



UDK: 336.012.23:004..738.5
DOI: 10.1515/jcbtp-2017-0017

Journal of Central Banking Theory and Practice, 2017, 2, pp. 149-167
Received: 18 September 2016; accepted: 21 January 2017

Andrija Đurović*

* *Societe Generale Montenegro AD, Montenegro*

Email:
djandrija@gmail.com.

Estimating Probability of Default on Peer to Peer Market – Survival Analysis Approach

Abstract: Arguably a cornerstone of credit risk modelling is the probability of default. This article aims to search for the evidence of relationship between loan characteristics and probability of default on peer-to-peer (P2P) market. In line with that, two loan characteristics are analysed: 1) loan term length and 2) loan purpose. The analysis is conducted using survival analysis approach within the vintage framework. Firstly, 12 months probability of default through the cycle is used to compare riskiness of analysed loan characteristics. Secondly, log-rank test is employed in order to compare complete survival period of cohorts. Findings of the paper suggest that there is clear evidence of relationship between analysed loan characteristics and probability of default. Longer term loans are more risky than the shorter term ones and the least risky loans are those used for credit card payoff.

Key words: Peer-to-peer market, Probability of default, Survival analysis, Vintage framework

JEL: G11, G23, C41

Introduction

P2P is a type of financing that involves the facilitation of loan originations outside the traditional consumer banking system by connecting borrowers directly with investors, through an Internet platform. This type of lending assumes individuals lending money to other individuals

without a banking intermediary. With P2P lending, applicants submit an application to the lending service. The lending service then decides if they will offer a loan and under what terms and conditions, which include interest rate, loan length and payment schedule. On other side, individual investors can manually pick which loan to invest in and how much money they will invest or they can use automated procedure based on investor defined criteria to build up a portfolio. Once the full amount for a loan is raised, the loan is fully funded and the borrower will receive their money less an origination (service) fee.

The history of P2P in finance can be traced back to the launch of two companies, the UK-based Zopa in 2005 and the US-based Prosper in 2006. Since then, P2P lending has developed rapidly, especially in the USA and the UK. Still, P2P lending represents a small fraction, less than 1%, of the stock of bank lending. In the UK – but not elsewhere – it is an important source of loans for smaller companies. Milne and Parboteeah (2016) argue that P2P lending is fundamentally complementary to, and not competitive with, conventional banking. Therefore, they expect banks to adapt to the emergence of P2P lending, either by cooperating closely with third-party P2P lending platforms or offering their own proprietary platforms.

Overall, many studies conclude that P2P lending is more risky than traditional banking. However, it is important to recognize that a constant conclusion would be misleading. P2P platforms have evolved and changed their appearance markedly over time, which implies that although final conclusion of increased riskiness through P2P markets remains valid over time, it is based on different arguments at different points in time (Käfer, 2016).

Loans channelled via P2P platforms involve higher interest rates than loans channelled via the traditional banking sector. When adjusted for risk, the interest rates are comparable. Moreover, analysis of the different segments of the bank credit market and P2P lending shows that, after having controlled for interest rate and risk differences, the bank lending volumes are negatively correlated with the P2P lending volumes. De Roure's et al. (2016) finding suggests that high-risk borrowers substitute bank loans for P2P loans since banks are unwilling or unable to supply this slice of the market.

Comparative study of online P2P lending practices in the USA and China was conducted by Chen and Han (2012). They found that two categories of credit information, “hard” and “soft” information, may have profound influences on lending outcomes in both countries, but lenders in China rely more on “soft” information. “Hard credit information” refers to the credit information that can be

accurately quantified, easily stored and efficiently transmitted. In the P2P lending context, hard credit information includes a credit profile of a borrower, such as a borrower's debt-to-income ratio, credit ratings, the number of credit inquiries made in the past, and the number of credit cards held by the borrower. In contrary to "hard credit information", such as credit scores or the financial condition of the borrower, "soft credit information" refers to information about borrowers that is fuzzy and hard to quantify. In P2P lending, soft credit information may be obtained from social networks of borrowers.

The same phenomena is studied by Duarte et al. (2012) who find that borrowers who appear more trustworthy have higher probabilities of obtaining a loan and pay lower interest rates than borrowers who appear less trustworthy. They also find that borrowers who appear more trustworthy indeed have better credit scores and lower default rates than those less trustworthy.

These results show that soft information sources possess at least as much significance about the borrower on the P2P lending markets as financial data. Those findings play an important role in lending outcome, but also mirror to investor's credit risk.

Using a sample of 11,752 loans from the Prosper P2P lending marketplace, Golubnicijis (2012) employs a 5-stage methodology in order to analyse and compare the attractiveness of the P2P lending market with traditional investment alternatives in terms of risk remuneration. Results present the evidence of high potential of P2P lending market as an investment alternative to the stock market, assuming maximum diversification opportunities and lender's efficiency in interest rate setting. Also, he arrives to the conclusion that, at its current level of development, the P2P loan market offers attractive investment risk remuneration particularly for lenders with longer investment horizon or with lower financial literacy.

