China Economic Review 78 (2023) 101944

Contents lists available at ScienceDirect

China Economic Review

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

Big data pricing in marketplace lending and price discrimination against repeat borrowers: Evidence from China

Geran Tian [a,b], Weixing Wu [b,*]

[a] School of Finance, Capital University of Economics and Business, Zhangjia Lukou 121, Fengtai, 100070 Beijing, China

[b] School of Banking and Finance, University of International Business and Economics, Huixindongjie 10, Chaoyang, 100029 Beijing, China

ARTICLE INFO

JEL codes:
D14
G21
G23
G51

Keywords:
Big data
Fintech
Marketplace lending
Peer-to-peer lending
Price discrimination
Fair lending
Behavior-based price discrimination

ABSTRACT

In this work, we systematically investigate the pricing mechanism change from auction to big data pricing on one of the major marketplace lending platforms in China. We find that big data pricing reduces the average interest rate while the borrowers with delinquency or default histories are assigned higher interest rates. However, repeat borrowers are also faced with growing interest rates, even though they have been paying their debts on time. Further analysis shows that repeat borrowers have lower income and education levels. Moreover, investor returns become less dispersed after pricing with big data, which can be a result of homogeneous loans on the market. The implications of the above findings are discussed.

1. Introduction

Since customer information was not as ubiquitous and available as it is today, the assessment of loan applicants mainly relied on credit scores, such as FICO. However, it is impossible for some loan applicants to be funded because they have unsatisfactory credit scores or they have no credit scores at all. Nowadays, with the development of financial technologies, modern loan providers can adopt big data to upgrade the risk management procedure and assess customer creditworthiness, by analyzing behavioral indicators and spending patterns. It is believed that big data can help to extract the value of data and it will be less risky for financial companies when predicting which clients will be successful in their payments. As a result, more people could have access to credit loans (Pérez-Martín, Pérez-Torregrosa, & Vaca, 2018).

Promising as it may seem, there is a lack of empirical evidence on how these technologies affect creditors and debtors. The adoption of big data pricing by a marketplace lending (also known as peer-to-peer lending) platform in China offers us an opportunity to investigate what role big data plays. This marketplace lending platform changed the pricing rule from auction design in favor of posted prices to big data pricing on 14 October 2015 (Dong, 2017), after which the loan interest rates are assigned by the platform rather than borrowers themselves. This is not uncommon for marketplace lending platforms. Other leading platforms such as Funding Circle and Prosper switched their pricing mechanism likewise (Wei & Lin, 2017).

Since the essence of big data is to assess credit risk by analyzing recorded behavioral information, a natural thought is that

* Corresponding author.
E-mail addresses: tiangeran@cueb.edu.cn (G. Tian), wxwu@uibe.edu.cn (W. Wu).

https://doi.org/10.1016/j.chieco.2023.101944
Received 4 August 2022; Received in revised form 25 November 2022; Accepted 23 February 2023
Available online 28 February 2023

Table 1
Variable definitions.

Variable	Definition
Portfolio return (%)	Annual return of investor's portfolio.
Portfolio default rate	The number of default notes divided by total number of notes in the portfolio.
Portfolio default rate (weighted by face value)	The number of default notes divided by total number of notes, weighted by face values.
Repeat	Repeat = 1 if it is not the very first loan of the borrower; Repeat = 0 if the loan is the very first loan of the borrower.
Annual percentage rate (APR)	Annual percentage interest rate of a loan.
Amount (yuan)	Total borrowing amount of a loan.
Transaction amount (10,000 yuan)	Transaction amount of a loan in the secondary market.
Repayment terms (month)	Loan repayment terms in month.
Title length	The number of Chinese characters in the loan listing title.
Description length	The number of Chinese characters in the loan description divided by 100.
Gender	Gender of a borrower. 1 = male and 0 = female.
Age	Age of the borrower.
Education	Education level of the borrower, including the following categories: middle/high school; college graduate; university graduate; postgraduate.
Income (yuan)	Monthly income of the borrower, including the following categories: <1000; 1001–2000; 2001–5000; 5001–10,000; 10,001–20,000; 20,001–50,000; >50,000.
Married	Marital status of the borrower. 1 = married; 0 = single, widowed, or divorced.
Work experience	Work experience of the borrower, including the following categories: no work experience or <1 year; 1 to 3 years; 3 to 5 years; >5 years.
Credit risk	Credit risk is measured according to the credit rating assigned by the marketplace lending platform. Credit risk = 1 if credit rating is "AA"; 2 if credit rating is "A"; 3 if credit rating is "B"; 4 if credit rating is "C"; 5 if credit rating is "D"; 6 if credit rating is "E"; 7 if credit rating is "HR", where "HR" represents "high risk".
Homeowner	1 if the borrower owns real estate properties; 0 otherwise.
Mortgage	1 if the borrower has outstanding mortgages; 0 otherwise.
Car owner	1 if the borrower owns cars; 0 otherwise.
Car loan	1 if the borrower has outstanding car loans; 0 otherwise.

borrowers with punctual repayment history should enjoy lower, or at least not higher, interest rates because they have exhibited creditworthiness. On the other hand, returning borrowers with delinquency or default records should face higher interest rates. However, it has been reported that some Internet enterprises utilize big data to price discriminate against existing users and this phenomenon has triggered many discussions in academic, legal, and government sections around the world (e.g., Bar-Gill, 2019; Botta & Wiedemann, 2019; Gillis & Spiess, 2019; Liu, Long, Xie, Liang, & Wang, 2021; Shiller, 2016; Townley, Morrison, & Yeung, 2017; Xinhua, 2021). It is unclear which effect is true - risk identification or price discrimination. If both effects exist, it is ambiguous which effect dominates the other.

We observe the pricing mechanism change from auction-style to big data pricing on a major peer-to-peer lending market in China and we seize this opportunity to systematically investigate how borrowers and investors are respectively affected by big data pricing. We find that after pricing with big data, the platform only allows borrowers with favorable credit ratings to enter the market and therefore the loan applications are more likely to be successfully funded. Although first-degree price discrimination is unlikely to take place in the real world, big data has made personalized price discrimination possible. Loan interest rates become lower in general, but borrowers with delinquency or default history are required to pay higher interest rates. Meanwhile, returning borrowers are confronted with higher interest rates as well, providing supportive evidence of "big data price discrimination against existing customers" or "big data price discrimination". As for investors, the adoption of big data pricing reduces the default possibility, possibly due to initial screening. Moreover, both the mean and the variance of portfolio return among investors are reduced. In other words, although risk takers and risk avoiders have different preferences, their portfolio returns become less divergent after big data pricing.

The rest of this manuscript is organized as follows. Section 2 provides a brief review of price discrimination in financial markets and price discrimination against repeat customers. Section 3 introduces data and variables. In Section 4, we provide empirical evidence that some repeated borrowers are charged with higher interest rates after controlling for a series of variables such as loan characteristics and repayment history. In section 5, we further evaluate the influence of big data pricing in two aspects - loan default and investor performance. Section 6 summarizes the findings, discusses the implication of this work, and concludes.

2. Literature review

2.1. Price discrimination in financial markets

Price discrimination has been found in multiple types of financial markets including the loan market, the mortgage market, the peer-to-peer lending market, and the derivative market.

Some researchers find evidence of racial discrimination in loan markets (e.g. Bartlett, Morse, Stanton, & Wallace, 2022; Harkness, 2016; Kau, Keenan, & Munneke, 2012), while other researchers attribute different loan approval rates or different interest rates to other factors. For example, Courchane and Nickerson (1997) point out that the differentials occurred because of changes in lock dates or close dates, rather than the intentional behavior of the loan officer. Black, Boehm, and DeGennaro (2003) show that minorities pay larger mortgage overages due to differences among borrowers rather than racial discrimination. Similarly, Ladd (1998) believes that the observed differentials are profit-motivated statistical discrimination. Moreover, it is found that estimated racial disparity in loan approval rates declines with the length of borrower's credit history, and no significant racial disparity was found among borrowers with long credit histories (Han, 2011). Ferguson and Peters (1995) point out that if minorities have a lower level of creditworthiness, then "colorblind" lending will not result in equal denial or default rates across different segments of the population. Therefore, the disparities among races in the lending market are not necessarily racial discrimination unless creditworthiness has been controlled.

Other than race, some other factors can also be the source of discrimination, such as gender (Chen, Huang, & Ye, 2020; Chen, Li, & Lai, 2016; Harkness, 2016), marital status[1] (Elliehausen & Lawrence, 1990), regions (Meyer, 1967), loan size (Benston, 1964), nature of the investors[2] (Hau, Hoffmann, Langfield, & Timmer, 2021; Li, Zhao, & Zhong, 2019) and sophistication of the investors (Hau et al., 2021).

