The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0144-3585.htm

Factors determining default in P2P lending

Evangelia Avgeri
Department of Economics, University of Piraeus, Piraeus, Greece, and
Maria Psillaki
Department of Economics, University of Piraeus, Piraeus, Greece and Department of Accounting and Finance, Neapolis University Paphos, Pafos, Cyprus

Received 25 July 2023
Revised 17 August 2023
17 August 2023
Accepted 17 August 2023

Abstract

Purpose – The research documented in this paper aims to examine multiple factors related to borrowers' default in peer-to-peer (P2P) lending in the USA. This study is motivated by the hypothesis that both P2P loan characteristics and macroeconomic variables have influence on loan performance. The authors define a set of loan characteristics, borrower characteristics and macroeconomic variables that are significant in determining the probability of default and should be taken into consideration when assessing credit risk.

Design/methodology/approach – The research question in this study is to find the significant explanatory variables that are essential in determining the probability of default for LendingClub loans. The empirical study is based on a total number of 1,863,491 loan records issued through LendingClub from 2007 to 2020Q3 and a logistic regression model is developed to predict loan defaults.

Findings – The results, in line with prior research, show that a number of borrower and contractual loan characteristics predict loan defaults. The innovation of this study is the introduction of specific macroeconomic indicators. The study indicates that macroeconomic variables assessed alongside loan data can significantly improve the forecasting performance of default model. The general finding demonstrates that higher percentage change in House Price Index, Consumer Sentiment Index and S&P500 Index is associated with a lower probability of delinquency. The empirical results also exhibit significant positive effect of unemployment rate and GDP growth rate on P2P loan default rates.

Practical implications – The results have important implications for investors for whom it is of great importance to know the determinants of borrowers' creditworthiness and loan performance when estimating the investment in a certain P2P loan. In addition, the forecasting performance of the model could be applied by authorities in order to deal with the credit risk in P2P lending and to prevent the effects of increasing defaults on the economy.

Originality/value – This paper fulfills an identified need to shed light on the association between specific macroeconomic indicators and the default risk from P2P lending within an economy, while the majority of the existing literature investigate loan and borrower information to evaluate credit risk of P2P loans and predict the likelihood of default.

Keywords P2P, Marketplace lending, Loan default, United States

Paper type Research paper

1. Introduction

Peer-to-peer (P2P) or marketplace lending emerged soon after the global financial crisis and now is the leading alternative finance format globally. P2P credit allows direct matching of lenders' supply and borrowers' demand of funding, without the intermediation of traditional banking. In online marketplaces borrowers raise funding from multiple lenders (individuals or institutions). Despite the benefits of this new on-line lending channel, it remains a risky activity for lenders since the default risk in this lending market is still high. The aim of this study is to find what determines marketplace loan performance, investigating multiple factors.

JEL Classification — G23, G28, G14, G20, O16, D14

Funding: No funding was received for conducting this study.

Conflict of interest: The authors declare that they have no conflict of interest that are relevant to the content of this study.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Journal of Economic Studies

0144-3585
DOI 10.1108/JES-07-2023-0376

Credit risk stems from the possibility of the borrower defaulting payments, because of the inability or lack of willingness to pay them back. For investors it is essential to know the determinants of loan performance in order to focus on the most relevant influential factors when estimating whether a particular loan is worth an investment or not. Studying the default behavior is of great importance as the higher default rates in this market, compared to the corresponding rates in the banking system, may lead to financial stability disruptions and can evolve into a financial crisis, if the credit risk in the P2P market spread and contaminate the financial market.

Many studies that examine marketplace default risk use data of the earliest years of the operation of this market (see, e.g. Carmichael, 2014; Möllenkamp, 2017; Serrano-Cinca *et al.*, 2015). We examine 1,863,491 P2P loans with clear payment status outcome that were originated during the period 2007-2020Q3 via LendingClub, the largest online lender in the USA and worldwide.