Atz and Bholat (2016) provided the first map of the UK P2P lending market in 2013, showing asymmetry between area of borrowing and lending on P2P market. Secondly, they put stress on subjective factors, notably the motives animating lenders, borrowers and P2P operators creating this new market. For investors, main motivation is pecuniary, but also control over where the invested money will go as well as supporting alternative to the big banks. On the other side, interest rate is the major motivation for the borrowers participating in the market. Finally, they flesh out the implications of P2P lending for the conventional banking sector. Here they mark unsecured personal loans (reduced rates) and necessary shift of distributional channel as product and business strategies are most likely to be impacted.

P2P investors suffer a severe problem of information asymmetry because they are at a disadvantage facing the borrower. For this reason, P2P lending sites provide potential lenders with information about borrowers and their loan purpose. P2P lending is a risky activity for individual investors because the loans are granted by them instead of P2P companies, which transfer the credit risk. Credit risk is the dominant source of risk for traditional banks but also for investors in P2P market. It is commonly defined as the loss resulting from failure of obligors to honour their payments. Arguably a cornerstone of credit risk modelling is the probability of default. Therefore, understanding and modelling probability of default play a key role in financial markets.

Estimating default probabilities could be challenging mainly due to limitations on data availability. Fortunately, in last decades number of models for modelling probability of default have been developed which allow us to overcome these limitations.

In this study we focus on modelling the probability of default using survival analysis techniques. In order to write the probability of default in terms of the conditional distribution function of the time to default, non-parametric approach is used. The main objective of this research is to use the survival model to generate probabilities of default at various points in time and to compare probability of default for different loan characteristics.

The idea of using survival analysis techniques for constructing credit risk models is not new. It started with the paper by Narain (1992) and it was later developed by Carling et al. (1998), Roszbach (2003), Glennon and Nigro (2005), Allen and Rose (2006), Baba and Goko (2006), Malik and Thomas (2006). A common feature of all these papers is that they use parametric or semi-parametric regression techniques for modelling the time to default (duration models), including exponential models, Weibull models and Cox's proportional hazards models, which are very common in this literature. The model established for the time to default is then used for modelling probability of default or constructing the scoring discriminant function.

Serrano-Cinca et al. (2015) study P2P lending and factors explaining loan default using the survival analysis and the logistic regression model. Their study results show that there is a clear relationship between the grade assigned by Lending Club and the probability of default. The interest rate assigned depends on the grade assigned and the higher the interest rate, the higher the default probability is. Loan purpose is also a factor explaining default: wedding is the less risky loan purpose and small business is the riskiest. Borrower characteristics, such as an-

nual income, current housing situation, credit history, and borrower indebtedness are relevant variables. No statistically significant differences are found in loan amount or length of employment.

To the best of our knowledge, this is the first study explaining defaults in Lending Club platform, using a database large enough to monitor the credit risk development over years.

Until recently, this was not possible due to data availability on the loan status. Our results show that, changes, in methodology used to connect investors and borrowers, influence level of credit risk and provide more possibilities to investors to distinguish between less and more risky loans. Both analyzed loan characteristics, term length and loan purpose, have a significant relationship with default within the analysed data.

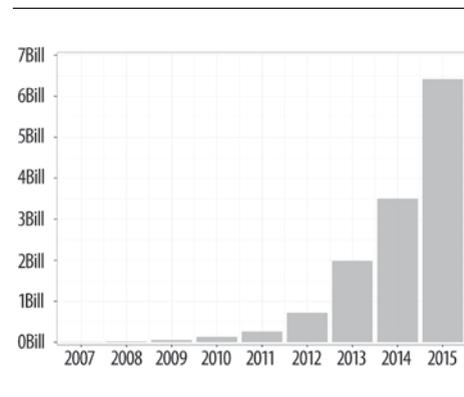
The remainder of the paper is organized as follows. Section 2 presents Lending Club portfolio description and its development over years. Section 3 gives empirical results, followed by conclusions.

Portfolio description

In this article, we analyse complete loan data¹ along with borrower's payments history², issued in the period from 2007 to 2015, from the world's largest online credit marketplace, Lending Club. Since Lending Club allows for delinquency status, as defaulted loans are marked all loans with current status *Default* or *Charged off*.

Total number of approved loans in the analysed period is 887,440 with the payment history of 16.6 million records. The number of loans with defaulted status is 69,165.