2.2. Behavior-based price discrimination

Behavior-based price discrimination is defined as follows (Fudenberg & Villas-Boas, 2006): "When firms are able to recognize their previous customers, they may be able to use their information about the consumers' past purchases to offer different prices and/or products to consumers with different purchase histories." For example, airline companies are found to price discriminate against business travelers by charging higher for tickets bought on weekdays (Puller & Taylor, 2012). Even in a competitive market for homogeneous goods, it is found that a food distributor enjoys market power coming from search costs and offers different prices to repeat buyers, and such price discrimination decreases total welfare (Marshall, 2020). One way to avoid being price discriminated is maintaining anonymity. However, Conitzer, Taylor, and Wagman (2012) establish a model with a monopolist and a continuum of heterogeneous consumers and reveals that the benefit to consumers is determined by the control right of the cost of maintaining anonymity. In order to understand consumer reaction to seller behaviors, Maxwell and Garbarino (2010) conduct an online survey and identify the following norms: for a given item a seller should charge the same price to all customers; a seller should not charge a higher price to either more loyal, more frequent customers, or new customers or infrequent customers; and a seller should not charge a lower price to infrequent purchasers.

3. Data and variables

The data used in this work is retrieved from Renrendai - one of the largest marketplace lending platforms in China. Established in 2010, this marketplace lending platform has attracted 48 million users to register. Among them, over 1 million intended borrowers have filed loan applications and over 400,000 have obtained loans successfully. Unlike some marketplace lending platforms that build lending relationships among acquaintances, this platform connects stranger borrowers and lenders.

The loan pricing mechanism before 14 October 2015 is analogous to the auction. The borrowers could post loan listings specifying the listing title, loan purpose, borrowing amount (multiples of 50), annual interest rate, and repayment terms. Additionally, in order to raise funds rapidly and to build lender trust, borrowers can also provide their education, income, marital status, etc. Afterward, the platform assigns a credit risk level to each borrower. Then borrowers can wait for investors to bid. Loans are split into 50 yuan notes. Investors can bid one or more notes on the loan. A loan listing is successful only if the bidding amount reaches the requested borrowing amount within seven days of listing.

After 14 October 2015, when the loan application is submitted, the platform analyzes the information provided by the intended borrower together with data collected in the past years and then assigns interest rates to the loan application. The assigned interest rate cannot be altered or bargained. If the borrower accepts the assigned interest rate, this loan listing will be on the market and can be bid upon. The time required to complete loan bidding and the face value of a single loan note remain unchanged.

The dataset used in this work contains information on borrowers, investors, loan applications, and transactions, from 2010 to 2017. With these datasets, it is feasible to combine each investor's bidding information with the corresponding transaction records to compute the holding months of each loan note. Next, we break down each borrower's repayment record into 50-yuan notes. Afterward, we map the holding history with decomposed repayment records and develop cash inflows and outflows for each investor. The cash flows have taken default records into account. Next, we compute the monthly return such that the net present value of cash outflow (i.e. investments) equals the net present value of cash inflow (i.e., repayments). Then, we multiply the monthly return by 12 to obtain the portfolio annual return.

Furthermore, two indicators are adopted to measure the defaults in the portfolios. The first is the portfolio default rate, which equals the number of default notes divided by the total number of notes in the portfolio. The second is the weighted average portfolio default rate, weighting the number of notes by the corresponding face values. The rest of the variables in this work are either loan-level (e.g., loan interest rate) or borrower-level (e.g., borrower credit risk). Table 1 shows the definition of variables, including investor-level, loan-level, and borrower-level variables. Note that in this work, we only include loans for which borrowers repay with equal monthly installments because this type of loan constitutes over 96% of the entire sample. Another repayment structure is "interest only" (borrowers repay interest only for each month and repay the principal at the end of the loan term). We exclude "interest only"

[1] Elliehausen and Lawrence (1990) find that lenders may have discriminated against single borrowers of both genders and against widows.

[2] Li et al. (2019) show evidence of price discrimination against retail investors. Hau et al. (2021) find that when trading with the same dealer, median nonfinancial clients pay more than the large bule-chip companies in the over-the-counter market.

Table 2
Summary statistics: before and after big data pricing.

	Successful loans before 13 October 2015					Successful loans after 15 October 2015				
Variable	N	Mean	Std. Dev.	Min	Max	N	Mean	Std. Dev.	Min	Max
Loan variables										
APR	193,646	12.25	1.25	3	24.4	256,121	10.23	0.49	8	13.2
Loan amount	193,646	57,857.14	34,859.97	3000	500,000	256,121	84,534.72	50,308.65	1000	500,000
Repayment terms	193,646	26.15	10.2	3	36	256,121	34.07	6.81	3	48
Title length	193,646	5.03	3.18	0	50	256,121	4.18	1.67	0	29
Loan description length	193,646	1.12	0.59	0	4.79	256,121	0.99	0.28	0	4.63
Borrower variables										
Gender	193,646	0.74	0.44	0	1	256,121	0.66	0.47	0	1
Age	193,646	37.81	8.42	20	73	256,121	36.09	8.35	0	63
Income										
Income = <1000	193,646	0.00	0.04	0	1	255,561	0.02	0.15	0	1
Income = 1001–2000	193,646	0.00	0.05	0	1	255,561	0.07	0.05	0	1
Income = 2001–5000	193,646	0.24	0.43	0	1	255,561	0.18	0.39	0	1
Income = 5001–10,000	193,646	0.36	0.48	0	1	255,561	0.32	0.47	0	1
Income = 10,001–20,000	193,646	0.18	0.39	0	1	255,561	0.21	0.41	0	1
Income = 20,001–50,000	193,646	0.12	0.33	0	1	255,561	0.14	0.35	0	1
Income = >50,000	193,646	0.08	0.28	0	1	255,561	0.06	0.23	0	1
Education										
Education = middle/high school	193,628	0.26	0.44	0	1	255,561	0.09	0.28	0	1
Education = college graduate	193,628	0.50	0.50	0	1	255,561	0.44	0.50	0	1
Education = university graduate	193,628	0.22	0.42	0	1	255,561	0.39	0.49	0	1
Education = postgraduate	193,628	0.01	0.12	0	1	255,561	0.02	0.14	0	1
Work experience										
Work experience = <1 year	193,520	0.67	0.47	0	1	253,317	0.09	0.28	0	1
Work experience = 1 to 3 years	193,520	0.17	0.38	0	1	253,317	0.44	0.50	0	1
Work experience = 3 to 5 years	193,520	0.06	0.24	0	1	253,317	0.39	0.49	0	1
Work experience = >5 years	193,520	0.09	0.29	0	1	253,317	0.08	0.14	0	1
Credit level	193,646	2.51	1.43	1	7	256,120	2.06	0.49	1	7
Homeowner	193,646	0.53	0.5	0	1	255,746	0.45	0.5	0	1
Mortgage	193,646	0.35	0.48	0	1	255,746	0.22	0.41	0	1
Car owner	193,646	0.29	0.45	0	1	255,746	0.18	0.39	0	1
Car loan	193,646	0.08	0.27	0	1	255,746	0.04	0.19	0	1

Note: In this table, we only include loans that repay with equal monthly installments. *** indicates statistical significance at 1% level. See Table 1 for variable definitions.

loans because the platform may price these loans very differently. Table 2 shows the summary statistics of the variables.

As shown in Table 2, the average APR dropped from 12.25% to 10.23% after big data pricing. The average loan amount decreased from 57,857.14 yuan to 84,534.72 yuan after the big data pricing, resulting in an elevated amount of 26,677.58 yuan. In some other articles using Renrendai data (Chen, Huang, & Shaban, 2020; Chen, Huang, & Ye, 2018; Chen, Huang, & Ye, 2020; Li, Deng, & Li, 2020), the loan amount ranges from 53,103 to 68,599 yuan, confirming the representativeness of our sample in this work. The loan term was prolonged from 26.15 to 34.07 months. Both loan title and loan description become shorter. As for borrowers, the sample becomes more gender-balanced and younger. The monthly income of 5001–10,000 category still covers the highest proportion of borrowers. In contrast, the education level of borrowers is greatly improved – borrowers with middle school or high school education are reduced from 26% to 9% and university graduates increase from 22% to 39%. Most borrowers (67%) before big data pricing have <1 year of working experience, but after big data pricing this category drops to only 9%. Notably, the average credit risk of borrowers decreases from 2.51 to 2.06. A lower proportion of borrowers have real estate properties, cars, mortgages, or car loans. Meanwhile, the repayment term becomes 7 months longer and the average annual percentage rate (APR) of loans is 2.02% lower.

4. Big data pricing and its effects on borrowers

In order to decipher how big data pricing has influenced the borrowers, we explore two effects – extensive effects (borrow or not borrow) and intensive effects (how much is borrowed, for how long, at what price, etc.).