Our analysis confirms that loan and borrower characteristics can indeed be used to predict probability of default. More specifically, our results show that characteristics explaining defaults are, among others, loan amount, loan maturity, number of delinquency incidences and recent credit inquiries. The most important factor of loan performance is the credit grade that is assigned by LendingClub. Higher credit grade loan is associated with lower default risk.

Our study aims to extend the P2P literature by investigating specific macroeconomic factors as determinants of default risk. We specifically examine the impact of GDP growth rate, Unemployment rate, House Price Index (HPI), Consumer Sentiment Index (CSI) and S&P500 Index on borrowers' probability of default.

Our results confirm the theoretical predictions, i.e. macroeconomic factors play a crucial role in consumer default. The empirical evidence shows a positive relationship between Unemployment rate and an increase in delinquencies and between GDP growth rate and delinquencies. On the other hand, we find that the HPI, CSI and S&P500 Index are negatively correlated with the defaults. Our results are supported by the literature on banks and traditional financial services (see, e.g. Skarica, 2014; Mateus, 2020; Crook and Banasik, 2012; Fallanca *et al.*, 2020; Ghosh, 2017). The findings of our empirical study reveal that the inclusion of these indicators improves model fit and prediction of default.

This paper contributes to the growing literature on P2P lending in two important ways. Firstly, this study shed light on the association between specific macroeconomic indicators and the default risk from P2P lending within an economy, while the majority of the existing literature investigates loan and borrower information to evaluate credit risk of P2P loans and predict the likelihood of default. Secondly, our conclusions are more robust and updated since, unlike previous studies, a more comprehensive dataset of default determinants is used spanning 14 years (from 2007 to 2020).

The rest of this paper is organized in the following order: Section 2 presents the literature review. Section 3 shows the statistical analysis for P2P loans. Section 4 provides empirical results for the determinants of loan default. The final section is outlying the conclusion and policy implications.

2. Literature review

The majority of FinTech lending literature has mainly focused on the explanation of the emergence and expansion of P2P credit and on the P2P loan performance and default risk.

Balyuk and Davydenko (2018) document that marketplace lending have evolved from trading venues into credit intermediaries and the most P2P loans being funded by institutional investors, such as banks. Rau (2020) investigates the determinants of the crowdfunding development and finds that the quality of regulation, the financial system inefficiency and the ease of Internet access all provide very robust links to crowdfunding volumes. Havrylchyk *et al.* (2019) examine the drivers of the expansion of P2P lending in the USA and find evidence that counties that were more affected by financial crisis, with weak banking competition and higher population density have more P2P loans per capita. Oh and Rosenkranz (2020) find that financial institutions' efficiency, financial literacy and lower branch and ATM penetration are positively related with the expansion of P2P lending.

Many researchers examine the risk related to the marketplace lending (see, e.g. Käfer, 2016; Đurović, 2017; Lenz, 2017; Setyaningsih *et al.*, 2019). Suryono *et al.* (2019) mention six core problems associated to P2P lending, namely information asymmetry, determination of borrower credit scores, moral hazard, investment decisions, platform feasibility and immature regulations. Zhao *et al.* (2021) discuss the credit risk contagion of P2P lending and find that the platform correlations, the susceptible immune rate, the elimination rate of the P2P platforms by regulatory agencies, the saturation coefficient and other factors affect the risk contagion in the Internet financial market.

Di Maggio and Yao (2021), using a consumer credit panel dataset, document that FinTech borrowers are more likely to default and exhibit higher indebtedness than borrowers from traditional financial institutions. Lots of studies in the FinTech literature focus on the determinants of P2P loan default by examining the performance of these loans. Most of these studies explore loan and borrower characteristics to evaluate credit risk of P2P loans and predict the likelihood of default (see, e.g. Carmichael, 2014; Möllenkamp, 2017). Serrano-Cinca *et al.* (2015), using a sample of 24,449 loans, issued through LendingClub from 2008 until 2014 find that factors that best explain default are loan purpose, annual income, current housing situation and indebtedness. They conclude that credit grade has the highest predictive probability, but their model can be improved by including other factors. Similarly, Jagtiani and Lemieux (2019) find that the assigned rating grades perform well in predicting loan performance over the two years after origination. Emekter *et al.* (2015) indicates that credit score, debt-to-income (DTI) ratio and FICO score play an important role in loan defaults. Đurović (2017) analyzes two loan characteristics, loan term length and loan purpose and finds that longer term loans are more risky than the shorter term ones and the least risky loans are those used for credit card payoff.