Figure 1: Amount of approved loans per year



¹ data source: <https://www.lendingclub.com/info/download-data.action>

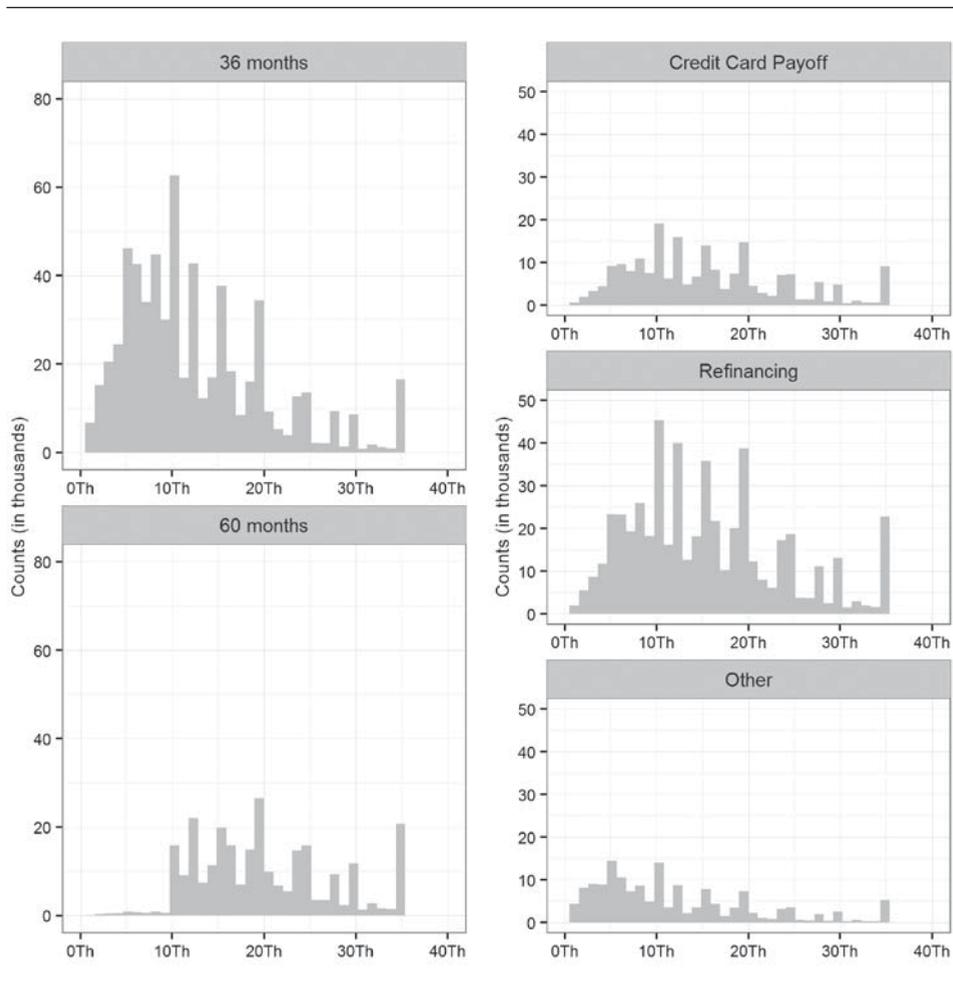
² data source: <http://additionalstatistics.lendingclub.com>

Figure 1 presents the amount of issued loan (in billions of dollars) per years.

The increasing trend of loans issued from year to year is easily noticeable, ranging from minimum 4.97 million in 2007 to 6.41 billion in 2015. Approved amount is in range of 500 to 35,000 of dollars and it is stable over the years.

Term length and loan purpose present an important factor in the process of loan approval, thus further analysis will be based on different groups regarding these characteristics. The next figure, Figure 2, presents the histogram of approved amount per different term length and loan purpose groups.

Figure 2: Histogram of approved loans per term length and loan purpose



Regarding the loan structure, two term lengths are available: 36 and 60 months. Longer term loans started being approved from 2010, with the share of 40.79% in the amount and 30.00% in the number of issued loans. Descriptive statistics registered mean value of issued loans of 12,481 and 20,059 of U.S. dollars for 36 and 60 months terms, respectively. Based on their share in total portfolio, loan purposes can be divided in three categories: 1) credit card pay off, 2) refinancing and 3) other loans. The group “other” contains loans classified as loans for buying cars, small business, and home improvement. The main share in the issued portfolio was of loans used for refinancing, with the share of 61.77% in the amount and 59.07% in the number of issued loans. Credit card payoff loans accounted for 24.15% and 23.23% in the issued amount and the number, respectively, while the share of loans for other purposes was 14.07% and 17.69%. An average issued amount for loan purpose categories range from 15,400 (Refinancing and Credit Card Payoff) to 11,740 (Other) U.S. dollars.

Table 1 and Table 2 (see Appendix) present yearly portfolio development of issued loans based on analyzed characteristics. In the past four years, longer term loans recorded an increasing trend in the share of all issued amounts starting from 29.32% in 2012 to 43.49% in 2015. On the other hand, the share of loans regarding loan purpose was stable over years, keeping the same structure as for the complete analyzed period.