4.1. Extensive effects

We first examine the extensive effect of big data pricing, namely how loan application success rate is influenced. Fig. 1 shows that before the pricing mechanism change, both successful and unsuccessful loan listings increase over time, and the unsuccessful loan listings greatly outnumber successful loan listings, resulting in a relatively low success rate. However, after the platform begins pricing loans with big data, the number of unsuccessful loan listings exhibits a sharp decline to almost zero while the number of successful loans slightly declines first and shows a steady increase after several months. Fig. 1 provides graphical evidence that big data pricing technology works as a filter by preventing some borrowers from entering the market in the first place. There has been a similar case that the loans are funded with a higher probability on Prosper Marketplace when the pricing mechanism changes from auction to platform-mandated (Wei & Lin, 2017). In fact, Table 2 can provide some suggestive evidence on what kind of borrowers are turned away by big data pricing. Borrowers with lower education levels (middle/high school) and few years of experience (<1 year) are substantially reduced.

4.2. Intensive effects

In this section, we investigate the intensive effect of big data pricing on borrowers. We attempt to find some general patterns first and proceed to examine the heterogenous effects on specific types of borrowers.

4.2.1. Overall effect

Table 2 in Section 3 reveals that big data pricing brings along longer maturity and lower interest rate. In fact, longer maturity and lower interest rate are also observed in the entire peer-to-peer lending market in China. Appendix A (Fig. A1 and Fig. A2) shows that APR decreases from over 20% in early 2014 to slightly below 10% by the end of 2019, and the loan maturity increases from <6 months to approximately 16 months.

In order to minimize the confounding effect brought by the overall market trend, we identify the relationship between the pricing mechanisms (whether the loan is priced by big data or by auctions) and the resulting loan interest rates by exploiting a sharp time discontinuity. Under the framework of regression discontinuity design, the day that the pricing mechanism change took place is regarded as the cutoff and the time window is from 365 days before the big data pricing (i.e. 14 October 2014) to 365 days after the big data pricing (i.e. 13 October 2016). This time window minimizes other potential shocks that lead to market rate disturbances while allowing a reasonable quantity of loans to be included in the analysis for robust and reliable results. Fig. 2 shows the graphical result of the regression discontinuity design using a linear fit. The loan interest rate exhibits a sharp drop by the time big data pricing is introduced. In fact, the fitted value from the left side of the cutoff is 11.698 and the fitted value from the right side is 11.194. Thus, the discontinuous jump is 0.504.

Additionally, we conduct a placebo test (or falsification test) using another time cutoff as our artificial cutoff. If the observed decrease in loan interest on 14 October 2015 is solely caused by big data pricing itself, then a similar reduction in loan interest rates should not be observed around our artificial time cutoff. Therefore, we move the entire time window one year earlier as a placebo test. In other words, the hypothetical cutoff is 14 October 2014 and the time window covers from 365 days before (i.e. 14 October 2013) to 365 days after the hypothetical cutoff (i.e. 14 October 2015). We choose not to move the time window and the cutoff to one year after the actual pricing mechanism change because there were two major regulatory changes[3] around 14 October 2016 and they may introduce significant noises into the placebo test. Fig. 3 shows the results. No significant difference can be observed at the artificial cutoff. The fitted values from the two sides are 12.84674 and 12.84597 respectively. The difference is trivial.

4.2.2. Heterogeneous effect

Next, we attempt to determine whether the big data prices certain types of borrowers differently. First, we generate a dummy variable *BDP* (big data pricing) which equals 1 if the loan is originated after the pricing mechanism change and 0 otherwise, and then regress *BDP*, together with loan and borrower characteristics, on the loan interest rate. The regression model is shown in Eq. (1):

$$APR_j = \alpha + \beta BDP_j + \gamma X_i + \varepsilon_j \tag{1}$$

where j denotes each loan and i denote each investor. X_i is a strand of investor control variables and ε_j represents the error term. Column (1) of Table 3 shows the benchmark regression by including *BDP* as the main explanatory variable. It shows that after controlling for the borrower and loan characteristics, loan interest rate is overall reduced by 2.056% after the pricing mechanism change and this coefficient is statistically significant at 1% level.

We further examine how big data prices loans by incorporating relevant factors into the regression analysis. First, we construct a dummy variable *Repeat*, which equals 0 if the loan is the very first loan of the corresponding borrower and equals 1 if it is a repeated loan from a returning borrower. We also construct a variable *Default record*, which equals 1 if the borrower has delinquency or default records at the time of pricing mechanism change and 0 otherwise. The corresponding models are shown below:

$$\mathrm{AP}R_j = \alpha + \beta BDP_j + \varphi BDP_j{*}Default\ record_i + \mu Default\ record_i + \gamma X_i + \varepsilon_j \tag{2}$$

$$APR_j = \alpha + \beta BDP_j + \varphi BDP_j{*}Repeat_j + \mu Repeat_j + \gamma X_i + \varepsilon_j \tag{3}$$

Column (2) of Table 3 corresponds to Eq. (2). The coefficient of the interaction term *BDP* Default record* is 1.485 and is significant at 1% level. It means after big data pricing, borrowers with delinquency or default records have to pay on average 1.485% higher interest

[3] In October 2016, "Interim Measures for the Administration of the Business Activities of Online Lending Information Intermediaries" was issued. In November 2016, "Guidelines for the Administration of Filing and Registration of Online Lending Information Intermediaries" was issued. Combining with "Guidelines for Online Loan Fund Depository Business" issued in February 2017 and "Guidelines for Information Disclosure of Business Activities of Online Lending Information Intermediaries" issued in August 2018, these four regulations are considered the most impactful policies (often referred to as the "3 + 1" regulatory framework) for marketplace lending business in China (Ding, Kavuri, & Milne, 2020).

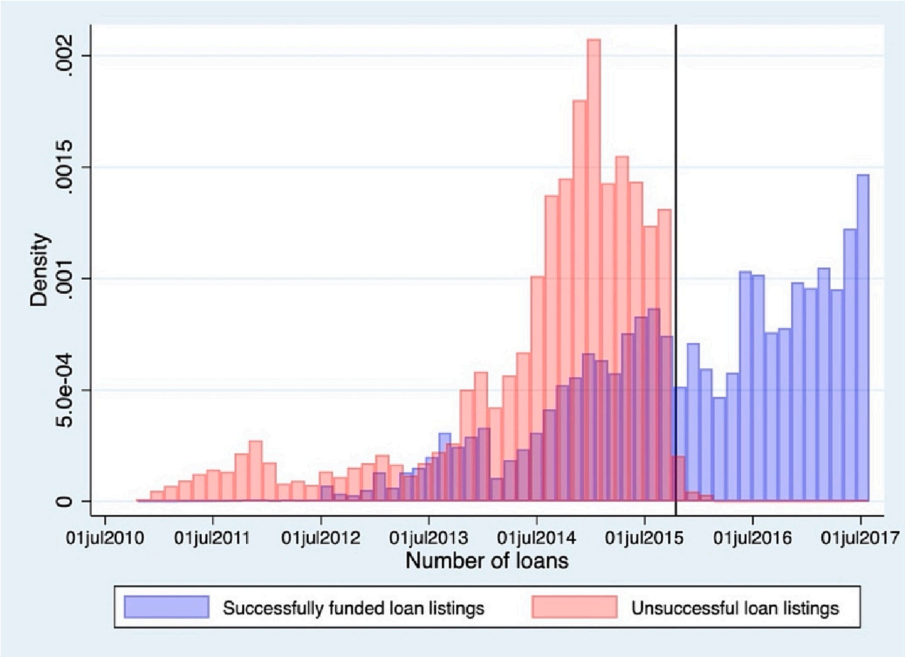

Fig. 1. Successfully-funded loan listings and unsuccessful loan listings before and after big data pricing.
Note: This figure includes 477,592 loans originated from 12 October 2010, to 31 July 2017. The black vertical line marks the pricing mechanism change on 14 October 2015.

