Predicting prepayment and default risks of unsecured consumer loans in online lending

Zhiyong Li

School of Finance, Southwestern University of Finance and Economics, China

Ying Tang

Southwestern University of Finance and Economics, China

Outline

- Introduction (what is P2P lending?)
- P2P literature (what have been done in academics?)
- Basic stats (what we observed?)
- Previous research on prepayment (Why prepayment matters?)
- Method (how we model?)
- Data (what we've got?)
- Results (what we found?)
- Conclusions

Introduction

- What is P2P lending?

Peer-to-Peer (People-to-People) lending is the practice of investors lending money to individuals or businesses through online services that match lenders directly with borrowers. Since the P2P lending companies offering these services operate entirely online, they can run with lower overhead and provide the service more cheaply than traditional financial institutions. As a result, lenders often earn higher returns compared to savings and investment products offered by banks, while borrowers can borrow money at lower interest rates, even after the P2P lending company has taken a fee for providing the match-making platform and credit checking the borrower.

P2P in academics

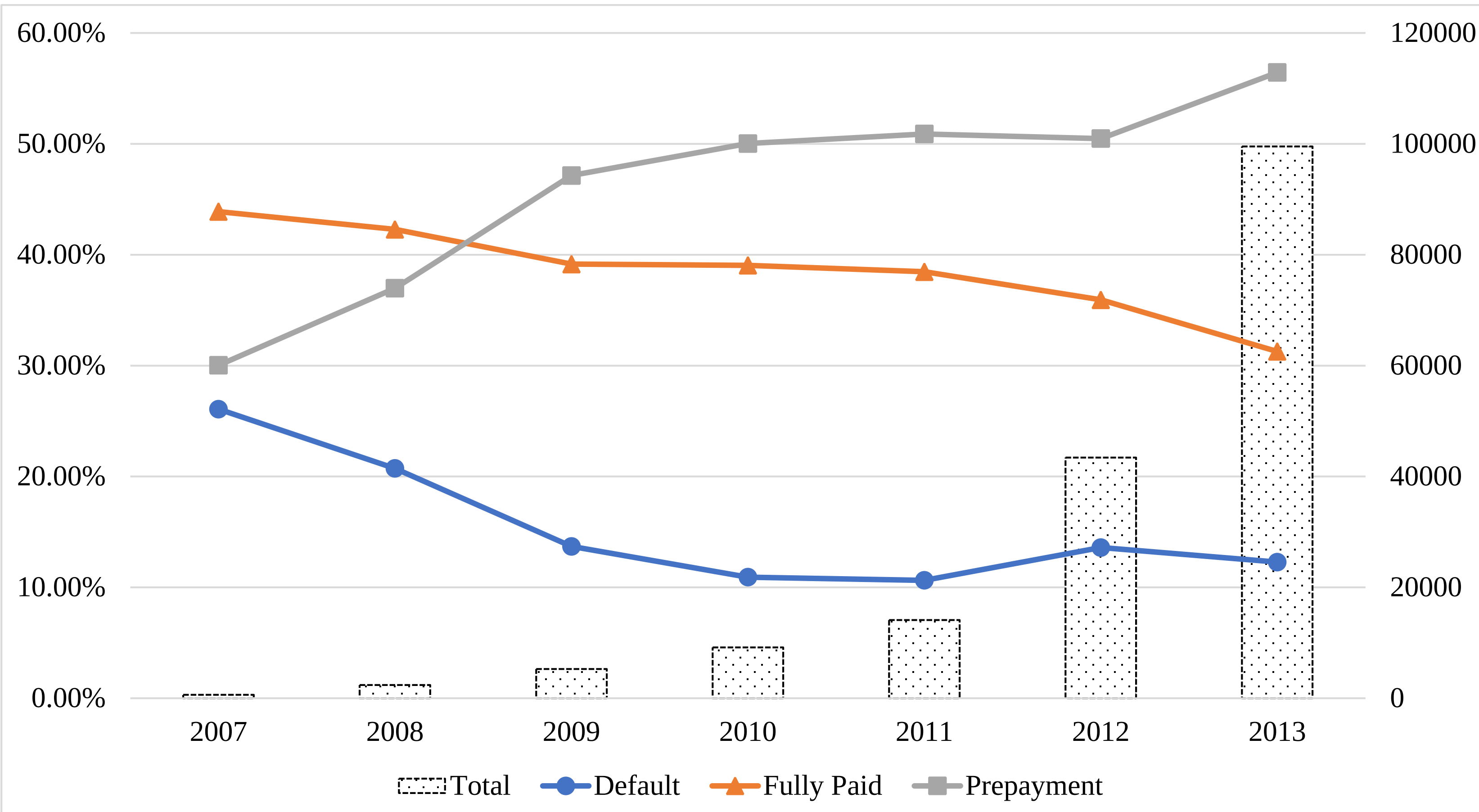
- NINE top journal papers (4 STAR)