Methodology

Following the vintage framework, cohorts (pools) are formed based on the origination date. Further segmentation is done by term length and loan purpose, separately. In total we have 103 cohorts³ for each segment – term length and loan purpose. For all the cohorts, probability of default is calculated using the survival analysis approach.

Usually, there are four different ways of presenting probability function in the survival analysis (Buis, 2006). First one is the probability function which gives the probability that an observed subject will survive past time t :

$$S(t) = \Pr(T > t) \quad (1)$$

Its complementary function, cumulative probability function, gives probability that the observed subject will not survive past time t :

³ 103 cohorts based on month of approval, starting from June 2007 until December 2015

$$F(t) = 1 - S(t) \quad (2)$$

The third way is the probability density function. This function gives the probability for every point of time that the event will occur at that point of time:

$$f(t) = \Pr(T = t) \quad (3)$$

Probability density function does not take into account that event for observed subject occur and the subject is no longer at risk. Hazard function takes care of this omission and presents instantaneous rate of event occurring at time t given the probability that the subject survived until that time:

$$h(t) = \frac{f(t)}{S(t)} \quad (4)$$

In this study, cumulative density function is used to present the probability of default for different points of time. Before estimating the probability of default, time of default is defined by using the known variables. For all loans with given time of default in payment history data, exact time is used while for loans with missing default time (887 loans) and no recovery amount, time of default is defined using the following formula:

Time of default = last payment date + 31 days (payment due date) + 121 days
(default point)⁴

For loans with missing default time and recovery amount greater than zero (36 loans), time of default is defined as follows:

Time of default = [(total payments-recoveries)/instalment]*31 + 31 days
(payment due date) + 121 days (default point)

An important feature of the survival analysis is that it can deal with censoring and truncation. Censored observation is the subject for whom the event of interest does not occur during the follow-up period. We should distinguish between terms censoring and truncation even though the two events are handled the same way analytically (Stevenson, 2009). A truncation period means that the outcome of interest cannot possibly occur. There are two types of censoring and truncation: 1) right and 2) left. Subject is right/left censored if it known that the event of interest occurs sometime after/before the recorded follow-up period. Subject is left truncated if it enters the population at risk at some stage after the start of

⁴ Lending Club P2P platform classifies a loan as defaulted starting from 121 days of delay

the follow-up period while it is right truncated if it leaves population at risk at some stage after the study start. A common type of censoring and truncation in risk environment is right due to prepayments, refinancing, call options, etc. also called exits (Xu, 2016). Analysed loans are characterized by high right censoring/truncation rates. Censoring/truncation time is presented by last payment dates.

A number of methods are available to model survival time. Methods include non-parametric, parametric and semi-parametric approaches. Estimating the probabilities without making any assumption on its shape is called non-parametric analysis. There are three non-parametric methods for the estimation of survival probability: Kaplan-Meier, Nelson-Aalen, and life table method. In parametric survival analysis all parts (baseline function and influence of covariates) of the model are specified. The most used underlying distributions are gamma, exponential, Weibull, log-normal. Parametric models are usually used in circumstances where prediction is the object of the exercise. Semi-parametric model presents intermediate technique whereby only assumption is made only about the way how covariate influences the risk of survival. Semi-parametric model is also called Cox-regression. In order to estimate probability of default we employed the Kaplan-Meier method (product-limit estimator). It is based on individual survival times and assumes that censoring/truncation is independent of survival times. At any point of time we have two groups: numbers of events occurred $d(t_i)$ and the population at risk $r(t_i)$. Then, survival function is:

$$\widehat{S}_{KM} = \prod_{t_i < t} \frac{r(t_i) - d(t_i)}{r(t_i)} = \prod_{i=1}^j \widehat{Pr}(T > t_i | T \geq t_i)$$

Following the aim of the paper, comparison of probability of default for different cohorts is performed by using log-rank test. The test compares estimates of hazard function of the groups at each observed event time. It is constructed by computing the observed and expected number of events in one of the groups at each observed event time and then adding these to obtain an overall summary across all-time points where there is an event:

$$\chi^2 = \sum_{i=1}^j \frac{(O_i - E_i)^2}{E_i}$$

This test statistic is approximately distributed as a chi-square random variable with one degree of freedom.

The cumulative density function is calculated using obtained survival probabilities. Then, taking the simple mean of all cumulative densities, for each cohort and for n periods after origination date, we obtain the so-called probability of default through the cycle (PD TTC). PD TTC is used to compare average probability of default between analysed segments for certain point of time, while log-rank test compares complete survival period.