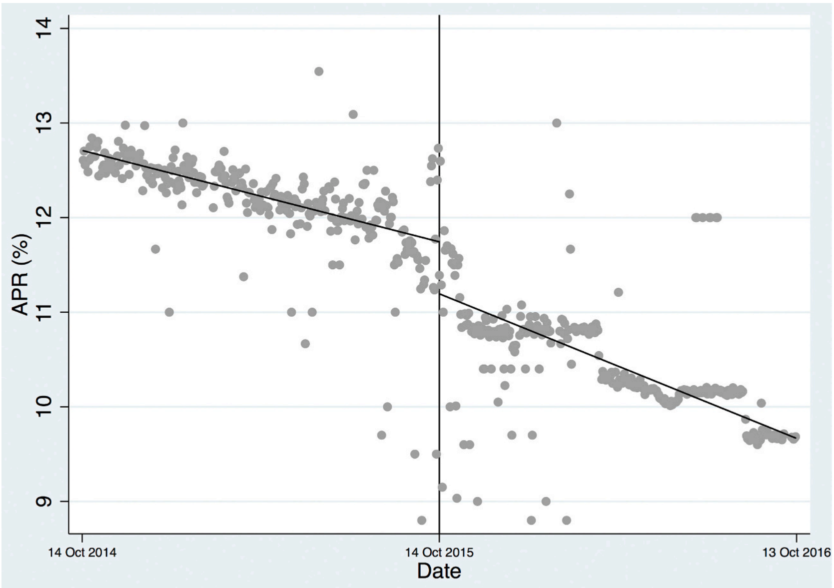

Fig. 2. Loan interest rates before and after big data pricing.
Note: This figure incorporates 240,752 successfully-funded loans that originated from 365 days before big data pricing (i.e. 14 October 2014) to 365 days after big data pricing (i.e. 13 October 2016). The vertical line marks the pricing mechanism change on 14 October 2015.

rate in order to be funded. This is in accordance with economic intuition. However, the coefficient estimate of *Default record* is negative. Existing literature indicates that this may due to borrowers' manipulation of information. It is found that information disclosure by borrowers can significantly improve the funding probability and this effect is even more prominent for borrowers with lower credit ratings (Chen, Huang, & Ye, 2020; Li, So, & Yuan, 2018). Moreover, voluntary information disclosure can decrease the loan interest rate by approximately 0.2% (Li et al., 2018). Nevertheless, the loan default rate increases with the information disclosure

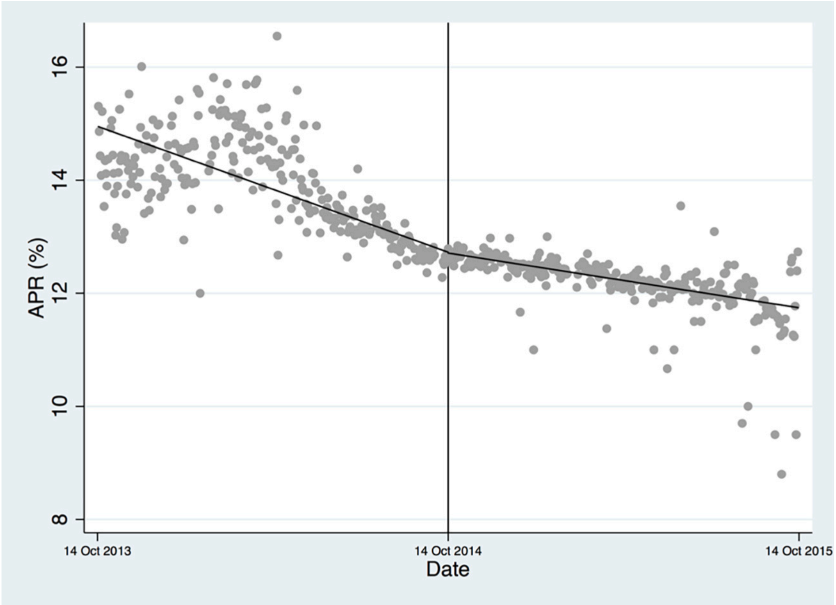

Fig. 3. A placebo test.
Note: This figure incorporates 168,375 successfully-funded loans that originated from 365 days before the hypothetical cutoff (i.e. 14 October 2013) to 365 days after the hypothetical cutoff (i.e. 14 October 2015). The vertical line marks the hypothetical cutoff 14 October 2014.

Table 3
Loan interest rate before and after pricing with big data.

	(1)	(2)	(3)	(4)
	APR	APR	APR	APR
BDP	−2.056***	−2.071***	−2.073***	−2.074***
	(−512.17)	(−520.64)	(−527.39)	(−527.92)
BDP* Default record		1.485***		
		(18.02)		
Default record		−0.135***		−0.0527**
		(−5.68)		(−2.24)
BDP*Repeat			0.345***	0.342***
			(17.64)	(17.38)
Repeat			0.223***	0.223***
			(13.29)	(13.29)
Transaction amount	−0.00276***	−0.00256***	−0.00315***	−0.00313***
	(−6.40)	(−5.94)	(−7.29)	(−7.25)
Repayment terms	0.0468***	0.0472***	0.0481***	0.0482***
	(157.46)	(158.64)	(168.18)	(168.24)
Payment structure	−1.077***	−1.074***	−1.057***	−1.046***
	(−13.41)	(−13.29)	(−12.89)	(−12.67)
Title length	0.0377***	0.0374***	0.0363***	0.0363***
	(35.98)	(35.70)	(35.02)	(34.94)
Description length	−0.0387***	−0.0385***	−0.0292***	−0.0288***
	(−7.54)	(−7.50)	(−5.74)	(−5.67)
Credit risk	0.383***	0.391***	0.357***	0.362***
	(127.10)	(103.61)	(107.92)	(90.32)
Other borrower characteristics	Yes	Yes	Yes	Yes
N	449,767	449,767	449,767	449,767

Note: Robust t-statistics are reported in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level, respectively. *BDP* is a dummy variable that equals 1 if the loan is originated after the pricing mechanism change and 0 otherwise. *Repeat* is a dummy variable that equals 0 if the loan is the very first loan of the borrower on this marketplace lending platform and 1 if it is a repeated loan from a returning borrower.

amount, indicating the possibility of information manipulation by the borrowers (Chen, Huang, & Ye, 2020). Therefore, some borrowers with unfavorable repayment records are frequent borrowers who have accumulated experience on how to obtain favorable interest rates for their own benefit and this may have led to the negative coefficient of *Default record*.

Column (3) of Table 3 corresponds to Eq. (3). The coefficient of the interaction term of *BDP*Repeat* is 0.345, significant at 1% level.

To eliminate the possibility that returning borrowers are less trustworthy in nature, we include the repayment history variable *Default record* in Column (4), and the coefficient of *BDP*Repeat* is still positive and significant at 1% level. This indicates that with big data pricing, returning borrowers have to undertake higher interest rates.

Since *Repeat* can only distinguish new borrowers and returning borrowers, the information contained in this variable is compromised and so are the corresponding coefficients. To address this concern together with the unexpected sign of the coefficient of *Default record* in Columns (3) and (4) in Table 3, we instead use the discrete variable *Borrow n-th time* as an alternative measure for *Repeat*, which equals the number of times that the user has borrowed on the platform.

On the other hand, we also establish a series of dummy variables *Borrow x-th time* ($x = 2, 3, 4, 5$ or $x > 5$[4]) to dissect how the marketplace lending platform charges borrowers with different numbers of past borrowing experiences. To control for the creditworthiness of borrowers, only borrowers with punctual repayments (no delinquency or default records) are considered. Moreover, credit risk is included in the regressions together with other loan and borrower characteristics as control variables.

Table 4 shows that, of all the 256,121 loans originated after big data pricing, most are one-time borrowing, and 75,141 of them are repeated loans from existing borrowers, accounting for 29.3% of the loan sample. The mean loan interest rate increases monotonically with how many times it is for the same user to borrow after big data pricing. The average loan interest rate for a user to borrow for the first time after big data pricing is 10.01% but for the second time the loan interest rate increases by 0.41% and reaches 10.42%. When the user keeps borrowing, the interest rate increases by 0.04% - 0.42% each time.

However, summary statistics only show the unconditional mean of loan interest rates without controlling for other variables. This result likely comes from borrowers' unfavorable characteristics. Since the interest rate on marketplace lending platforms is usually considerably higher than interest rates offered by formal financial institutions such as banks, those who do not have access to formal finance are more likely to turn to marketplace lending than those who do have access to less expensive financing channels. Therefore, we include borrower characteristics and loan characteristics in the regressions to control for this factor. The regression model is shown in Eq. (4).

$$APR_j = \alpha + \beta BDP_j + \mu Borrow\ n\ th\ time_j + \gamma X_i + \varepsilon_j \tag{4}$$

In column (1) of Table 5, only *Borrow n-th time* and investor characteristics are included in the model. The coefficient of *Borrow n-th time* is positive and significant at 1% level. It shows that the loan interest rate increases with how many times the user has borrowed. The magnitude of this coefficient does not seem large and it is possibly due to the nonlinear effect brought by extremely frequent borrowers, who will be grouped in column (3). In column (2), *BDP*, *Borrow n-th time*, and their interaction term *BDP * Borrow n-th time* are included in the regression. After controlling for *BDP*, the coefficient of *Borrow n-th time* becomes insignificant but the coefficient of the interaction term is positive and significant at 1% level. It reveals that frequent borrowers are faced with higher interest rates after the pricing mechanism change. In column (3), the discrete variable *Borrow n-th time* is replaced by dummy variables *Borrow x-th time* to show decomposed pricing scheme imposed by the platform. If we take borrowing for the first time as the benchmark, borrowers encounter significantly higher interest rates when borrowing for the second, third, fourth, fifth time or more. The increase in interest rate is greater when borrowing more times. After grouping those who borrowed more than five times, the coefficients are not only statistically significant but become economically significant as well.

