

# Default Data Manipulation in Marketplace Lending

Yutong Hu<sup>LBS</sup> and Shasha Li<sup>IWH Halle and OVGU</sup>  
ZEW 2024

June 2024

# Motivation

- Puzzling low default rates in **online lending platforms**, e.g. Zopa 0.5% since 2005.
  - Online lending platform: individuals directly participate, mainly previously unbanked group;
  - Large amount of transaction activity data including **repayment performance (loan default)**;
  - Platforms can decide what information to disclose.
- Call for transparency: lack platform information disclosure regulation.
  - In 2016, the U.S. Treasury Department released a white paper to call for greater transparency in online lending;
  - In 2019, EU introduced transparency obligations for online platforms.

# Research Question

## This paper

- ① Detect data manipulation on a typical online lending platform
  - Find that the platform hid actual default rates (bad loan repayments)
- ② Impacts of default data hiding
  - on market efficiency?
    - Can interest rates predict default rates?
  - on market stability?
    - How do loan transactions respond to dramatic negative news in the P2P industry? (Event study, 2015 Ezubao Scandal)

**Data:** 2010-2017, detailed loan-level data from a leading Chinese online lending platform

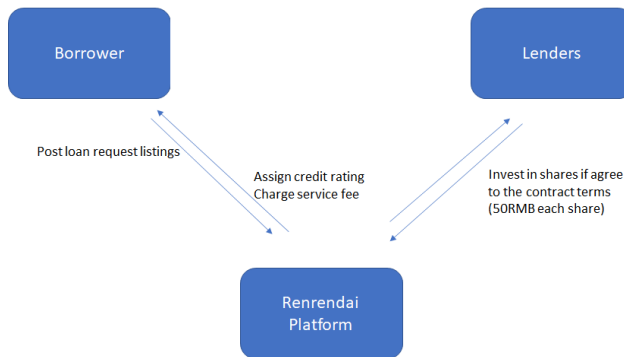
## Default

90+ days payments overdue on any repayment

# Renrendai Platform

Loan application, repayment, and transfer, online

- Founded in May 2010, one leading online marketplace lending platform in China
- Information Intermediary: charge service fees tied to the trading volume



# Data

## Sample

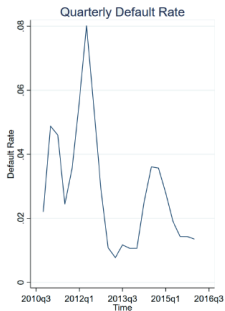
- 1,033,465 loan applications on Renrendai platform, Oct 2010 - Jan 2017.
- Observe monthly repayment performance until Sep 2018.
- Observe loan transfer data from Oct 2013 to Sep 2018.
- 40% of the total applications were successful.

**Table:** Summary Statistics: Loan Characteristics

Variable	Mean	Median	SD	Min	Max
A. Loan characteristics					
Loan size (applied) (thousand yuan)	68.9	50	73.9	0	3000
Loan size (granted) (thousand yuan)	75	66.7	51.8	1	3000
Maturity (months)	31	36	9	1	48
Annual Interest rate (%)	10.51	10.2	1.30	3	24.4

# Loan default data hiding

- 2012 Nov: a sudden drop of loan default rates.
- What happens?
  - Renrendai collaborated with its sister company, Ucredit, to **integrate offline sourced loan applications ("field" type) into the online market.**
  - The default rates for "field" type loans consistently show as zero on the online market.
- How to hide?
  - Ucredit's safeguard fund covers offline defaults; online records show no delinquency for such cases.



# Empirical Analysis: estimate underreported default rates

Method: DID, introductions of Ucredit offline branches, city-year month level

- Set the introduction time of Ucredit offline branches as the event time. Take 6-month pre- and 6-month post-period windows. Run DID regression below,

$$\begin{aligned} Default_{it} = & \beta_0 + \beta_1 \times Treat_{ij} \times Post_{ij} + \beta_2 \times Post_{ij} + \beta_3 \times Treat_{ij} \\ & + \beta_4 \times X_1^{Borrower}_{it} + \beta_5 \times X_2^{Loan}_{it} + \beta_6 \times X_3^{other}_{ij} + \eta_{it} \end{aligned}$$

- $Treat_{ij}$  equals 1 if observation  $i$  is in the treatment group regarding offline branch  $j$ ;
- $Post_{ij}$  equals 1 if loan  $i$  is originated after the corresponding offline branch  $j$  opens.
- Estimation results: under-report monthly default rate by 5.26% (6-m window).

