimura, H., Smith, J., Todd, P., 1998. Characterizing selection bias using experimental data. National Bureau of Economic Research.

Hertzberg, A., Liberti, J.M., Paravisini, D., 2010. Information and incentives inside the firm: Evidence from loan officer rotation. Journal of Finance 65, 795–828.

Huang, J., Kisgen, D.J., 2013. Gender and corporate finance: Are male executives overconfident relative to female executives? Journal of Financial Economics 108, 822–839.

Huberman, G., 2001. Familiarity breeds investment. Review of Financial Studies 14, 659–680.

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

Jurkus, A.F., Park, J.C., Woodard, L.S., 2011. Women in top management and agency costs. Journal of Business Research 64, 180–186.

Ladd, H.F., 1998. Evidence on discrimination in mortgage lending. Journal of Economic Perspectives 12, 41–62.

Leary, M.T., Roberts, M.R., 2014. Do peer firms affect corporate financial policy? Journal of Finance 69, 139–178.

Lennox, C.S., Francis, J.R., Wang, Z.T., 2012. Selection models in accounting research. Accounting Review 87, 589–616.

Lin, M.F., Viswanathan, S., 2016. Home bias in online investments: An empirical study of an online crowdfunding market. Management Science 62, 1393–1414.

Little, R.J., 1985. A note about models for selectivity bias. Econometrica 53, 1469–1474.

Mantel, N., Haenszel, W., 1959. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 22, 719–748.

Michels, J., 2012. Do unverifiable disclosures matter? Evidence from peer-to-peer lending. Accounting Review 87, 1385–1413.

Moro, A., Wisniewski, T.P., Mantovani, G.M., 2017. Does a manager's gender matter when accessing credit? Evidence from European data. Journal of Banking & Finance 80, 119–134.

Muravyev, A., Talavera, O., Schafer, D., 2009. Entrepreneurs' gender and financial constraints: Evidence from international data. Journal of Comparative Economics 37, 270–286.

Paglia, J.K., Harjoto, M.A., 2014. The effects of private equity and venture capital on sales and employment growth in small and medium-sized businesses. Journal of Banking & Finance 47, 177–197.

Phelps, E.S., 1972. The statistical theory of racism and sexism. American Economic Review 62, 659–661.

Pope, D.G., Sydnor, J.R., 2011. What's in a picture? Evidence of discrimination from Prosper.com. Journal of Human Resources 46, 53–92.

Powell, M., Ansic, D., 1997. Gender differences in risk behaviour in financial decision-making: An experimental analysis. Journal of Economic Psychology 18, 605–628.

Ravina, E., 2012. Love & loans: The effect of beauty and personal characteristics in credit markets. Working Paper, SSRN 1101647.

Rosenbaum, P.R., 2002. Observational studies. Springer.

Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

Seasholes, M.S., Zhu, N., 2010. Individual investors and local bias. Journal of Finance 65, 1987–2010.

Shahriar, A.M., 2016. Exposure to intimate partner violence and repayment of microcredit: Evidence from field experiments in Bangladesh. Journal of Banking & Finance 72, 1–14.

Staiger, D., Stock, J.H., 1994. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

Steinbuks, J., 2015. Effects of prepayment regulations on termination of subprime mortgages. Journal of Banking & Finance 59, 445–456.

Sundén, A., Surette, B., 1998. Gender differences in the allocation of assets in retirement savings plans. American Economic Review 88, 207–211.

Tourani-Rad, A., Gilbert, A., Chen, J., 2016. Are foreign IPOs really foreign? Price efficiency and information asymmetry of Chinese foreign IPOs. Journal of Banking & Finance 63, 95–106.

Wiswall, M., Zafar, B., 2018. Preference for the workplace, investment in human capital, and gender. Quarterly Journal of Economics 133, 457–507.

World Economic Forum, 2017. The Global Gender Gap Report 2017. Available at: https://www.weforum.org/reports/the-global-gender-gap-report-2017.

Zhao, Z., 2004. Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence. Review of Economics and Statistics 86, 91–107.