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: the role of appearance in peer-to-peer lending. ***Review of Financial Studies***, 25(8), 2455-2484.
Zhang, J., & Liu, P. (2012). Rational herding in microloan markets. ***Management Science***, 58(5), 892-912.
Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. ***Management Science***, 59(1), 17-35.
Rigbi, O. (2013). The effects of usury laws: Evidence from the online loan market. ***Review of Economics and Statistics***, 95(4), 1238-1248.
Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2015). Credit Scoring with Social Network Data. ***Marketing Science***, forthcoming
Liu, D., Brass, D., Lu, Y., & Chen, D. (2015). Friendships in online peer-to-peer lending: Pipes, prisms, and relational herding. ***MIS Quarterly***, 39(3), 729-742.
Lin, M., & Viswanathan, S. (2015). Home bias in online investments: An empirical study of an online crowdfunding market. ***Management Science***, forthcoming
Miller, S. (2015). Information and default in consumer credit markets: Evidence from a natural experiment. ***Journal of Financial Intermediation***, 24(1), 45-70.
Iyer, R., Khwaja, A. I., Luttmer, E. F., & Shue, K. (2015). Screening peers softly: Inferring the quality of small borrowers. ***Management Science***, forthcoming

P2P literature

Paper	Data source	Sample size	Data period
Duarte et al (2012)	Prosper	5,950	2006-2008
Zhang & Liu (2012)	Prosper	49,693	2006-2008
Lin et al (2013)	Prosper	56,584	2007-2008
Rigbi (2013)	Prosper	114,902 listings, 9,969 loans	2007-2008
Wei et al (2015)	No data no scoring		
Liu et al (2015)	PPDai	12,514	2009-2011
Lin & Viswanathan (2015)	Prosper	29,422	2008-2011
Miller (2015)	Prosper	12,117	2006
Iyer et al (2015)	Prosper	194,033 listings, 17,212 loans	2007-2008

Outcomes of P2P loans



Outcomes of P2P loans

- What we found?

Volumes increases dramatically in recently years (institutional investors enter this market too!);

High default risk;

Even higher proportions of early repayment

- What is default?

“a borrower failing to meet its obligations in accordance with agreed terms ”(Basel Committee on Bank Supervision,   2002)

Prepayment

- Consequence of prepayment?

Loss of interests

- Research on prepayment?

Most literature is from the secured consumer loan market where collaterals are involved, such as mortgages and auto loans.

Subprime Automobile loans: Heitfield and Sabarwal (2004)

Mortgages: Ciochetti et al. (2002), Ciochetti et al. (2003) and Pennington-Cross (2010)

Prepayment

- **Option Theory**

Deng, Y., Quigley, J.M., van Order, R., 2000. Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options. **Econometrica** 68, 275-307.

Deng et al. (2000) tested the option theory in the mortgage market, where the property holder can excise the call option by early repayment and refinancing the mortgage, if the market value of the property exceeds its original value; and further, that a put option can be excised by defaulting on the mortgage, if the market value goes below the original value. They also commented that transaction costs would be a significant factor in the excising decision. In commercial mortgages, creditors usually implement penalty clauses to prevent refinancing in the case of prepayment (Steinbuks, 2015; Varli and Yildirim, 2015).