Polena and Regner (2018) define four loan risk classes (based on the assigned credit grade) and find that the borrower's and loan's information that identified as determinants for default in previous studies are only significant in specific loan classes. Chen *et al.* (2022) use transaction data from LendingClub originated from 2015 until 2018 to estimate the gross rate of return (ROR) on an individual loan base and their results reveal that borrowers' credit rating, loan interest rate, loan status and paid-month are the most critical factors to influence investors' ROR.

Lots of studies in banking literature show that economic conditions play a crucial role in consumer delinquency and thus should be taken into consideration when assessing loan performance. Louzis *et al.* (2012) investigating the determinants of non-performing loans in Greek banking system show that GDP, unemployment, interest rate and public debt have a strong effect on the level of NPLs in all loan categories. Similar study on the Italian banking system was conducted by Foglia (2022). Their empirical findings show that GDP and public debt have a negative impact on NPLs and unemployment rate and domestic credit a positive one. Wadud *et al.* (2020) provide evidence on the determinants of household loan delinquency for mortgages, credit card and auto loans in the USA. They report positive effect of unemployment rate on delinquency rates and adverse effect of current consumer sentiment and per capita income. Research of Mateus (2020) demonstrates that consumer sentiment and the S&P 500 index impact negatively on delinquency and default. Bofondi and Ropele (2011) examine the macroeconomic determinants of banks' loan quality in Italy and find that the growth rates of GDP, the stock prices index and house price index are negatively correlated with probability of defaulting on loans.

Recent studies are beginning to expand the research on loan default in P2P lending industry by investigating factors, other than loan and borrower characteristics. Nigmonov *et al.* (2022), utilizing a probit regression analysis, investigate the macroeconomic factors that influence default risk and show that a higher interest rate and inflation increase the probability of default in P2P lending. Croux *et al.* (2020) by including in their model, aside from the data provided by LendingClub, variables such as GDP growth, VIX Index and Russell 2000 Index, show that macroeconomic conditions also impact the likelihood of P2P loan default.

3. Preliminary analysis of P2P loans

3.1 Data selection

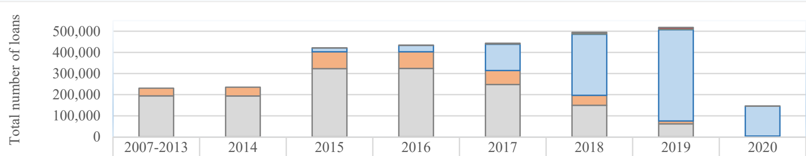
This study uses data from the LendingClub platform. Our data set contains information about 2,925,440 loans issued between June 2007 and September 2020 [1]. Over the loan origination study period LendingClub lent around $45bn to borrowers. The loans in our data set have seven different statuses: Fully Paid, Charged Off, Current, Default, Late (31–120 days), Late (16–30 days) and In Grace Period. Figure 1 shows the number of loans in each loan status per year.

Only the loans with exact ending resolution of the payment are useful, in order to distinguish between "good" and "bad" loans and estimate the probability of a loan default. Thus, we have removed the categories Current, Default, Late (31–120 days), Late (16–30 days) and In Grace Period, since they include loans that do not yet have a clear payment status outcome. Our final sample consist of funded loans whose outcome is known, i.e. "Charged Off" or "Fully Paid".

