Determinants of Loan Performance in P2P Lending

Nilas Möllenkamp
University of Twente
n.moellenkamp@me.com

ABSTRACT
This research paper investigates the influential factors of loan performance in the online P2P lending industry. The study analysed 143,654 P2P loans that were funded on the P2P lending platform Lending Club between 2012 and 2013 and found evidence that the assigned credit grade of loans is the most influential factor on loan success and default. Furthermore, loan amount and annual income are significant predictors. The variables debt-to-income ratio, inquires in the last 6 months, open credit lines and revolving credit balance were only found significant for some credit grade classes.

Keywords
P2P lending, P2P loan performance, loan success, loan risk

INTRODUCTION
Peer-to-peer ('P2P') lending has been a topic of high interest in the past few years. The U.S. Treasury (2016) expects the market to be worth around $90 billion in 2020. As it is one of the most promising trends in the modern online banking industry, research provides some studies on various fields of P2P lending. However, papers on ex-post P2P loan performance are mostly vague and do not deliver coherent outcomes. This research finds the determinants and influential factors of loan success and failure and uses most recent data available.

P2P lending is based on the idea of crowd funding. Borrowers apply for funding their unsecured loans for specific purposes and investors can lend a portion of these loans. While borrowers pay an interest rate to the lender and a transaction fee to the P2P lending platform, the microloan interest rate depends on the default risk that investors are facing. It is generally lower than with banks. P2P loans are usually between $1.000 and $25.000 and therefore are classified as microloans – a sector that is normally not targeted by traditional banks. The first P2P platform, Zopa.com, emerged in 2006. Therefore, the trend is relatively new and still barely regulated by governments.

Default risk for investors is omnipresent. The most relevant issue is information asymmetry. Credit institutes and banks precisely monitor loan applicants in order to secure loan repayment. In P2P lending, monitoring efforts are lower which increases moral hazard and therefore information asymmetry and ex-post loan default risk.

Usually, P2P platforms provide investors with a lot of information concerning the specific borrower so that the investor can decide whether a loan is promising and likely to deliver a full return. A credit grade that is assigned by the platform for every specific borrower is expected to be the most relevant one. It shows the creditworthiness of an individual borrower. At Lending Club, the platform of study, the credit grades range from A (lowest risk) to G (highest risk). Other data just like the debt-to-income ratio and the annual income are also given. This study

investigates which of those factors are influential for the default and success probability of a loan and therefore finds the determinants of loan performance. Hence, this research follows the research question:

What are the determinants of loan performance in P2P lending?

Table 1: Frequency Distribution by Credit Grade

Grade	Frequency (%)	Amount (%)	Defaulted Loans (%)	Amount defaulted loans (%)	Fully paid loans (%)	Amount fully paid loans (%)	Ratio defaulted to all loans (%)
A (lowest risk)	27,767 (19.3)	$377,978,925 (21.27)	1,554 (8.52)	$19,910,225 (9.12)	26,213 (20.90)	$358,068,700 (22.97)	5.60
B	57,050 (39.7)	$703,508,625 (39.58)	6,046 (33.13)	$72,764,000 (33.34)	51,004 (40.67)	$630,744,625 (40.46)	10.60
C	34,531 (24.0)	$416,220,100 (23.42)	5,510 (30.19)	$64,254,575 (29.44)	29,021 (23.14)	$351,965,525 (22.57)	15.96
D	19,552 (13.6)	$222,941,150 (12.54)	4,029 (22.08)	$47,303,400 (21.67)	15,523 (12.39)	$175,637,750 (11.27)	20.61
E	4,011 (2.8)	$48,212,575 (2.71)	922 (5.05)	$11,768,725 (5.39)	3,089 (2.46)	$36,443,850 (2.33)	22.99
F	706 (0.5)	$7,589,300 (0.43)	181 (0.99)	$2,084,800 (0.96)	525 (0.42)	$5,504,500 (0.35)	25.64
G (highest risk)	37 (0.0)	$924,625 (0.05)	7 (0.04)	$174,525 (0.08)	30 (0.02)	$750,100 (0.05)	18.92
Total	143,654 (100.0)	$1,777,375,300 (100.0)	18,249 (100.0)	$218,260,250 (100.0)	125,405 (100.0)	$1,559,115,050 (100.0)	12.70

Previous literature on the determinants of P2P loan performance is rather rare and lead, in the case of Lending Club, to vague results. Only four papers investigated further on this topic. However, there are many issues. Data is rather old, uses immature loans and only guesses loan defaults, does not reflect current economic situations, may be biased due the financial crisis or the introducing phases of P2P loans to public, or used small samples. Moreover, previous studies on P2P lending did not explain the loan performance on the basis of the different credit grades. This paper focuses on the seven different credit grades and gives insight into the specific determinants that are influential for loan performance in each grade.