However, in online lending markets, prepayment generates no fee, unlike a borrower's transaction cost, so the prepayment rate is relatively very high.

Prepayment

- What is competing risks?

The method used to differentiate causes of termination is rather vividly called ‘competing risks’, whereby all possible events which could lead to the exit of an account compete against one another to be the first to happen. Whichever happens first may halt or indeed stop other events from occurring. Competing risks are common in studying mortality rates. In the credit scoring context, default, attrition, prepayment and closure will all lead to a borrower stopping using the existing account.

Modelling method: Survival Analysis

In this paper, we used multinomial logistic regression.

Method

- Multinomial logistic regression?

Multinomial logistic regression treats the outcome as a discrete choice variable. It assumes mutual independence of choices for a given record during an observation period.

$$\ln\left(\frac{P(d_i = j) \mid \mathbf{x}_i)}{P(d_i = 0) \mid \mathbf{x}_i)}\right) = \boldsymbol{\beta}_j^T \mathbf{x}_i, \quad j = 1, 2$$

$$P(d_i = 0) \mid \mathbf{x}_i) = \frac{1}{1 + \sum_{j=1}^{2} \exp(\boldsymbol{\beta}_j^T \mathbf{x}_i)}$$

$$P(d_i = j) \mid \mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}_j^T \mathbf{x}_i)}{1 + \sum_{j=1}^{2} \exp(\boldsymbol{\beta}_j^T \mathbf{x}_i)}$$

Finally, for a given record the predicted category is found to be

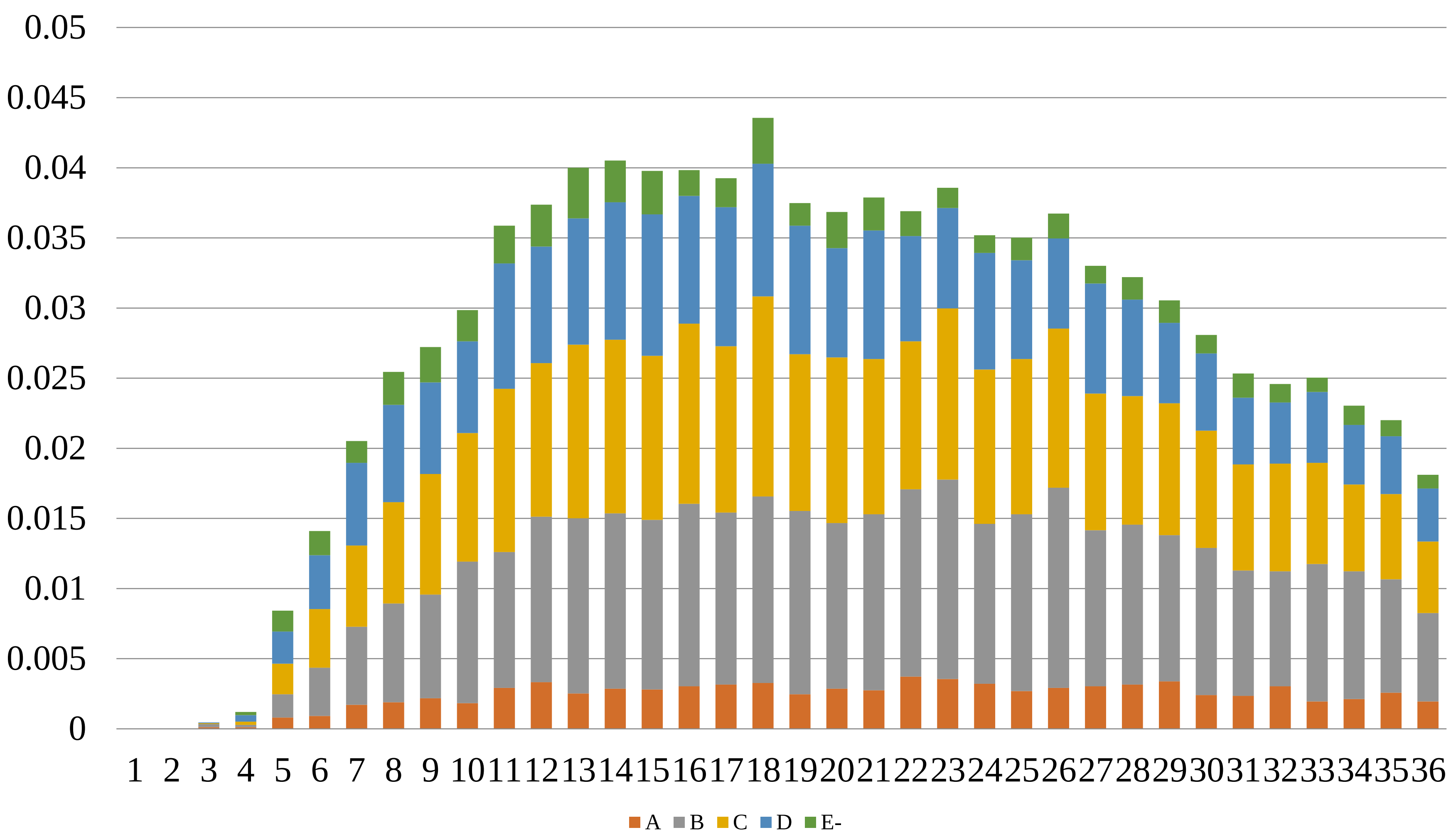
$$\hat{j} = \arg\max\{\hat{P}(d_i = j), j = 0, 1, 2\}$$