3.2 Statistical analysis

Table 1 reports the loan temporal distribution of the selected sample and presents statistics on the total number and amount of funded loans with known outcome across time, the status outcomes of these (fully paid or charged off) and the corresponding loan default rate per year (default rate is given by dividing the defaulted loans by total number of matured loans).

From the selected 1,863,491 loans, 363,826 defaulted and 1,499,759 repaid fully. The fully paid status has the largest share (80.5% of issued loans on average), while the percentage of default status increasingly grows every year, from 15.6% in 2007–2013 to 24% in 2018.



	2007-2013	2014	2015	2016	2017	2018	2019	2020
Late			464	714	3,349	6,518	7,243	576
In Grace Period			290	467	1,795	3,093	3,695	689
Current			17,504	30,120	124,067	288,448	431,741	141,162
Charged Off	36,102	41,748	79,708	79,083	66,246	47,226	13,564	58
Fully Paid	194,614	193,870	323,123	324,022	248,119	149,945	61,837	4,228

Source(s): Authors' calculations based on LendingClub database

Figure 1.
Distribution of loan statuses

Table 1. Loan temporal distribution

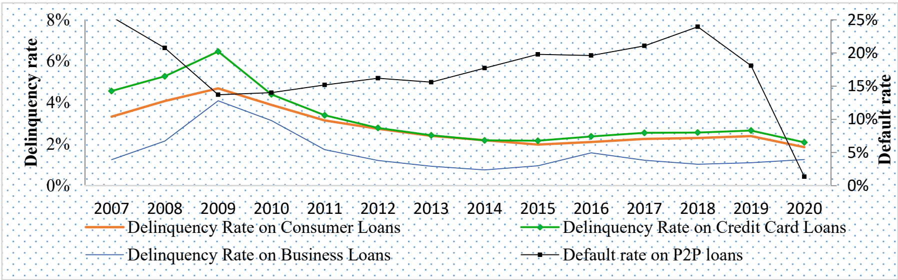
Year	2007–2013	2014	2015	2016	2017	2018	2019	2020	Total
Total number of loans	230,716	235,618	402,831	403,103	314,367	197,171	75,401	4,286	1,863,491
(Percentage)	12.4%	12.6%	21.6%	21.6%	16.9%	10.6%	4.0%	0.2%	100.0%
Amount of loans (000s)	3,172,878	3,503,640	6,048,843	5,761,210	4,474,635	3,023,308	1,126,882	59,130	27,170,525
(Percentage)	11.7%	12.9%	22.3%	21.2%	16.5%	11.1%	4.1%	0.2%	100.0%
Number of charged off loans	36,102	41,748	79,708	79,081	66,247	47,226	13,564	58	363,734
(Percentage)	9.9%	11.5%	21.9%	21.7%	18.2%	13.0%	3.7%	0.0%	100.0%
Amount of charged off loans (000s)	533,271	652,172	1,269,502	1,229,651	1,051,285	812,509	228,968	1,057	5,778,414
(Percentage)	9.2%	11.3%	22.0%	21.3%	18.2%	14.1%	4.0%	0.0%	100.0%
Number of fully paid loans	194,614	193,870	323,123	324,022	248,120	149,945	61,837	4,226	1,499,757
(Percentage)	13.0%	12.9%	21.5%	21.6%	16.5%	10.0%	4.1%	0.3%	100.0%
Amount of fully paid loans (000s)	2,639,607	2,851,468	4,779,341	4,531,559	3,423,350	2,210,799	897,913	58,073	21,392,111
(Percentage)	12.3%	13.3%	22.3%	21.2%	16.0%	10.3%	4.2%	0.3%	100.0%
Loan default rate	15.6%	17.7%	19.8%	19.6%	21.1%	24.0%	18.0%	1.4%	19.5%

Source(s): Authors' calculations based on LendingClub database

Overall, the ratio default per total loans is 19.5% with total amount of defaulted loans $ 5.8bn. The highest rate of default was recorded in 2018 (24%) with total losses of $ 812m.