The study finds a positive relationship between the credit grade and loan performance. With a higher credit grade, the risk of loan default decreases. Furthermore, there is evidence for more determinants of loan success. Loan amount and annual income are significant influential factors in all credit grades. Debt-to-income ratio as well as inquires in the last 6 months are significant influential factors in most grades. Open credit lines and total credit lines are found out to be only predictive in some credit grades while revolving credit balance has no influence in subsamples of credit grades.

LITERATURE REVIEW AND HYPOTHESIS

The most relevant risk in lending for lenders is always default risk. P2P lending is accompanied with some unique risks compared to traditional lending that make it riskier for lenders to invest. For example, platform default, fraudulent activities and cybercrime can decrease profitability due to loan default (Kirby & Worner, 2014). Some papers suggest more regulation as well as bank involvement for P2P platforms in order to decrease potential risks (e.g. Galloway, 2009). P2P lending also differs from traditional lending in terms of monitoring of borrowers. While banks and other financial intermediaries can observe bank account activities somehow (Gorton & Winton, 2003), P2P platforms do not. Information asymmetry in moral hazard increases and must be handled by investors. This construct is expected to influence the ex-post risk of loan success and default.

Although there is some general literature on P2P lending, research on the influential factors of P2P loan performance is quite rare. For Lending Club, only four papers (Emekter, Jirasakuldech & Lu, 2015; Li, Yao, Wen & Yang, 2016; Carmichael, 2014; Serrano- Cinca, Gutiérrez-Nieto &

Table 2: Impact of Credit Grades

	A	B	C	D	E	F	G
				β ($\exp(\beta)$)			
Intercept	2.775*** (16.044)	2.288*** (9.854)	1.932*** (6.906)	1.644*** (5.174)	1.482*** (4.402)	1.563*** (4.774)	1.408 (4.088)
Grade A		.487*** (1.628)	.843*** (2.323)	1.132*** (3.101)	1.293*** (3.645)	1.212*** (3.361)	1.367 (3.925)
Grade B	-.487*** (.614)		.355*** (1.427)	.644*** (1.904)	.806*** (2.239)	.725*** (2.064)	.880 (2.411)
Grade C	-.843*** (.430)	-.355*** (.701)		.289*** (1.335)	.450*** (1.569)	.369** (1.447)	.524 (1.690)
Grade D	-1.132*** (.322)	-.644*** (.525)	-.289*** (.749)		.162** (1.175)	.080 (1.084)	.236 (1.266)
Grade E	-1.293*** (.274)	-.806*** (.447)	-.450*** (.637)	-.162** (.851)		-.081 (.922)	.074 (1.077)
Grade F	-1.212*** (.298)	-.725*** (.484)	-.369** (.691)	-.080 (.923)	.081 (1.085)		.155 (1.168)
Grade G	-1.367 (.255)	-.880 (.415)	-.524 (.592)	-.236 (.790)	-.074 (.929)	-.155 (.856)	
Loan Amount	-.000026*** (.999974)	-.000026*** (.999974)	-.000026*** (.999974)	-.000026*** (.999974)	-.000026*** (.999974)	-.000026*** (.999974)	-.000026*** (.999974)
Annual Income	.000008*** (1.000008)	.000008*** (1.000008)	.000008*** (1.000008)	.000008*** (1.000008)	.000008*** (1.000008)	.000008*** (1.000008)	.000008*** (1.000008)
Debt-to-income	-.014*** (.986)	-.014*** (.986)	-.014*** (.986)	-.014*** (.986)	-.014*** (.986)	-.014*** (.986)	-.014*** (.986)
Inquires in the Last 6 Months	-.080*** (.923)	-.080*** (.923)	-.080*** (.923)	-.080*** (.923)	-.080*** (.923)	-.080*** (.923)	-.080*** (.923)
Open Credit Lines	-.018*** (.982)	-.018*** (.982)	-.018*** (.982)	-.018*** (.982)	-.018*** (.982)	-.018*** (.982)	-.018*** (.982)
Revolving Credit Balance	.000008** (1.000008)	.000008** (1.000008)	.000008** (1.000008)	.000008** (1.000008)	.000008** (1.000008)	.000008** (1.000008)	.000008** (1.000008)
Cox-Snell R^2	.028	.028	.028	.028	.028	.028	.028