To compare difference between segment's PD TTC t-test and ANOVA are employed. Two-sample t-test formula is as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1}{n_1} + \frac{S_2}{n_2}}}$$

where \bar{X}_1 and \bar{X}_2 represent mean return, S_1 and S_2 standard deviation of samples, and n_1 and n_2 are the number of observations in samples.

Total variability of an observed variable can be divided into factorial and residual sum of squares. ANOVA assumes comparison of factorial and residual sum of squares offering a general evaluation of group mean differences. Factorial variance is given by:

$$V_f = \frac{\sum_{i=1}^r n_i (\bar{X}_i - \bar{\bar{X}})^2}{r - 1}$$

where V_f stands for factorial variance, n_i number of observation in i -th group, \bar{X}_i is mean return in i -th group, $\bar{\bar{X}}$ overall mean return, and r number of groups, while

$$V_r = \frac{\sum_{i=1}^r \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^r n - r}$$

where V_r residual variance, X_{ij} present j -th value of returns in i -th group, and \bar{X}_i mean return in i -th group and n is number observations in each groups. Ratio of those two variances follow F-distribution with $r - 1$ i $n^*r - r$ degrees of freedom:

$$F = \frac{V_f}{V_r}$$

If ANOVA shows statistically significant results, then we use Tukey's honest significant difference (Tukey's HSD) test for comparing group means. According

to the Mason et al. (2003), two averages \bar{y}_i and \bar{y}_j based on n_i and n_j observations respectively, are significantly different if

$$|\bar{y}_i - \bar{y}_j| > TSD,$$

where

$$TSD = q(\alpha; k, v) \left(MSE \frac{n_i^{-1} + n_j^{-1}}{2} \right)^{1/2}$$

in which $q(\alpha; k, v)$ is the studentized range statistic, k is the number of averages being compared, MSE is the mean squared error from an ANOVA fit for the data based on v degrees of freedom, and α is the experiment-wise error rate.

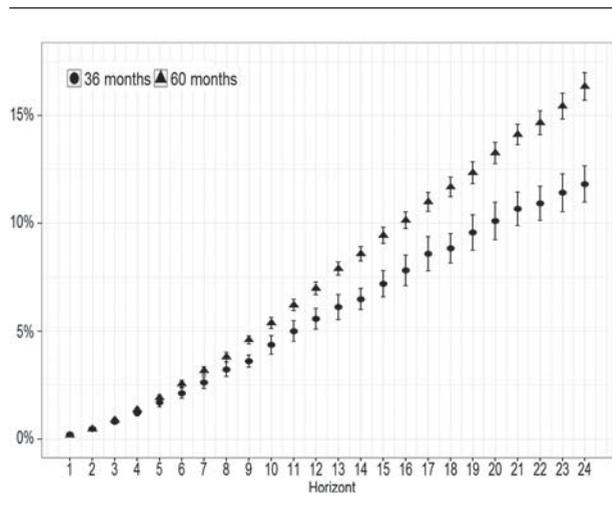
Empirical results

Based on PD TTC for 24 months after the origination date, we find statistically significant difference between loans of different term length, as well as for different loan purposes.

The figure 3 presents 95% interval of PD TTC for both term lengths for 24 months after the origination date.

With the widening of the analysing period, the difference between probabilities of default of two term lengths becomes more evident. Starting from the ninth month following the origination date the difference becomes statistically significant, leading to the conclusion that 60-month term loans are more risky than 30-month loans if measured by PD TTC. Twelve-month probability of default for 30-month loans is 5.58%, with 95%

Figure 3: 95% mean confidence interval of PD TTC per term length



confidence interval of the mean of 5.09% - 6.06%. The same probability of default for 60-month term loans is 6.98% with 95% confidence interval of the mean of 6.69% - 7.27%. Mean difference of 12-month PD TTC for two term lengths, tested by two sample t-tests, is statistically significant with p-value of 2.814e-06.