Figure 2 presents a comparison of the default rate of P2P loans and delinquency rates on consumer loans, credit card loans and business loans over the time under consideration. Default rates in P2P lending are reflected over time at higher levels compared to traditional loans (average default rate 17.3 vs 2.78%) and peak in 2018, while the corresponding rate of traditional lending peaks in 2009.

Table 2 presents the self-reported loan purpose default statistics. It indicates that the loans that are supposed to be used for debt consolidation and credit card debts are the most frequent. The highest default rate is observed in loans for small business funding (28.7%), followed by loans regarding house buying (22.2%) and loans for moving purposes (22.1%). The less risky loan purpose for lenders is wedding loans and car purchase (repayment rate 87.6 and 85.8% respectively).



Source(s): Authors' calculations based on data provided by Board of Governors of the Federal Reserve System (US) and LendingClub

Figure 2.
Comparison of the level of default rates on P2P loans and delinquency rates on loans at all US commercial banks

Table 2.
Loan distribution by loan purpose

	Total loans		Charged off loans		Fully paid loans	
Loan purpose	Number	percentage of the total	Number	Default rate	Number	Repayment rate
Car	20,286	1.1%	2,871	14.2%	17,415	85.8%
Credit card	413,270	22.2%	69,725	16.9%	343,545	83.1%
Debt consolidation	1,064,797	57.1%	220,047	20.7%	844,750	79.3%
Educational	424	1.3%	88	20.8%	336	79.2%
Home improvement	124,100	5.4%	21,428	17.3%	102,672	82.7%
House	11,152	1.0%	2,481	22.2%	8,671	77.8%
Major purchase	41,880	2.0%	7,742	18.5%	34,138	81.5%
Medical	22,640	1.1%	4,635	20.5%	18,005	79.5%
Moving	13,249	1.8%	2,932	22.1%	10,317	77.9%
Other	114,039	4.9%	22,894	20.1%	91,145	79.9%
Renewable energy	1,263	0.2%	275	21.8%	988	78.2%
Small business	20,682	1.1%	5,940	28.7%	14,742	71.3%
Vacation	13,355	0.6%	2,382	17.8%	10,973	82.2%
Wedding	2,354	0.1%	293	12.4%	2,061	87.6%
Total	1,863,491	100.0%	363,733	19.5%	1,499,758	80.5%

Source(s): Authors' calculations based on LendingClub database

LendingClub evaluate borrowers' riskiness to default and classifies them in credit grades from A (low risk borrowers) to G (high risk borrowers). The grades are used to assign the interest rate of the funded loan (borrowers in grade G exhibit higher likelihood of delinquency and thus they are charged higher interest rate). Table 3 shows loans temporal distribution among the seven credit grades. The majority (29.4% of total loans) of issued loans belong to grade B, closely followed by credit grade C (28.4%). Credit grade G, as expected, presents the highest default rate (48.8%). The second worst loan performance is observed in F-graded loans (44.3% default rate). Finally, all loan categories reached their peak during the issued year 2018. Only 6.5% of total loans stem from credit grade A defaulted.

4. The empirical study

4.1 Variables and model

The research question in this study is to find the significant explanatory variables that are essential in determining the probability of default for LendingClub loans. We employ binary logistic regression to assess the capability of determinants analyzed to predict the loan default.

(1) *Dependent variable*

Since two outcomes are possible, the dependent variable is binary (or dichotomous) and presents the status of loan payment "Charged Off".

(2) *Independent variables*

The factors that usually predict the repayment of a loan or its default are expected to be loan and individual borrower characteristics. There are more than 100 variables in the data set of LendingClub, but not all are of interest for our analysis. Our variables were selected based on the results of previous studies on P2P default behavior and have been proved to play essential role in borrower solvency. Table A1 summarizes the P2P variables explanation. Except for *term, grade, employment length, home ownership* and *loan purpose*, all the other variables are continuous.