Table 3: Binary Logistic Regression Results

	A	B	C	D	E	F	G
				β (exp (β))			
Intercept	2.481*** (11.947)	2.128*** (8.400)	2.092*** (8.100)	1.767*** (5.852)	1.166*** (3.209)	2.093** (8.109)	
Grade A	2.481*** (11.947)						
Grade B		2.128*** (8.400)					
Grade C			2.092*** (8.100)				
Grade D				1.767*** (5.852)			
Grade E					1.166*** (3.209)		
Grade F						2.093** (8.109)	
Grade G							
Loan Amount	-.000034*** (.999966)	-.000019*** (.999981)	-.000020*** (.999980)	-.000036*** (.999964)	-.000030*** (.999970)	-.000018 (.999982)	
Annual Income	.000006** (1.000006)	.000008*** (1.000008)	.000008*** (1.000008)	.000008*** (1.000008)	.000007** (1.000007)	.000008 (1.000008)	
Debt-to-income	-.026** (.975)	-.016*** (.984)	-.012*** (.988)	-.016*** (.984)	-.001 (.999)	-.011 (.989)	
Inquires in the Last 6 Months	-.047 (.954)	-.067** (.935)	-.107*** (.899)	-.054** (.948)	-.098** (.907)	-.139 (.870)	
Open Credit Lines	-.028 (.972)	-.003 (.997)	-.023*** (.977)	-.026*** (.975)	-.019 (.981)	-.071 (.931)	
Revolving Credit Balance	.000005 (1.000005)	.000008 (1.000008)	.000014 (1.000014)	.000008 (1.000008)	-.000006 (.999994)	.0000041 (1.000041)	
Cox-Snell R^2	.011	.009	.011	.020	.017	.048	
N	27,767	57,050	34,531	19,552	4,011	706	37

López-Palacios, 2016) that investigate the determinants of loan performance emerged in literature. All found different determinants of loan success and default. Overall, 24 distinctive factors were stated as significant, but different ones in each of the papers. Some, such as credit grade, debt-to-income ratio and annual income were found to be significant in at least three of these papers. This study focuses (1) on the impact and influence of the given credit grade on loan performance and (2) on other determinants that are expected to influence the loan performance. Therefore, two hypotheses are tested:

H1: The higher the credit grading, the less likely is P2P loan default.

H2: The borrower and loan characteristics loan amount, annual income, debt-to-income ratio, inquires in the last 6 months, the number of open credit lines, revolving credit balance and the number of total credit lines are significant predictors of loan success in all risk classes.

METHODOLOGY AND VARIABLES

In accordance to past research papers, this study uses logistic regressions and the estimated maximum likelihood method for identifying the effect of independent variables on the loan performance. Since only two outcomes (loan success and loan default) are possible, the binary logistic regression is chosen.

DATA

In order to secure data is most recent and already has a fixed loan status that is either default or success, only matured loans from the years 2012 and 2013 are used. For this study, 143,654 loans with an overall loan amount of $1.7 billion from the P2P lending platform Lending Club were chosen. Data is publicly available on www.lendingclub.com. Detailed frequency distributions are displayed in Table 1. Most loans belong to credit grade B, while less than 1% of all loans are in credit grade F and only 37 loans were counted in grade G. Data should be handled with caution. All values of the descriptive statistics are in line with previous studies by Emekter et al. (2015), Serrano-Cinca et al. (2015) and Li et al. (2016).

Furthermore, correlation tests between all variables were made. While the correlation between all other variables is minor, open credit lines and total credit lines correlate by 0.667. In order to avoid multicollinearity, the number of total credit lines is not used for regressions.