Figure 4: p-value of log-log rank test for each cohort for different term length

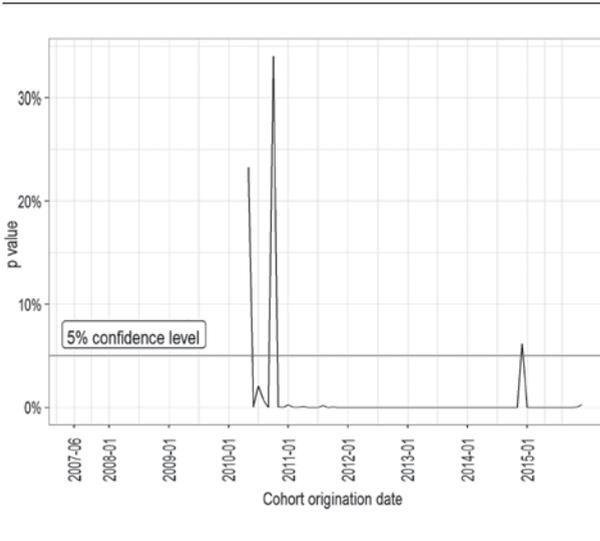
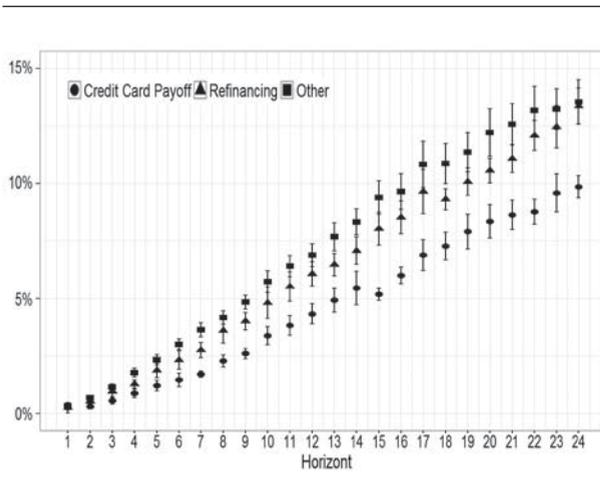


Figure 5: 95% mean confidence interval of PD TTC per loan purpose



The same probability of default for 60-month term loans is 6.98% with 95% confidence interval of the mean of 6.69% - 7.27%. Mean difference of 12-month PD TTC for two term lengths, tested by two sample t-tests, is statistically significant with p-value of 2.814e-06. In order to compare survival probabilities for each cohort log-rank test is employed. Figure 4 presents p-value of log-rank test for each cohort.

Longer term loans started to issue from 2010, therefore data for previous cohorts are missing. Only few initial cohorts (2010-2011) show statistically nonsignificant difference in survival probabilities. Starting from 2011, we find statistically significant difference of total survival probability between these two term lengths for almost each cohort, except for loans approved in January 2015.

With regard to loan purpose, same conclusions can be drawn. Twelve months probability of default for refinancing is 6.07% (95% confidence interval 5.54%-6.59%), for credit card payoff it is 4.33% (95% confidence interval 3.89%-4.77%), while for other purposes

it is 6.88% (95% confidence interval 6.38%-7.38%). Figure 5 presents 95% confidence interval for PD TTC 24 months after the origination date.

ANOVA test confirmed visual findings that there is at least one pair of mean differences statistically significant from zero with p-value of 3.21e-10, while Tukey’s HSD test revealed there is statistically significant difference between credit card payoff PD TTC and two other PD TTC at 5% level of significance.

Table 3 presents results of Tukey’s HSD test.

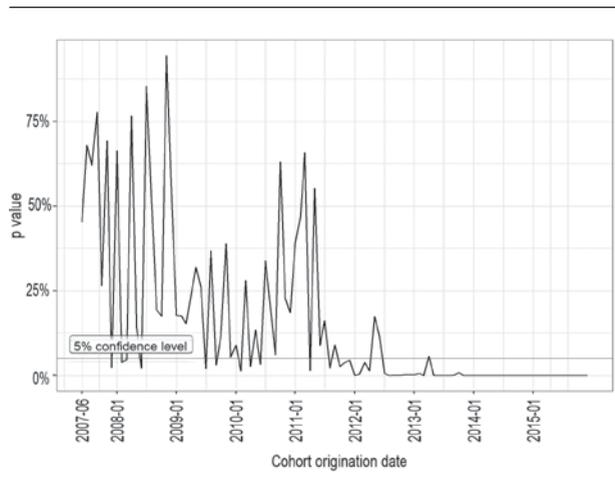
Table 3: Results of Tukey’s HSD test for all pairs of loan purposes

Pair	Mean difference	Lower limit	Upper limit	p-adj
Refinancing-Credit Card Payoff	1.73%	0.86%	2.61%	0.00143%***
Other-Credit Card Payoff	2.55%	1.67%	3.42%	0.00000%***
Other-Refinancing	0.81%	-0.01%	1.63%	5.22891%

Signif. codes: 0'***' 0.1 ' ' 1

Even if the share of credit card payoff in total issued amount is significantly lower than the financing share, loans approved for this purpose are less risky. Formally, the difference between PD TTC for different loan purpose is confirmed with log-rank test on cohort level. Figure 6 presents results of log-rank test for each cohort and per loan purpose.

Figure 6: p-value of log-log rank test for each cohort for different loan purpose



Unlike term length, log-rank tests for loan purpose show statistically significant difference in recent years, starting from middle of 2012. These variations of log-rank test results over the years are closely related to P2P market growth and with Lending Club platform changes. Namely, the increasing number of P2P borrowers and investors caused changes in Lending Club approval process in order to decrease investor’s risk tightening credit policy.