We expect that loan performance not only depends on the borrower and loan information but also on the overall state of the economy. Based on the theory and results of the relevant banking literature we consider the following factors, reflecting the health of general economy, as potential explanatory variables: *real GDP growth rate, Unemployment rate, House Price Index, S&P500 Index* and *Consumer Sentiment Index.*

A lot of studies (e.g. Croux *et al.*, 2020; Louzis *et al.*, 2012; Skarica, 2014) suggest that an increase in Gross Domestic Product (GDP) positively affects loan quality and decreases the likelihood of default. Thus, we expect real GDP growth to have a negative effect on default rate.

The unemployment rate, which is linked to the uncertainty regarding future income, is an important indicator for signaling borrower solvency and is commonly used to interpret the default rate. Rise in unemployment rates, can be expected to increase the hazard. Louzis *et al.* (2012) argue that unemployment is an important factor that affects consumer NPLs in the Greek banking sector, implying that an increase in unemployment levels has a negative impact on borrowers' ability to settle their obligations. A lot of studies find strong positive correlation between unemployment and consumers' delinquency (Mateus, 2020; Tobbak *et al.*, 2014; Bellotti and Crook, 2009; among others).

In the literature the House Price Index (HPI) has been assessed as a factor that negatively affect non-performing loans [2]. Usually, a positive shock in house prices results in a fall in delinquency volume. Ghosh (2015) finds that a rise in housing price reduces NPLs. Crook and Banasik (2012) concludes that the greater the fall in house prices, the greater the increase in

Table 3.
Temporal distribution by credit grade

Grade		2007–2013	2014	2015	2016	2017	2018	2019	2020	Total
A	Total loans	38,763	36,108	72,728	69,348	57,111	49,591	21,293	1,503	346,445 (18.6%)
	Charged off loans	2,230	1,954	4,035	4,122	3,780	4,525	1,730	13	22,389 (6.2%)
	Default rate	5.8%	5.4%	5.5%	5.9%	6.6%	9.1%	8.1%	0.9%	6.5%
B	Total loans	75,012	61,934	113,304	127,321	94,190	53,736	20,419	1,154	547,070 (29.4%)
	Charged off loans	8,406	6,875	14,524	16,813	13,997	10,424	3,007	21	74,067 (20.4%)
	Default rate	11.2%	11.1%	12.8%	13.2%	14.9%	19.4%	14.7%	1.8%	13.5%
C	Total loans	58,749	66,563	113,999	119,800	99,212	51,193	18,468	906	528,890 (28.4%)
	Charged off loans	10,372	12,369	24,738	26,529	24,891	15,215	4,255	11	118,380 (32.5%)
	Default rate	17.7%	18.6%	21.7%	22.1%	25.1%	29.7%	23.0%	1.2%	22.4%
D	Total loans	33,908	42,987	59,015	53,399	40,830	30,952	13,955	723	275,769 (14.8%)
	Charged off loans	7,646	10,920	18,205	16,834	13,572	11,750	4,093	13	83,033 (22.8%)
	Default rate	22.5%	25.4%	30.8%	31.5%	33.2%	38.0%	29.3%	1.8%	30.1%
E	Total loans	15,638	20,118	32,503	22,690	15,289	9,585	1,243	0	117,066 (6.3%)
	Charged off loans	4,443	6,610	12,715	9,350	6,192	4,199	473	0	43,982 (12.1%)
	Default rate	28.4%	32.9%	39.1%	41.2%	40.5%	43.8%	38.1%		37.6%
F	Total loans	7,009	6,223	9,200	8,259	4,940	1,715	16	0	37,362 (2.0%)
	Charged off loans	2,412	2,285	4,420	4,178	2,373	896	6	0	16,570 (4.6%)
	Default rate	34.4%	36.7%	48.0%	50.6%	48.0%	52.2%	37.5%		44.3%
G	Total loans	1,637	1,685	2,082	2,286	2,793	399	7	0	10,889 (0.6%)
	Charged off loans	593	735	1,071	1,255	1,441	217	0	0	5,312 (1.5%)
	Default rate	36.2%	43.6%	51.4%	54.9%	51.6%	54.4%	0.0%		48.8%

Note(s): The numbers in parentheses are the percentages of each loan category to the total number of loans. The numbers are from authors' calculations based on LendingClub database
Source(s): Author's own creation/work

delinquent debt and Wadud *et al.* (2020) show that house prices have an adverse and significant impact on mortgage delinquency rate.