Data

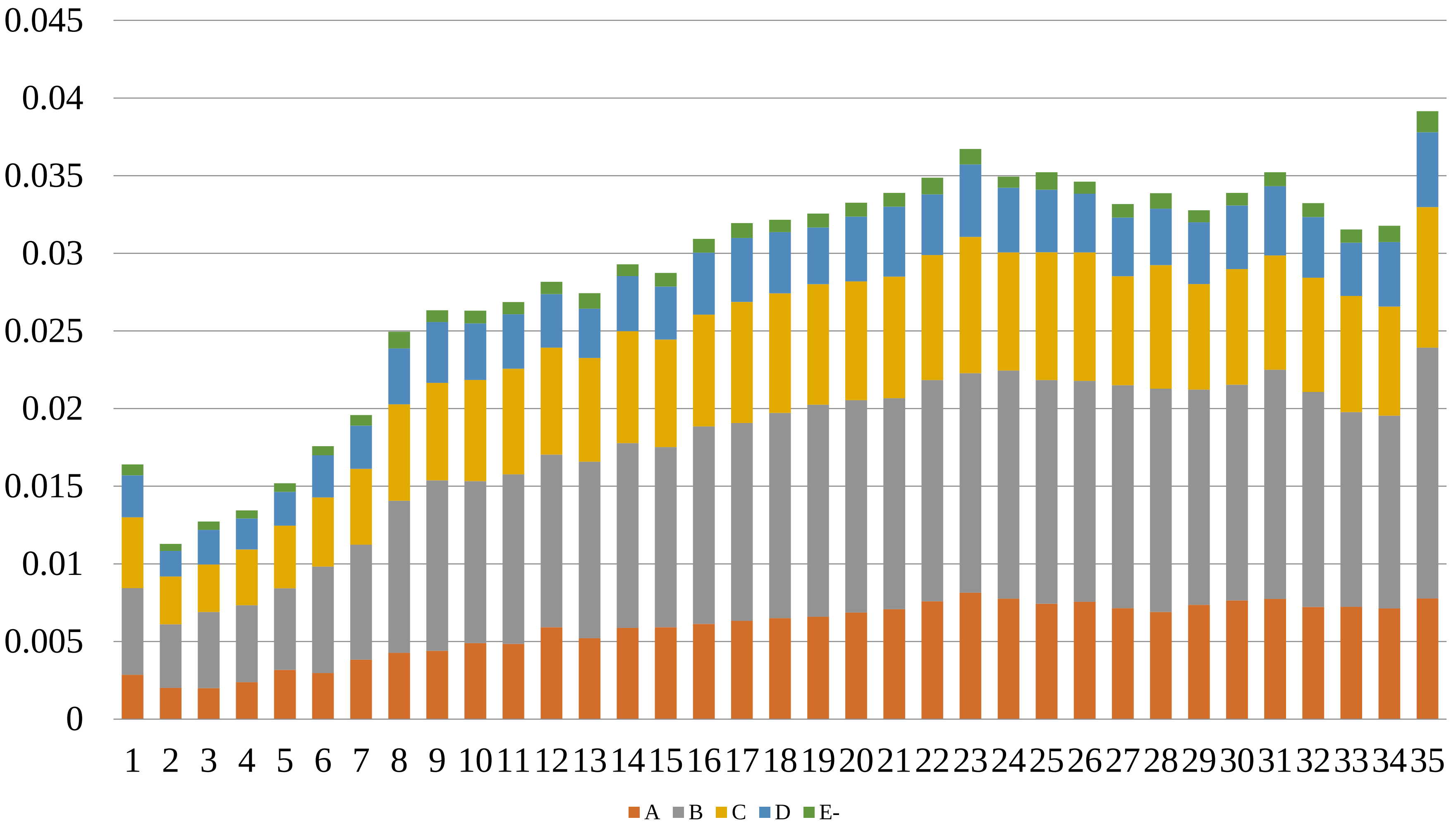
- 140,605 unsecured consumer loans from a P2P market