Our empirical results are in line with previous findings. A number of studies have estimated the elasticity of demand for microcredit based on data from China, India, the United States, and South Africa (Bell, Srintvasan, & Udry, 1997; Gross & Souleles, 2002; Karlan & Zinman, 2008; Kochar, 1997; Turvey, He, Ma, Kong, & Meagher, 2012; Weersink, Vanden Dungen, & Turvey, 1994), and the elasticity ranges from −0.14 to −0.84, indicating that microcredit consumers are insensitive to interest rate changes.

Note that the difference in interest rates between first-time borrowers and borrowers who borrowed >5 times is 1.18% as shown in Table 4, while the coefficient in column (3) of Table 5 is 0.829%. It seems that the differential effect is related to control variables, i.e. borrower and loan characteristics. Previous literature also shows that borrower heterogeneity plays a significant role in the demand for credit (Bogan, Turvey, & Salazar, 2015; Gross & Souleles, 2002; Lukas, 2017). Therefore, it is necessary to carefully examine the repeat borrowers in our sample – who on earth have borrowed multiple times in marketplace lending despite the growing financing cost.

To decipher which types of users borrow repeatedly after big data pricing, we regress borrower characteristics on *Borrow more than once after BDP* using the Probit model and OLS model, which are shown in Eq. (5) and Eq. (6) respectively.

$$Pr(Borrow\ more\ than\ once\ after\ BDP_i = 1) = \alpha + \gamma X_i + \varepsilon_i \tag{5}$$

$$Total\ number\ of\ borrowings\ after\ BDP_i = \alpha + \gamma X_i + \varepsilon_i \tag{6}$$

In Eq. (5), the dependent variable *Borrow more than once after BDP* is a dummy variable that equals 1 if the user has borrowed twice or more after big data pricing and 0 otherwise. The result in Column (1) of Table 6 shows that users who are older, married, with lower education levels, longer work experience, lower income levels, and higher credit risk are more likely to borrow repeatedly.

In Eq. (6), the OLS model is employed to place the Probit model, and the dummy variable *Borrow more than once after BDP* is replaced by the discrete variable *Total number of borrowings after BDP*. The result remains mostly unchanged, except that in the OLS model, the coefficient of gender becomes significant at 5% level. It reveals that males are more likely to engage in borrowing activities multiple times on the marketplace lending platform, although female borrowers are found to exhibit lower default probability (Chen, Huang, & Ye, 2020) and are subject to higher success rates (Chen et al., 2016).

[4] $x = 1$ is omitted to avoid multicollinearity.

Table 4
Summary statistics of loan interest rate after big data pricing.

	APR(%)				N
	Mean	Std. Dev.	Min	Max	
Borrow 1st time after big data pricing	10.01	2.63	3	24.4	180,980
Borrow 2nd time after big data pricing	10.42	2.99	3	24.4	33,710
Borrow 3rd time after big data pricing	10.84	3.10	3	24.4	18,181
Borrow 4th time after big data pricing	10.88	3.14	3	24.4	9093
Borrow 5th time after big data pricing	11.09	3.21	3	24.4	5559
Borrow >5 times after big data pricing	11.19	3.44	6	24.4	8598

Note: This table reports descriptive statistics of APR of loans originated after big data pricing.

Table 5
Borrowing times and loan interest rate.

	(1)	(2)	(3)
	APR	APR	APR
Borrow n-th time	0.0119***	−0.000190	
	(5.82)	(−0.11)	
BDP		−2.212***	−2.185***
		(−310.09)	(−632.22)
BDP * Borrow n-th time		0.1200***	
		(8.25)	
Borrow 2nd time			0.1253***
			(8.36)
Borrow 3rd time			0.444***
			(7.88)
Borrow 4th time			0.679***
			(6.77)
Borrow 5th time			0.813***
			(5.78)
Borrow >5 times			0.829***
			(9.49)
Borrower and loan characteristics	Yes	Yes	Yes
N	408,914	408,914	408,914

Note: This table shows the regression results based on a sample of borrowers without delinquency or default records. Observations with missing variables are excluded. *BDP* is a dummy variable that equals 1 if the loan is originated after the pricing mechanism change and 0 otherwise. *Borrow n-th time* is a discrete variable, which equals the number of times that the user has borrowed on the platform. Robust t-statistics are reported in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level, respectively.

Table 6
Repeat borrowers and their characteristics.

	(1)	(2)
	Probit	OLS
	Borrow more than once after BDP	Total number of borrowings after BDP
Gender	0.0139	0.00136**
	(1.16)	(2.01)
Age	0.0140***	0.000804***
	(19.88)	(17.15)
Education	−0.135***	−0.00690***
	(−15.95)	(−12.10)
Income	−0.162***	−0.00818***
	(−35.26)	(−31.37)
Married	0.0349***	0.00194**
	(2.74)	(2.45)
Work experience	0.0979***	0.00488***
	(17.18)	(14.67)
Credit risk	0.324***	0.0994***
	(47.02)	(21.02)
Controls	Yes	Yes
N	250,636	250,636

Note: Robust t-statistics are reported in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level, respectively. The dependent variable *Borrow more than once after BDP* is a dummy variable that equals 1 if the user's total number of borrowing after big data pricing is >1 and it equals 0 otherwise.

Table 7
Borrowing experience before BDP and loan interest rate after BDP.

	(1)	(2)	(3)	(4)	(5)	(6)
	APR	APR	APR	APR	APR	APR
Average APR before BDP	0.121***	0.130***	0.0609**			
	(9.23)	(12.91)	(2.22)			
Average APR before BDP * Borrow n-th time after BDP			0.0288**			
			(2.49)			
Number of borrowing records before BDP				0.2261***	0.1405***	0.0676***
				(11.46)	(6.08)	(7.45)
Number of borrowing records before BDP * Borrow n-th time after BDP						0.0332**
						(2.22)
Borrow n-th time after BDP			−0.3434			0.1091
			(−0.47)			(1.03)
Control	No	Yes	Yes	No	Yes	Yes
N	6082	6036	6036	256,121	256,100	256,100

Note: The dependent variable APR is the loan interest rate after big data pricing. In Columns (1)–(3) the regressions are conducted at the borrower-level and only include those who have borrowed at least once before big data pricing. *Borrow n-th time* is a discrete variable, which equals the number of times that the user has borrowed on the platform. In Columns (4)–(6) the regressions are conducted at the loan-level. Robust t-statistics are reported in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level, respectively.

The above results are not surprising because the users with lower education levels and the users with lower levels of income are less likely to access formal financing channels such as banks, and therefore they are more likely to remain in the peer-to-peer lending market despite the unfavorable financing cost. Empirical research from other regions also shows that low-income households are less sensitive to interest rate changes (Gross & Souleles, 2002; Lukas, 2017) and borrowers with less education are more inelastic demand for loans (Bogan et al., 2015; Turvey et al., 2012). Bogan et al. (2015) additionally find that micro-entrepreneurs with larger monthly sales have more elastic demand. These findings, together with ours, are all consistent with the idea that borrowers with limited resources are willing to take higher interest rates.

Thus far, we have investigated whether repeat borrowers will be charged higher interest rates under the big data pricing mechanism. Next, we examine whether the borrowing records before big data pricing have an impact on loan pricing after the mechanism change. For those who borrowed before the pricing mechanism change, the platform has recorded their previously-accepted interest rates. Therefore, the marketplace lending platform has the motivation of aligning the interest rate of a new loan to previously-accepted rates which are higher in general. By doing so, there are two potential benefits for the platform - higher interest rates can not only attract more investors but also bring higher service fees.

To answer the above question, we develop two indicators for borrowing records before big data pricing. The first is the average interest rate of loans originated before big data pricing and the second is the number of times that the investor borrowed before big data pricing. The regression model is established in Eq. (7):

$$\begin{aligned} APR_j = & \alpha + \beta Borrowing\ record\ before\ BDP_i + \omega Borrowing\ record\ before\ BDP_i {*} Borrow\ nth\ time\ after\ BDP_j \\ & + \rho Borrow\ nth\ time\ after\ BDP_j + \gamma X_i + \varepsilon_i \end{aligned} \tag{7}$$

where the Borrowing record before BDP can be the two indicators we constructed.

Columns (1) and (2) of Table 7 show the regression results of excluding and including control variables. Both coefficients are positive and significant at 1% level, suggesting that the average loan interest rate before the big data pricing can predict the loan interest rate after the big data pricing. Next, we include the interaction term of *Average APR before BDP* and *Borrow n-th time after BDP*, and Column (3) shows that the coefficient is positive and significant as well. It is revealed that those who have borrowed at higher interest rates before the big data pricing are faced with even higher financing costs as they borrow more times.

In Column (4)–(6) of Table 7, the average APR is replaced with the number of borrowing records as an alternative measure of borrowing experience before big data pricing. It is shown that the coefficients of the *Number of borrowing records before BDP* across different specifications are all positive and significant at 1% level and the coefficient of the interaction term *Number of borrowing records before BDP * Borrow n-th time after BDP* is positive and significant at 5% level. It is shown that the previous number of borrowing records is an effective predictor of subsequent loan interest rates. Moreover, those who have borrowed multiple times before big data pricing encountered higher interest rates when they borrow repeatedly on this platform. The above results provide suggestive evidence that the marketplace lending platform may have utilized previously collected data to identify repeat borrowers and charge them higher interest rates.