EMPIRICAL RESULTS

The empirical results of the binary logistic regressions are summarised in tables 2 and 2. Table 2 shows the impact of credit grades on the loan performance. A change from grade B, for example, to grade C decreases the loan success probability by the logarithm of exp (β), so 35.52%, assumed all other variables remain the same. The results are generally in line with expectations: the higher the credit grade, the lower is the loan default risk. However, this is not true for grades F and G in direct comparison to grade E. Further investigations with a higher amount of data should be underdone in order to clarify this unusual finding. However, at least for 95.5% of all loans, H1 is supported by the results.

The coefficients of all variables are significant at the 1%, respectively the 5% level (revolving credit balance). With these insights, H2 is supported for the full sample of all loans.. Loan amount, annual income, debt-to-income ratio as well as inquires in the last 6 months, open credit lines and revolving credit balance are indeed determinants of loan performance. Table 3 also shows how the probability for

loan success changes when adding one more unit of the different variables. For instance, with one more US-Dollar added to the loan amount, the actual probability of loan success decreases by the inverse of the exp β , so by 0.0026%, given all other variables remain constant. Besides the credit grade, the most influential factor is inquires in the last 6 months where one more inquiry decreases the loan success probability by 7.7%.

Table 3 gives a deeper understanding about the influence of specific variables in different credit grades. Due to small amounts of data for credit grade G, no regressions could be run. In grade F, none of the coefficients is significant. Data should be handled with caution. The Cox-Snell R^2 model fit also shows a distinct higher value in credit grade F (.048), compared to other credit grades.

For all other credit grades, there are some surprises. Although revolving credit balance is significant for the whole loan sample, the coefficients for every subsample per credit grade are not significant. Moreover, the number of open credit lines is only significant in credit grades C and D. Inquires in the last six months as well as debt-to-income ratio are significant in all credit grades besides grade A, respectively grade E. The only influential factors that are significant in all credit grades except for the special cases of grades F and G are loan amount and annual income. Therefore, it can be stated that hypothesis H2 is not supported. Only loan amount and annual income are indeed influential factors of loan performance for the full loan sample and credit grade subsamples. Debt-to-income ratio, inquires in the last 6 months and open credit lines cannot be used as determinants beyond doubt. The number of total credit lines is expected to behave very similar to open credit lines.

CONCLUSION

In previous research, four studies on Lending Club found different variables. Overall, 24 determinants were stated as significant and all variables differ between the studies. This research is the first that investigates the determinants of P2P loan performance on the basis of different credit grades. Results indicate a distinctive view on loan performance factors.

The credit grade is the most influential predictor of loan performance. However, only loan amount and annual income can further be stated as significant predictors of loan performance. All other variables lose significance in forecasting power when it comes to subsampling by credit grades. To get reliable results of predicting loan performance, a pre-selection of variables on the basis of credit grades should be underdone. More precise results can be expected. This could solve the discordant results that were delivered by previous research and give a deeper insight into the topic of ex-post risk in P2P Lending. Furthermore, larger datasets for credit grades F and G are to be used in order to get consistent results that are comparable to grades A to E.

ROLE OF THE STUDENT

For my Bachelor thesis, I was in close connection to Dr. X. Huang of the University of Twente in order to find a appropriate topic that has not been studied in research in excess. The study, including all calculations and data gathering, as well as the processing of results and formulation of results and conclusion was done by me.

REFERENCES

Carmichael, D. (2014). Modeling default for peer-to-peer loans. *Available at SSRN: http://ssrn.com/abstract=2529240*. ISSN 1556-5068. doi: 10.2139/ssrn.2529240.

Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics*, *47*(1), 54-70.

Galloway, I. (2009). Peer-to-peer lending and community development finance. *Community development investment center working paper*, (39), 19-23.

Gorton, G., & Winton, A. (2003). Financial intermediation. *Handbook of the Economics of Finance*, *1*, 431-552.

Kirby, E., Worner, S. (2014). Crowd-funding: An infant industry growing fast. *IOSCO, Madrid*.

Li, Z., Yao, X., Wen, Q., & Yang, W. (2016). Prepayment and Default of Consumer Loans in Online Lending.

Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of default in P2P lending. *PloS one*, *10*(10).

U.S. Treasury, (2016). Opportunities and Challenges in Online Marketplace Lending. *Washington, DC, http://bit.ly/27bRLDC*

SRC 2016, November 30, 2016, The Netherlands.