	A	B	C	D	E-	Total
Default	5.32%	10.28%	15.68%	20.39%	23.59%	17449
Fully Paid	36.36%	33.26%	29.60%	27.95%	26.63%	45074
Prepayment	58.31%	56.45%	54.71%	51.66%	49.78%	78082
Total	27414	55969	33615	18995	4612	140605

Data(Month on Book: Default)



Data (Month on Book: Prepayment)



Variables

	Default		Fully paid		Prepayment	
	Mean	Std	Mean	Std	Mean	Std
Borrower characteristics						
Debt to income	18.020	7.659	16.885	7.620	16.281	7.539
Employment length	5.657	3.607	6.013	3.615	5.740	3.615
Open credit lines	10.885	4.618	10.770	4.540	10.890	4.576
Total credit lines	22.945	10.853	22.876	10.650	24.716	11.171
Revolving credit utilisation	0.609	0.222	0.588	0.227	0.562	0.233
Public record	0.107	0.371	0.094	0.352	0.115	0.381
Last FICO	618.188	61.117	697.914	59.337	709.882	46.156
Tax lines	0.013	0.176	0.012	0.172	0.014	0.173
Charge off within 12 months	0.005	0.082	0.003	0.063	0.006	0.089
Public record bankruptcies	0.087	0.292	0.075	0.272	0.093	0.305
Inquiries 6 months	0.933	1.082	0.678	0.951	0.800	1.022
Loan features						
Interest rate	14.803	3.713	12.800	3.867	12.916	3.868
Payment due to income	0.175	0.107	0.082	0.044	0.079	0.041
Macroeconomic factors						
GDP growth rate	2.278	0.652	2.054	0.670	2.317	0.638
Federal funds	9.732	6.383	19.635	10.573	9.709	6.350
Bankruptcy fillings	-0.029	0.067	-0.011	0.055	-0.029	0.066

Results

	Fully paid vs. Default			Fully paid vs. Prepayment		
	Coefficient	Error	p-value	Coefficient	Error	p-value
Intercept	18.043***	0.285	<.0001	0.069	0.179	0.6999
Borrower characteristics						
Debt to income	-0.022***	0.002	<.0001	-0.005***	0.001	0.0001
Employment length	-0.001	0.005	0.8169	-0.027***	0.003	<.0001
Home ownership Mortgage	0.264***	0.061	<.0001	0.123***	0.034	0.0002
Home ownership Rent	0.200***	0.060	0.0008	0.013	0.034	0.7015
Open credit lines	-0.027***	0.005	<.0001	-0.030***	0.003	<.0001
Total credit lines	0.035***	0.002	<.0001	0.027***	0.001	<.0001
Revolving credit utilisation	-0.718***	0.083	<.0001	-0.515***	0.046	<.0001
Public record	0.195	0.156	0.2108	0.042	0.091	0.6457
Last FICO	-0.025***	0.000	<.0001	0.006***	0.000	<.0001
Tax lines	0.098	0.177	0.5799	0.064	0.107	0.5511
Charge-off within 12 months	0.315*	0.125	0.0116	0.449***	0.131	0.0006
Public record bankruptcies	0.477**	0.166	0.0041	0.384***	0.097	<.0001
Inquiries 6 months	0.128***	0.017	<.0001	0.006	0.010	0.5283