5. Big data pricing and its effects on repayments

Thus far, it has been shown that big data pricing rejects some intended borrowers, charges higher interest to repeat borrowers, and provides lower interest rates to the rest. Next, we probe into the next question - is loan default ameliorated by big data pricing?

To answer this question, we develop the Probit model in Eq. (8) and regress *BDP* on the dummy variable *default*, which equals 1 if the loan went default. Additionally, we include the *Loan issuance date* in the model to control for the time trend. Time trend is critical

Table 8
Big data and loan default.

	(1)	(2)	(3)
	Default	Default	Default
	Logit	Probit	OLS
BDP	−0.150***	−0.271***	−0.0163***
	(−11.68)	(−12.08)	(−4.50)
Loan issuance date	0.00180***	0.00312***	0.000489***
	(58.70)	(58.19)	(59.54)
Loan characteristics	Yes	Yes	Yes
Borrower characteristics	Yes	Yes	Yes
N	239,889	239,889	239,889

Note: Robust t-statistics are reported in parentheses. *** indicates statistical significance at 1% level. BDP is a dummy variable that equals 1 if the loan is originated after the pricing mechanism change and 0 otherwise. The sample in this table includes the loans originated from 14 October 2014 to 14 October 2016.

when investigating borrower default because since 2013, a number of marketplace lending platforms go bankrupt every once in a while, and relevant news can bring vital influence to market sentiment. In order to further minimize the effects brought by other events over time, we only include loans originated from 14 October 2014 to 14 October 2016 in the regressions.

$$Pr\left(\mathrm{Default}_j = 1\right) = \alpha + BDP_j + Loan\ issuance\ date_j + \gamma X_i + \varepsilon_j \tag{8}$$

Table 8 exhibits the regression results. Column (1) shows Probit model results. In Column (2), we replace the Probit model with the Logit model. Column (3) shows the linear probability model. All three models show consistent results. After controlling for loan and borrower characteristics, loan issuance date has a significant negative influence on loan default probability. This is in accordance with the deteriorating market condition for peer-to-peer lending in China.

In contrast, the coefficient estimates of *BDP* are negative and significant at 1% level. This indicates that despite the degenerating market conditions over time, big data pricing has a positive effect on loan default.

This finding is in contrast with what has been found on Prosper that switching from auction to a platform-mandated price mechanism increases default probability (Wei & Lin, 2017). In Wei and Lin's work (Wei & Lin, 2017), propensity score matching is employed and the posted-price mechanism is regarded as the treatment group and those priced by auctions are in the control group. The reason why propensity score matching is not applied here is that marketplace lending in China had been undergone gradual changes including the declining interest rate and prolonged loan maturity. Even if we try to match auction-priced loans with platform-priced loans, it is unlikely to entirely eliminate the market differences (declining interest rate, changing market sentiment, competition among marketplace lending platforms, etc.) that the corresponding borrowers face and these factors can influence borrowers' willingness to repay.

6. Big data pricing and its effects on investors

It has been shown that loan-level interest rate declines after the adoption of big data pricing. We then proceed to portfolio-level, and establish the following model:

$$y_{i,t} = \alpha + \beta BDP + \gamma X_i + \varepsilon_{i,t} \tag{9}$$

where i denotes each investor and t represents before ($t = 0$) and after ($t = 1$) big data pricing. The dependent variable $y_{i,\ t}$ may be portfolio return or portfolio default rate. X_i is a strand of investor control variables. Our main coefficient of interest is β, which describes the relationship between portfolio performance and big data pricing.

Columns (1)–(3) of Table 9 show that, after controlling for investor characteristics, the portfolio return is on average decreased by 1.01%, significant at 1% level. Meanwhile, the portfolio default rate also declines by approximately 0.25%. In other words, the portfolios overall become less profitable as a result of big data pricing.

We further examine whether big data pricing has differential effects on investors with varied investment objectives. The following two models are developed. Eq. (10) investigates how big data pricing affects risk-takers and risk-avoiders differently. Eq. (11) investigates how big data pricing affects high-return investors and low-return investors differently. The coefficient of interest is θ.

$$APR_{i,t} = \alpha + \beta BDP_t + \theta BDP_t \bullet Risk_i + \gamma X_i + \varepsilon_{i,t} \tag{10}$$

$$Risk_{i,t} = \alpha + \beta BDP_t + \theta BDP_t \bullet APR_i + \gamma X_i + \varepsilon_{i,t} \tag{11}$$

Column (4) corresponds to eq. (10). The coefficient of the interaction term BDP * Portfolio risk is −0.43, which is statistically significant at 1%. Before big data pricing (BDP = 0), the regression coefficient of Portfolio risk on Portfolio return is 0.08, consistent with the economic intuition that high risk is associated with high return; After big data pricing (BDP = 1), the coefficient becomes 0.08–0.43 = −0.35 and the "high risk, high return" pattern no longer exists.

Columns (5) and (6) correspond to eq. (11) and exhibit similar results regardless of how the portfolio default is measured. In

Table 9
Big data pricing, portfolio performance and differential effects on investors.

	(1)	(2)	(3)	(4)	(5)	(6)
	Portfolio return	Portfolio default rate	Portfolio default rate (weighted by face value)	Portfolio return	Portfolio default rate	Portfolio default rate (weighted by face value)
BDP	−1.01***	−0.25***	−0.28***	−0.66***	−0.10***	−0.09***
	(−106.75)	(−45.17)	(−49.93)	(−33.28)	(−19.05)	(−18.50)
BDP * Portfolio risk				−0.43***		
				(−10.17)		
Portfolio risk				0.08***		
				(4.10)		
BDP * Portfolio return					−0.05***	−0.06***
					(−5.42)	(−4.17)
Portfolio return					0.03***	0.02**
					(6.99)	(2.00)
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes

Note: Robust t-statistics are reported in parentheses. ***, ** and * indicate statistical significance at 1%, 5% and 10% level, respectively. Portfolio return is the annual percentage return of an investor's portfolio. Portfolio default rate is the number of default notes divided by the total number of notes in the portfolio. Portfolio default rate (weighted by face value) is an alternative measure, which improves portfolio default rate by considering the face value of default notes. Portfolio risk is the average borrower credit risk of the loans in a portfolio, weighted by investment amount and holding time.

Column (5) and Column (6), the coefficients of BDP * Portfolio return are −0.05 and − 0.06, respectively. Before the adoption of big data pricing (BDP = 0), the regression coefficient of Portfolio return on Portfolio default is 0.03 in Column (5) and 0.02 in Column (6). Combined with the results in Column (4), it is shown that with the auction pricing mechanism the investors holding high-risk assets encounter more defaults but still earn a risk premium. After big data pricing, the coefficient becomes 0.03–0.05 = −0.02 in Column (5) and 0.02–0.06 = −0.04 in Column (6). In both Column (5) and Column (6), the coefficient sign is reversed from positive to negative. It means that after big data pricing, investors holding high-risk assets experience more defaults and no longer have a risk premium. In other words, the reduction of portfolio return brought by big data pricing is more salient for risk-takers.

Table 10 shows the summary statistics of portfolio return before and after pricing with big data. Portfolio return on average declines by 2.1%, which is in accordance with the findings from previous tables. It should be noted that the standard deviation abates drastically from 1.7% to 1.01%. The range of portfolio returns also shrinks, mainly due to the disappearance of abnormally high returns above 20.2%. Fig. 4 shows the distribution of investor portfolio return before and after big data pricing.

To sum up the above observations from Table 9, Table 10, and Fig. 4, we conclude that the adoption of big data pricing reduces the differences in portfolio performance among investors despite their distinct preferences. This is consistent with the previous finding that highly sophisticated investors, identified by whether engaged in credit arbitrage activities, experience a greater reduction in portfolio return after big data pricing, compared with the rest of investors (Tian, Wang, & Wu, 2021) and that the extraordinary performance of the sophisticated investors shrinks when platform reduces information provision to the investors (Vallée & Zeng, 2019).

7. Discussion

Big data and novel ways of applying big data technology are transforming financial markets and have attracted the attention of some economists. For example, Cerchiello and Giudici (2016) present a systemic risk model based on big data stemming from financial tweets and Begenau, Farboodi, and Veldkamp (2018) establish a simple model in which investors gather information about firms, and the amount of information is positively related to the size of the firm. However, it remains unclear how financial companies have been utilizing big data. In this work, we present some empirical results from a marketplace lending platform.