# Empirical Analysis: Market Efficiency

Method, stacked regression

- If the market is efficient,  $\pi_i^e \approx \alpha^* + 1 \times \frac{r_i}{LGD_i^e}$  (Franks et. al, 2021); derived from lenders' participation constraint,  $1 + \rho = (1 - \pi_i^e)(1 + r_i) + \pi_i^e(1 - LGD_i^e)(1 + r_i)$ .
- Run regressions below

$$Default_{it} = \beta \times r_i^* \times Post + \kappa r_i^* + \eta Post + \theta \times FE\_Rating_i + \gamma \times FE\_Mlssue_i \\ + \lambda \times FE\_SLife_{it} + \delta \times X_{it} + \epsilon_{it}$$

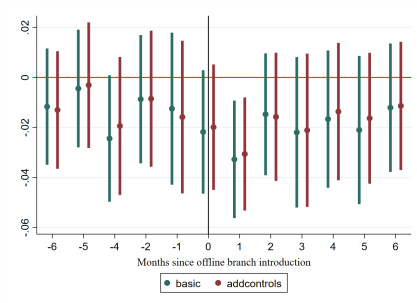
- *Post* equals 1 if the borrower's working city has an offline branch opening.
- If the market is efficient ex-ante:  $\kappa = 1$ .
- If the market is efficient after mixing with offline loan data:  $\kappa + \beta = 1$ .
- *FE\_Rating<sub>i</sub>* and *FE\_Mlssue<sub>i</sub>*: fixed effects (FEs) for credit ratings and issuing month of loan *i*.
- *FE\_SLife<sub>it</sub>*: fixed effects of loan *i*'s life cycle, into three equal stages, Early, Mid and Late.



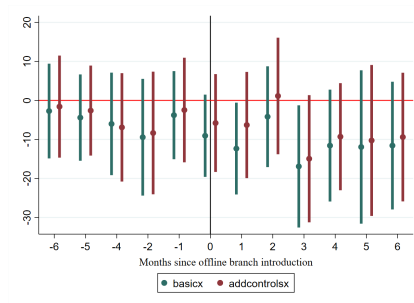
# Empirical Analysis: Market Efficiency

## Test Results

- $\kappa$  is positive, close to zero (the market is not efficient ex-ante).
- Data hiding moves the market further away from the efficiency ( $\beta < 0$ ).
- A decline in borrower credit quality.



a. Market Efficiency Test



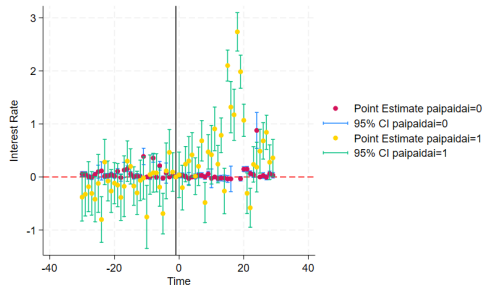
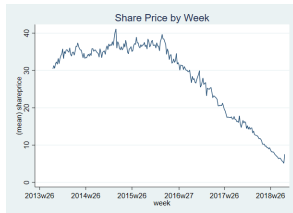
b. Latest Credit Score

# Empirical Analysis: Market Stability

## 2015 Dec Ezubao Scandal Shock

In 2015 Dec, Ezubao, a major P2P platform, was reported a \$7.6 billion Ponzi scheme and shocked online lending industry.

- Panic run evident in Renrendai's secondary loan market prices;
- Comparing to Paipaidai (no default data hiding), Renrendai is fragile. Renrendai's interest rates increased with the first negative news.



Event Study “Ezubao” scandal, Paipaidai vs Renrendai  
2015 Dec 16 news peak (time 0); 2015 Dec 04 first news

# Takeaways

- We find statistical evidence showing that the online platform under-reports monthly default rates by at least 5.26% on average.
- The default data manipulation drives price away from market efficiency and attracts lower quality borrowers.
- The market is fragile to the external negative shock.
- Regulation on information disclosure for Fintechs/digital platforms is important.

Thanks!

# Appendix. Empirical Analysis: market efficiency

## Method

- Following Franks, Serrano-Velarde and Sussman [2021]:
- In a risk-neutral world, lenders face the following participation constraint:

$$1 + \rho = (1 - \pi_i^e)(1 + r_i) + \pi_i^e(1 - LGD_i^e)(1 + r_i)$$

- $\rho$ , risk-free rate
- $\pi_i^e$ , loan  $i$ 's probability of default
- Loss Given Default (LGD)
- Rewrite, we get

$$\pi_i^e \approx \alpha^* + 1 \times r_i^*$$

- where  $\alpha^* = -\frac{\rho}{LGD_i^e}$  and  $r_i^* = \frac{r_i}{LGD_i^e}$ .