Conclusion

The aim of this paper was to analyse the probability of default, as a cornerstone of credit risk, at the P2P market and to compare loan characteristics from the point of view of investor's risk using the survival analysis in the vintage model framework. P2P platforms change their loans approval methodology and collection, constantly, which have an impact of investor's risk. Increasing trend of the number of participants in the market demands a closer and detailed attention to investment decisions. Based on the number of rejected application, it is obvious that interest in P2P lending exists and a further growth can be expected. However, growing velocity in the following years deserves a closer look. On one side, a vast majority of people are not willing to lend money in an unsecured and uninsured manner in long term without having pronounced opportunities to liquidate their investments. On the other side, the appearance of institutional investors as substitutes for private lenders can facilitate growth perspectives even above the previous rates. From the borrowers' perspective, it is less important who provides them with loans. Nevertheless, it is open to debate whether it will ever become as important as the traditional financial system.

Having all this in mind, monitoring and controlling investor's risk is of high importance. Analysis of determinants of default can significantly help investors to identify good and bad loans and diversify their portfolios as much as possible to avoid adverse selection. Here, the probability of default for two loan characteristics: 1) term length and 2) loan purpose, were analysed using all available data from Lending Club P2P platform (loan data and payment history). Firstly, the study shows that there is a clear evidence of relationship between probability of default and analysed loan characteristics. Longer term loans are more risky than shorter term ones, and the least risky loans, regarding the loan purpose, are those for credit card payoff. Secondly, the study reveals evidence that this relationship becomes more evident with P2P market development. For recently issued loans (from 2012 till 2015) difference in the probability of default between analysed segments becomes more evident and that is formally confirmed with log-rank test for all cohorts.

Our results are similar to those found in the P2P lending market where determinants of default are documented: Serrano-Cinca's et al. (2015), Golubnicij's (2012) first stage of the research, Duarte et al. (2012) although we use longer analysed period and the vintage model framework that provide possibility to make conclusion about timely change of credit risk at the P2P lending market.

To the best to our knowledge, this is the first study analysing defaults at the P2P lending market using the survival approach within vintage framework. Findings of this paper encourage further research in the field of credit risk on P2P lending and the examination of other loan's and available borrower's characteristics as possible determinants of default.

References

1. Allen, L. N. and Rose, L. C. (2006). Financial survival analysis of defaulted debtors, *Journal of Operational Research Society*, 57, 630-636.
2. Atz, U., Bholat, D., (2016) Peer-to-peer lending and financial innovation in the United Kingdom, Bank of England, Staff Working Paper No. 598.
3. Baba, N. and Goko, H. (2006). Survival analysis of hedge funds, Bank of Japan, Working Papers Series No. 06
4. Buis, M. L. (2006) An introduction to Survival Analysis, Working paper, Retrieved from <http://maartenbuis.nl/wp/survival.pdf>
5. Carling, K., Jacobson, T. and Roszbach, K. (1998). Duration of consumer loans and bank lending policy: dormancy versus default risk, Working Paper Series in Economics and Finance No. 280, Stockholm School of Economics.
6. Chen, D., Han, C., (2012) A Comparative Study of online P2P Lending in the USA and China, *Journal of Internet Banking and Commerce*, 17, 1-15
7. De Roure, C., Pelizzon, L., Tasca, P. (2016) How does P2P lending fit into the consumer credit market?, Discussion Paper Deutsche Bundesbank, 30, Retrieved from https://www.bundesbank.de/Redaktion/EN/Downloads/Publications/Discussion_Paper_1/2016/2016_08_12_dkp_30.pdf?__blob=publicationFile
8. Duarte J., Siegel S., Young L. (2012) Trust and Credit: The Role of Appearance in Peer-to-peer Lending, Oxford University Press on behalf of The Society for Financial Studies, doi:10.1093/rfs/hhs071
9. Glennon, D. and Nigro, P. (2005). Measuring the default risk of small business loans: a survival analysis approach, *Journal of Money, Credit, and Banking*, 37, 923-947.
10. Golubnicij, D., (2012) Is Your Peer a Lemon? Relative Assessment of Risk Remuneration on the P2P Lending Market, (Master's thesis). Retrieved from <http://arc.hhs.se/download.aspx?MediumId=1529>
11. Käfer, B. (2016) Peer to Peer Lending – A (Financial Stability) Risk Perspective, Joint Discussion Paper Series in Economics by the Universities of Aachen · Gießen · Göttingen Kassel · Marburg · Siegen, Retrieved from <http://www.uni-marburg.de/fb02/makro/forschung/magkspapers>.
12. Malik, M. and Thomas L. (2006). Modelling credit risk of portfolio of consumer loans, University of Southampton, School of Management Working Paper Series No. CORMSIS-07-12.
13. Mason, R., Gunst, F.R. and Hess, J.L. *Statistical Design and Analysis of Experiments with Application to Engineering and Science*, Wiley-Interscience, 2003