In a marketplace lending platform, the pricing mechanism changed from auction to big data pricing. By analyzing the data before and after this change, it is found that: (1) big data pricing improves funding success rate; (2) interest rates are generally lower after pricing with big data but returning borrowers and the borrowers with default records are faced with higher interest rates; (3) after big data pricing, the investor return is lower and more converged; Investors that undertake higher risk face a greater reduction in portfolio return.

The first observation is related to the work of Franks, Serrano-Velarde, and Sussman (2020) which attributes information inefficiency on Funding Circle to liquidity shocks arising from the mismatch of funds that flow into and out of the platform. Franks et al. (2020) believe that such a mismatch is a result of algorithmic "autobid" designed to delegate the orders of passive investors who lack the time and expertise to participate in auctions. Our first observation may be due to a similar rationale that, by using big data pricing the platform dynamically adjusts the assigned loan interest rate to an equilibrium price such that the market can clear quickly. Moreover, it is believed that a peer-to-peer platform is a substitute for not only another peer-to-peer platform (Mariotto, 2016) but also microfinance companies (Wang, Yu, Yang, & Zhang, 2021) and banks (Tang, 2019) because they can provide similar microcredits. Since China is one of the most prosperous markets for marketplace lending and microfinance institutions (Britzelmaier, Kraus, & Xu,

Table 10
Portfolio return before and after big data pricing.

Portfolio return	N	Mean	Std. Dev.	Min	Max
Before pricing with big data	366,620	12.8%	1.70%	8.9%	26.3%
After pricing with big data	366,620	10.7%	1.01%	8.9%	20.2%

Notes: The difference between the two means is 2.1%, statistically significant at 1%.

Fig. 4. Distribution of portfolio return.

2013), the competition among these financial institutions provides borrowers with bargaining power. The adoption of big data pricing may serve as a tool to flexibly adjust platform interest rates to make it more attractive to borrowers.

Our second observation is twofold. On one hand, it seems that the big data technique delivers its job of risk identification by offering higher prices to less creditworthy borrowers. On the other hand, it charges the repeated borrowers with a higher financing price despite that these repeated borrowers have been repaying their debts on time. Before the era of Fintech, economists focus more on the second- and third-degree of price discrimination because, in reality, first-degree discrimination is almost impossible to achieve. However, big data has made personalized price discrimination possible. The debate follows - whether it is ethical and legal to charge high willingness-to-pay consumers higher prices. Financial products and services such as loans are special because people that are faced with high interest rates in the marketplace lending market are often those who do not have access to bank loans and regard marketplace lending as the last resort, let alone the fact that these repeat borrowers have not shown any default records.

In this regard, China regulatory departments have issued “Provisions on Internet information service recommendation algorithm”,[5] effective as of March 1, 2022. It is regulated that if an algorithm is used to recommend products or services to consumers, the preferences and habits of consumers should not be used in the algorithm to avoid unreasonable price discrimination. Under this provision, it might be easier to eliminate price discrimination in the product market but it remains difficult to identify price discrimination in the credit market due to the nature of this market, which contains the following two aspects. First, the credit market is built upon “credit” - borrowers with punctual repayment records are rated with higher credit scores and can enjoy higher credit limits. In other words, the utilization of customers’ past information is the foundation of credit market. Second, even if the pricing of loans is solely based on demographic information, it is still difficult to make loan prices entirely transparent to potential borrowers because the pricing models are often extremely complex and are updated dynamically according to market conditions.

Our third observation reveals a potential drawback of big data pricing in this case. Due to initial screening by big data technology, a market with homogenous loans cannot provide the investors with different types of asset choices and this may hurt investor return and willingness to participate.

[5] For the full text, please refer to http://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm

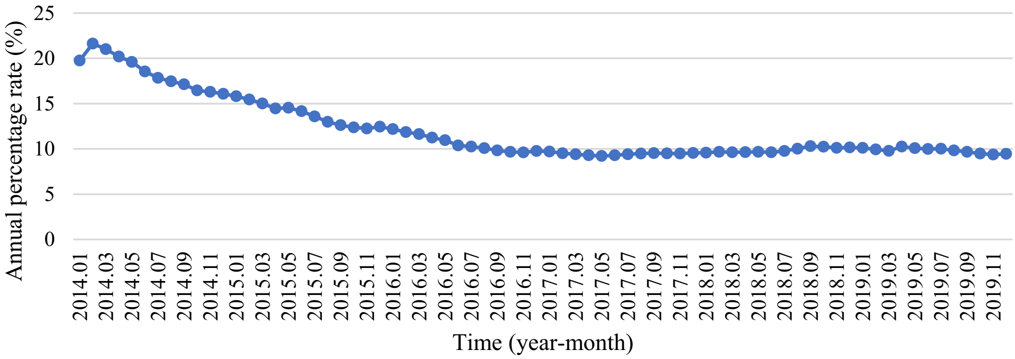

Fig. A1. Average annual percentage rate of marketplace lending in China
Data source: Home of Marketplace Lending (Wang Dai Zhi Jia), URL: https://www.wdzj.com.

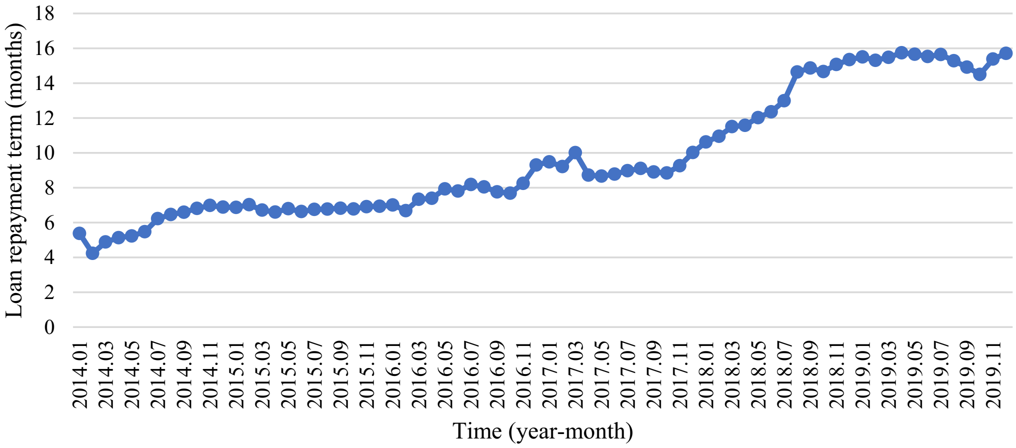

Fig. A2. Loan repayment term of marketplace lending in China.
Data source: Home of Marketplace Lending (Wang Dai Zhi Jia), URL: https://www.wdzj.com.

Data availability

Data will be made available on request.

Acknowledgments

Geran Tian gratefully acknowledges financial support by Research Initiation Fund Project for New Youth Teachers (grant number: XRZ2023048) from Capital University of Economics and Business. Weixing Wu gratefully acknowledges financial support from National Social Science Fund of China (Grant No. 16ZDA033).

References

Bar-Gill, O. (2019). Symposium: Algorithmic Price discrimination when demand is a function of both preferences and (Mis)perceptions. *The University of Chicago Law Review, 86*(2), 217–254.

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the Fintech era. *Journal of Financial Economics*. https://doi.org/10.1016/j.jfineco.2021.05.047

Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics, 97*, 71–87. https://doi.org/10.1016/j.jmoneco.2018.05.013

Bell, C., Srintvasan, T. N., & Udry, C. (1997). Rationing, spillover, and interlinking in credit markets: The case of rural Punjab. *Oxford Economic Papers, 49*(4), 557–585. https://doi.org/10.1093/oxfordjournals.oep.a028625

Benston, G. J. (1964). Commercial Bank Price discrimination against small loans: An empirical study. *The Journal of Finance, 19*(4), 631. https://doi.org/10.2307/2977116

Black, H. A., Boehm, T. P., & DeGennaro, R. P. (2003). Is there discrimination in mortgage pricing? The case of overages. *Journal of Banking & Finance, 27*(6), 1139–1165. https://doi.org/10.1016/s0378-4266(02)00250-9

Bogan, V. L., Turvey, C. G., & Salazar, G. (2015). The elasticity of demand for microcredit: Evidence from Latin America. *Development and Policy Review, 33*(6), 725–757. https://doi.org/10.1111/dpr.12131

Botta, M., & Wiedemann, K. (2019). To discriminate or not to discriminate? Personalised pricing in online markets as exploitative abuse of dominance. *European Journal of Law and Economics, 50*(3), 381–404. https://doi.org/10.1007/s10657-019-09636-3

Britzelmaier, B., Kraus, P., & Xu, Y. (2013). Microfinance institutions in China: Development and challenges to their sustainability. *Strategic Change, 22*(1–2), 67–78. https://doi.org/10.1002/jsc.1922