P2P default probability may be affected by the performance of the US equity market. The impact of such Indices on loan delinquencies was assessed, for example, by Altman *et al.* (2005), Ghosh (2017) and Fallanca *et al.* (2020). A decrease in financial wealth is expected to increase the borrowers' probability of defaulting on loans, since the ability to service debts also decreases. We use historical data from S&P 500 Index [3], which reflects the overall return characteristics of the stock market as a whole, as a measure of changes in financial wealth.

Consumer Sentiment Index (CSI) is another important economic indicator and measures how optimistic consumers are about their financial situation and the overall economic outlook [4]. The effect of CSI is expected to be ambiguous. Increased households' optimism is likely to cause less loan defaults, but high consumer optimism may be expected to increase the demand for loans and, as a consequence, an increasing debt may lead to high levels of loan delinquencies. Mateus (2020) reports that an increase in consumer sentiment causes delinquency to decrease in auto loans, credit cards, mortgages and student loans in the USA.

Table A2 presents the macroeconomic variables, their sources and the expected signs.

The probability of default for a P2P loan can be described through the following equation:

$$\begin{aligned} \mathrm{D} = {} & \beta_0 + \beta_1 \mathrm{loan\_amnt} + \beta_2 \mathrm{term} + \beta_3 \mathrm{grade} + \beta_4 \mathrm{emp\_length} + \beta_5 \mathrm{home\_ownership} \\ & + \beta_6 \mathrm{annual\_inc} + \beta_7 \mathrm{purpose} + \beta_8 \mathrm{edti} + \beta_9 \mathrm{deling\_2yrs} + \beta_{10} \mathrm{unq\_last\_6mths} \\ & + \beta_{11} \mathrm{mths\_since\_last\_deling} + \beta_{12} \mathrm{mths\_since\_last\_record} + \beta_{13} \mathrm{open\_acc} \\ & + \beta_{14} \mathrm{revol\_bal} + \beta_{15} \mathrm{revol\_util} + \beta_{16} \mathrm{hargeoff\_within\_12\_mths} + \beta_{17} \mathrm{GDP} + \beta_{18} \mathrm{HPI} \\ & + \beta_{19} \mathrm{UR} + \beta_{20} \mathrm{S\&P500} + \beta_{21} \mathrm{CSI} + \varepsilon \end{aligned} \tag{1}$$

Here, D is a binary variable and represents probability of default (1 if the funded loan has been defaulted and 0 otherwise). Equation (1) analyzes the determinants of probability of default.

4.2 Descriptive statistics

Table 4 reports the summary statistics for all variables used in this study. Their number, mean, extreme values and median are reported.

Table A3 shows the correlation matrix table (Pearson's correlation coefficients) of all non-categorical variables. We can observe that the chosen independent variables are not highly correlated to each other and multicollinearity problems do not arise.

4.3 Empirical results

We first carry out non-parametric test in order to examine if there are differences in the chosen variables between two subsamples of loan status ("Charged off" and "Fully paid"). The Mann–Whitney U (Wilcoxon rank sum) test is used for comparing the two groups, where the null hypothesis is that the two samples come from identical populations (i.e. have the same median). Table A4 shows the results of the non-parametric test. We see that there are statistically significant differences between two groups and thus the null hypothesis is rejected. To further determine the exact effect of one of each variable on the probability of a P2P loan to default, we perform 3 logistic regression models. The estimation results of the models are reported in Table 5. Model 1 uses the variables provided by LendingClub (individual borrower and loan characteristics), model 2 contains only the macroeconomic