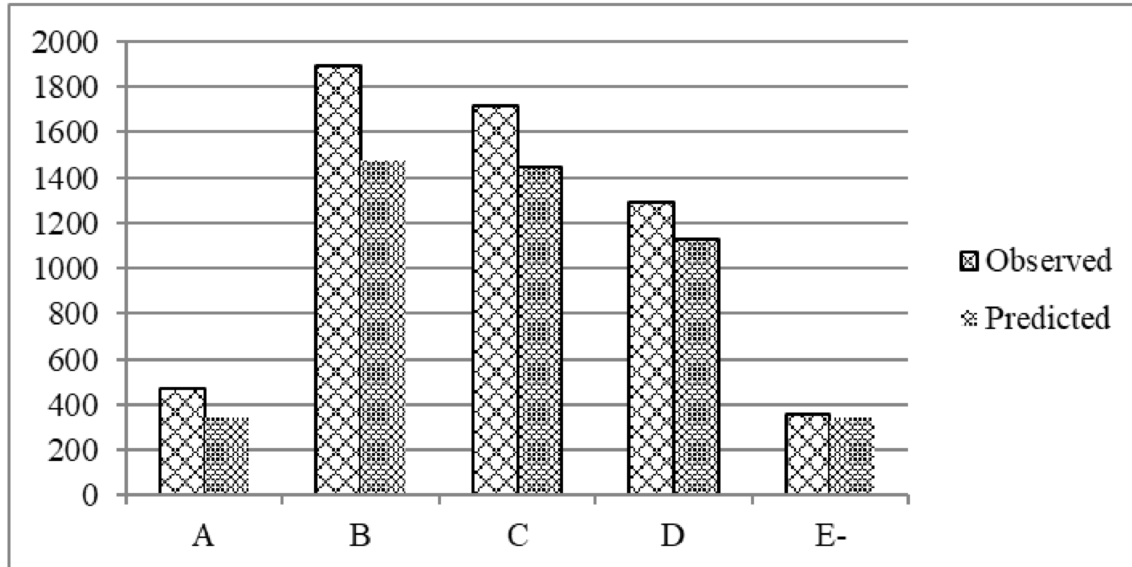
Results

Loan features						
Interest rate	0.229***	0.015	<.0001	0.298***	0.010	0.0001
Payment due to income	24.200***	0.318	<.0001	-1.015***	0.215	<.0001
Grade B	0.227**	0.082	0.0058	0.317***	0.041	<.0001
Grade C	0.339**	0.124	0.0061	0.410***	0.066	<.0001
Grade D	0.503**	0.167	0.0027	0.439***	0.092	<.0001
Grade E	0.705**	0.222	0.0015	0.547***	0.126	<.0001
Macroeconomic factors						
GDP growth rate	-1.318***	0.031	<.0001	-0.910***	0.019	<.0001
Bankruptcy fillings	-2.298***	0.258	<.0001	-2.277***	0.144	<.0001
Federal funds	-0.183***	0.003	<.0001	-0.163***	0.001	<.0001