Cerchiello, P., & Giudici, P. (2016). Big data analysis for financial risk management. *Journal of Big Data, 3*(1). https://doi.org/10.1186/s40537-016-0053-4
Chen, D., Li, X., & Lai, F. (2016). Gender discrimination in online peer-to-peer credit lending: Evidence from a lending platform in China. *Electronic Commerce Research, 17*(4), 553–583. https://doi.org/10.1007/s10660-016-9247-2
Chen, X., Huang, B., & Shaban, M. (2020). Naïve or sophisticated? Information disclosure and investment decisions in peer to peer lending. *Journal of Corporate Finance, 101805*. https://doi.org/10.1016/j.jcorpfin.2020.101805
Chen, X., Huang, B., & Ye, D. (2018). The role of punctuation in P2P lending: Evidence from China. *Economic Modelling, 68*, 634–643. https://doi.org/10.1016/j.econmod.2017.05.007
Chen, X., Huang, B., & Ye, D. (2020). Gender gap in peer-to-peer lending: Evidence from China. *Journal of Banking & Finance, 112*, Article 105633. https://doi.org/10.1016/j.jbankfin.2019.105633
Conitzer, V., Taylor, C. R., & Wagman, L. (2012). Hide and seek: Costly consumer privacy in a market with repeat purchases. *Marketing Science, 31*(2), 277–292. https://doi.org/10.1287/mksc.1110.0691
Courchane, M., & Nickerson, D. (1997). Discrimination resulting from overage practices. *Journal of Financial Services Research, 11*, 133–151. https://doi.org/10.1007/978-1-4615-6147-7_8
Ding, C., Kavuri, A. S., & Milne, A. (2020). Lessons from the rise and fall of Chinese peer-to-peer lending. *Journal of Banking Regulation, 22*(2), 133–143. https://doi.org/10.1057/s41261-020-00132-2
Dong, C. (2017). Is posted-pricing mechanism better? Evidence from online lending market. *Operations Research Management Science, 26*, 8 (In Chinese).
Elliehausen, G. E., & Lawrence, E. C. (1990). Discrimination in consumer lending. *The Review of Economics and Statistics, 72*(1), 156. https://doi.org/10.2307/2109753
Ferguson, M., & Peters, S. (1995). What constitutes evidence of discrimination in lending? *The Journal of Finance, 50*(2), 739–748. https://doi.org/10.1111/j.1540-6261.1995.tb04803.x
Franks, J., Serrano-Velarde, N., & Sussman, O. (2020). Marketplace lending, information aggregation, and liquidity. *The Review of Financial Studies, 34*(5), 2318–2361. https://doi.org/10.1093/rfs/hhaa101
Fudenberg, D., & Villas-Boas, J. M. (2006). Behavior-based price discrimination and customer recognition. In T. Hendershott (Ed.)*, Vol. 1. Economics and information systems, handbooks in information systems* (pp. 377–436). Amsterdam: Elsevier.
Gillis, T. B., & Spiess, J. L. (2019). Big data and discrimination. *The University of Chicago Law Review, 86*, 459–488.
Gross, D. B., & Souleles, N. S. (2002). Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data. *The Quarterly Journal of Economics, 117*(1), 149–185. https://doi.org/10.1162/003355302753399472
Han, S. (2011). Creditor learning and discrimination in lending. *Journal of Financial Services Research, 40*(1–2), 1–27. https://doi.org/10.1007/s10693-011-0101-3
Harkness, S. K. (2016). Discrimination in lending markets. *Social Psychology Quarterly, 79*(1), 81–93. https://doi.org/10.1177/0190272515623459
Hau, H., Hoffmann, P., Langfield, S., & Timmer, Y. (2021). Discriminatory pricing of over-the-counter derivatives. *Management Science, 67*(11), 6660–6677.
Karlan, D. S., & Zinman, J. (2008). Credit elasticities in less-developed economies: Implications for microfinance. *American Economic Review, 98*(3), 1040–1068. https://doi.org/10.1257/aer.98.3.1040
Kau, J. B., Keenan, D. C., & Munneke, H. J. (2012). Racial discrimination and mortgage lending. *The Journal of Real Estate Finance and Economics, 45*, 289–304. https://doi.org/10.1007/s11146-011-9330-3
Kochar, A. (1997). An empirical investigation of rationing constraints in rural credit markets in India. *Journal of Development Economics, 53*(2), 339–371. https://doi.org/10.1016/s0304-3878(97)00020-5
Ladd, H. F. (1998). Evidence on discrimination in mortgage lending. *Journal of Economic Perspectives, 12*(2), 41–62.
Li, X., Deng, Y., & Li, S. (2020). Gender differences in self-risk evaluation: Evidence from the Renrendai online lending platform. *Journal of Applied Economics, 23*(1), 485–496. https://doi.org/10.1080/15140326.2020.1797338
Li, Y., So, J., & Yuan, J. (2018). Voluntary verifiable information disclosure and loan funding performance: Evidence from Paipaidai in China. *The Singapore Economic Review, 65*(02), 419–441. https://doi.org/10.1142/s0217590818500066
Li, Y., Zhao, C., & Zhong, Z. (2019). Price discrimination against retail investors: Evidence from mini options. *Journal of Banking & Finance, 106*, 50–64. https://doi.org/10.1016/j.jbankfin.2019.05.012
Liu, W., Long, S., Xie, D., Liang, Y., & Wang, J. (2021). How to govern the big data discriminatory pricing behavior in the platform service supply chain? An examination with a three-party evolutionary game model. *International Journal of Production Economics, 231*, Article 107910. https://doi.org/10.1016/j.ijpe.2020.107910
Lukas, M. (2017). Estimating interest rate elasticities in consumer credit. *Economics Letters, 156*, 155–158. https://doi.org/10.1016/j.econlet.2017.05.004
Mariotto, C. (2016). Competition for lending in the internet era: The case of peer-to-peer lending marketplaces in the USA. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2800998
Marshall, G. (2020). Search and wholesale price discrimination. *The Rand Journal of Economics, 51*(2), 346–374. https://doi.org/10.1111/1756-2171.12317
Maxwell, S., & Garbarino, E. (2010). The identification of social norms of price discrimination on the internet. *The Journal of Product and Brand Management, 19*(3), 218–224. https://doi.org/10.1108/10610421011046193
Meyer, P. A. (1967). Price discrimination, regional loan rates, and the structure of the banking industry. *The Journal of Finance, 22*(1), 37. https://doi.org/10.2307/2977298
Pérez-Martín, A., Pérez-Torregrosa, A., & Vaca, M. (2018). Big data techniques to measure credit banking risk in home equity loans. *Journal of Business Research, 89*, 448–454. https://doi.org/10.1016/j.jbusres.2018.02.008
Puller, S. L., & Taylor, L. M. (2012). Price discrimination by day-of-week of purchase: Evidence from the U.S. airline industry. *Journal of Economic Behavior & Organization, 84*(3), 801–812. https://doi.org/10.1016/j.jebo.2012.09.022
Shiller, B. (2016). *Personalized Price discrimination using big data*. Working Papers 108,. Brandeis University, Department of Economics and International Business School.
Tang, H. (2019). Peer-to-peer lenders versus banks: Substitutes or complements? *The Review of Financial Studies, 32*(5), 1900–1938. https://doi.org/10.1093/rfs/hhy137
Tian, G., Wang, X., & Wu, W. (2021). Borrow low, lend high: Credit arbitrage by sophisticated investors. *Pacific-Basin Finance Journal, 67*, Article 101544. https://doi.org/10.1016/j.pacfin.2021.101544
Townley, C., Morrison, E., & Yeung, K. (2017). Big data and personalized price discrimination in eu competition law. *Yearbook of European Law, 36*, 683–748. https://doi.org/10.1093/yel/yex015
Turvey, C. G., He, G., Ma, J., Kong, R., & Meagher, P. (2012). Farm credit and credit demand elasticities in Shaanxi and Gansu. *China Economic Review, 23*(4), 1020–1035. https://doi.org/10.1016/j.chieco.2012.05.004
Vallée, B., & Zeng, Y. (2019). Marketplace lending: A new banking paradigm? *The Review of Financial Studies, 32*(5), 1939–1982. https://doi.org/10.1093/rfs/hhy100
Wang, B., Yu, Y., Yang, Z., & Zhang, X. (2021). Microfinance institutions and peer-to-peer lending: What does microfinance competition bring? *Pacific-Basin Finance Journal, 67*, Article 101557. https://doi.org/10.1016/j.pacfin.2021.101557
Weersink, A., Vanden Dungen, M. J., & Turvey, C. G. (1994). Estimating the demand for farm operating and term credit. *Cahiers D'Economie Et Sociologie Rurales, 33*(1), 97–116. https://doi.org/10.3406/reae.1994.1433
Wei, Z., & Lin, M. (2017). Market mechanisms in online peer-to-peer lending. *Management Science, 63*(12), 4236–4257. https://doi.org/10.1287/mnsc.2016.2531
Xinhua. (2021, August 13). Draft Chinese law targets excessive collection of personal data by apps. http://www.xinhuanet.com/english/2021-08/13/c_1310125736.htm.