14. Milne, A., Parboteeah, P (2016) The Business Models and Economics of Peer-to-Peer Lending, European Credit Research Institute (ECRI), Research Report No 17
15. Narain, B. (1992): Survival Analysis and the Credit Granting Decision. In: Thomas, L. C. – Crook, J. N. – Edelman, D. B. (eds): Credit Scoring and Credit Control. Oxford, Oxford University Press, 1992, 109-122.
16. Roszbach, K. (2003). Bank lending policy, credit scoring and the survival of loans, Sverriges Riksbank Working Paper Series No. 154
17. Serrano-Cinca C, Gutiérrez-Nieto B, López- Palacios L (2015) Determinants of Default in P2P Lending. PLoS ONE 10(10): e0139427. doi:10.1371/journal.pone.0139427
18. Stevenson, M. (2009) An Introduction to Survival Analysis, Retrieved from http://www.massey.ac.nz/massey/fms/Colleges/College%20of%20Sciences/Epicenter/docs/ASVCS/Stevenson_survival_analysis_195_721.pdf
19. Xu, X. (2016) Estimating Lifetime Expected Credit Losses Under IFRS 9, Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2758513

APPENDIX

Table 1: Share of term length in total portfolio of approved loans per year

Year of approval	Term	Mean	Min	Max	SD	Share in amount
2007	36 months	4,977,475	500	25,000	6,160	100.00%
2008	36 months	21,119,250	500	25,000	5,740	100.00%
2009	36 months	51,928,250	1,000	25,000	5,995	100.00%
2010	36 months	89,740,900	1,000	25,000	6,363	67.99%
2010	60 months	42,251,650	1,000	25,000	6,826	32.01%
2011	36 months	132,847,450	1,000	35,000	6,477	50.77%
2011	60 months	128,836,375	1,000	35,000	8,729	49.23%
2012	36 months	507,799,125	1,000	35,000	7,081	70.68%
2012	60 months	210,611,900	1,000	35,000	7,574	29.32%
2013	36 months	1,272,091,475	1,000	35,000	7,356	64.16%
2013	60 months	710,673,800	1,000	35,000	7,176	35.84%
2014	36 months	2,046,040,750	1,000	35,000	7,843	58.39%
2014	60 months	1,457,799,425	1,000	35,000	7,438	41.61%
2015	36 months	3,626,461,100	1,000	35,000	8,061	56.51%
2015	60 months	2,791,147,075	8,000	35,000	7,333	43.49%

Table 2: Share of loan purpose in total portfolio of approved loans per year

Year of approval	Term	Mean	Min	Max	SD	Share in amount
2007	Credit Card Payoff	732,125	500	25,000	5,839	14.71%
2007	Refinancing	1,879,325	700	25,000	5,706	37.76%
2007	Other	2,366,025	500	25,000	6,483	47.53%
2008	Credit Card Payoff	3,412,875	725	25,000	5,332	16.16%
2008	Refinancing	9,383,825	700	25,000	5,321	44.43%
2008	Other	8,322,550	500	25,000	6,178	39.41%
2009	Credit Card Payoff	6,364,775	1,000	25,000	5,680	12.26%
2009	Refinancing	24,367,150	1,000	25,000	5,756	46.92%
2009	Other	21,196,325	1,000	25,000	6,036	40.82%
2010	Credit Card Payoff	18,205,750	1,000	25,000	6,545	13.79%
2010	Refinancing	68,689,300	1,000	25,000	6,560	52.04%
2010	Other	45,097,500	1,000	25,000	6,143	34.17%
2011	Credit Card Payoff	34,899,475	1,000	35,000	7,480	13.34%
2011	Refinancing	144,057,250	1,000	35,000	8,177	55.05%
2011	Other	82,727,100	1,000	35,000	7,968	31.61%
2012	Credit Card Payoff	135,468,775	1,000	35,000	7,439	18.86%
2012	Refinancing	445,182,300	1,000	35,000	7,962	61.97%
2012	Other	137,759,950	1,000	35,000	8,472	19.18%
2013	Credit Card Payoff	482,708,150	1,000	35,000	7,672	24.35%
2013	Refinancing	1,251,639,675	1,000	35,000	8,008	63.13%
2013	Other	248,417,450	1,000	35,000	8,332	12.53%
2014	Credit Card Payoff	853,482,525	1,000	35,000	8,306	24.36%
2014	Refinancing	2,203,845,875	1,000	35,000	8,310	62.90%
2014	Other	446,511,775	1,000	35,000	8,561	12.74%
2015	Credit Card Payoff	1,627,463,800	1,000	35,000	8,507	25.36%
2015	Refinancing	3,939,344,925	1,000	35,000	8,401	61.38%
2015	Other	850,799,450	1,000	35,000	8,679	13.26%