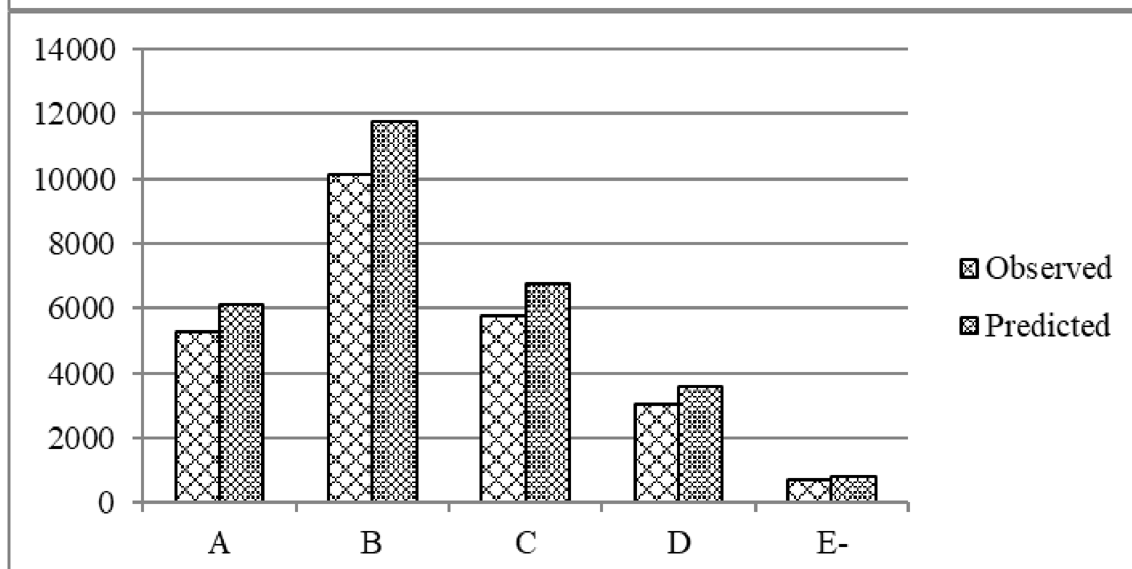
*, **, *** indicates that the coefficients are significant at 5%, 1% and 0.1% level respectively.

Results

Default



Prepayment



Prediction (Contingency table)

		Actual				Accuracy
		Default	Fully paid	Prepayment	Total	
Training Set						
Predicted	Default	7376	820	3025	11221	65.73%
	Fully paid	1062	17137	11069	29268	58.55%
	Prepayment	838	5162	41849	47849	87.46%
	Total	9276	23119	55943	88338	75.12%
Test Set						
Predicted	Default	3756	427	1553	5736	65.48%
	Fully paid	514	8729	5850	15093	57.83%
	Prepayment	410	2636	21558	24604	87.62%

Prediction (Accuracy)

	A	B	C	D	E-
Default	53.07%	63.24%	66.74%	71.32%	67.71%
Fully Paid	54.86%	57.74%	58.19%	61.47%	64.78%
Prepayment	86.06%	88.12%	88.05%	87.98%	86.50%
Total	72.59%	75.19%	75.32%	76.63%	75.98%

Conclusions

- We use multinomial logistic regression to model the three levels of outcomes of a loan: fully-paid, prepayment and default. Given the observable information of borrower characteristics and loan features, combined with the influence of macroeconomy, both default and prepayment can be accurately predicted, although the predictive performance for default is slightly poorer than that for prepayment.
- We found high interest rates of the loan does not only indicate large probability of default but also increase the probability of prepayment as borrowers do not wish to bear high interests. The borrower characteristics such as the debt-to-income ratio and the FICO score have significant impact on both outcomes, where a large FICO score imply that the borrower has a large chance to early repay the loan. Macroeconomic factors the GDP growth, the Federal fund rates and the personal bankruptcy rate can influence the occurrence of two events.

Conclusions

- Considering that the volume of payday loans has grown rapidly in recent years, with expensive annualised percentage rates (Bhutta, 2014), borrowers seem to be using P2P lending as an alternative to payday loans. Without penalties, P2P lending is typically used as a short-term but low cost loan, even though it is designed for 36-month or longer terms. Prepayment is therefore much more likely to happen and there may be an arbitrage opportunity of abusing P2P loans. It is suggested that lenders should pay attention to this and the platform may consider charging a penalty for prepayment, in order to compensate potential losses to their underlying loan portfolios. P2P companies such as Lending Club now repackage the loan portfolios and resell them to other financial institutions. Its appropriate pricing is important for investors and regulators who want to avoid any more disasters like the sub-prime crisis.

Thank for listening!

Comments appreciated!