Variables	Number of observations	Mean	Std. Dev	Min	Max	Median
loan_amnt	1,863,491	14,580.44	8,969.92	500	40,000	12,000
Annual income	1,863,491	77,360.57	117,753.9	0	1.10e+08	65,000
dti	1,863,491	18.561	13.086	−1	999	17.71
Grade	1,863,491	2.691	1.271	1	7	3
delinq_2 yrs	1,863,491	0.313	0.875	0	42	0
inq_last_6 mths	1,863,491	697.658	32.705	610	845	690
mths_since_last_delinq	1,863,491	701.658	32.705	614	850	694
mths_since_last_record	1,863,491	0.618	0.918	0	33	0
open_acc	916,981	34.427	21.907	0	226	31
revol_bal	1,863,491	11.604	5.575	0	90	11
revol_util	1,863,491	0.208	0.589	0	86	0
chargeoff_within_12_mths	1,863,491	4.652	3.196	0	64	4
Employment length over 10 years (Y/N)	1,863,491	0.328	0.469	0	1	0
Employment length under 10 years (Y/N)	1,863,491	0.605	0.488	0	1	1
Term (maturity 36 months)	1,863,491	0.749	0.433	0	1	1
GDP	1,863,491	2.239	0.575	−3.4	2.92	2.28
HPI	1,863,491	4.794	1.688	−5.61	5.57	5.11
UR	1,863,491	5.251	1.231	3.6	9.55	4.9
S&P500	1,863,491	9.797	10.743	−38.49	29.6	9.54
CSI	1,863,491	4.724	5.353	−18.3	28.19	5.63

Note(s): Credit grade “1” is the loan category of A, which is the least risky class of loans. Credit grade “7” is the loan category of G (high risk borrowers)
Source(s): Author’s own creation/work

Table 4.
Descriptive statistics of all variables

variables (exogenous economic factors) and the overall model 3 uses all the explicative variables.

Firstly, we use goodness-of-fit measures by means of the Hosmer–Lemeshow test and the method of Akaike information criterion (AIC) to compare the 3 models and determine which one is the best fit for the data. Both of them indicate that the model 3 is the most adequate in explaining the status loans. Consequently, the inclusion of macroeconomic variables clearly improves our model.

The majority of independent variables in model 3 are statistically significant at 1% level in explaining the probability of a loan default and indicate the expected association (positive or negative) with the dependent variable. The delinquency probability for a typical P2P loan can be determined using the Odds ratio reported. Credit grade is an important factor determining loan performance and default likelihood. As expected, the lower the credit grate the riskier the loan. The results show a descending order from A (low risk borrowers) to G (high risk borrowers). The estimated exponentiated coefficients for each credit grade are significant at the 1% level. The variable with the highest predictive capability of all in the study is one reporting whether a borrower was assigned with grade A. Going up from grade B to grade A is associated with a decrease of almost 80% in the odds of becoming a defaulted loan.

Another important factor determining the loan outcome is the self-recorded loan purpose. The default probability is high in loans used for small business funding, medical and moving purposes, while loans for wedding expenditures, car purchase and credit card bear lower default risk, with the odds ratio being 0.9213, 0.8541 and 0.9694 respectively.

Dependent variable: Defaulted loans	Model 1		Model 2		Model 3	
Variables	Odds ratio	Robust std. err	Odds ratio	Robust std. err	Odds ratio	Robust std. err
_cons	1.8361***	0.1024	1.2023	8.9308	2.5198***	3.9348
loan_amnt	1.0001***	3.6145			1.0001***	3.6155
dti	1.0065***	0.0005			1.0064***	0.0005
delinq_2 yrs	1.0215***	0.0027			1.0206***	0.0028
inq_last_6 mths	0.9947***	0.0001			0.9948***	0.0001
mths_since_last_delinq	1.0002***	0.0001			1.0002**	0.0001
mths_since_last_record	1.0590***	0.0032			1.0625***	0.0032
revol_bal	0.9938***	0.0006			0.9939***	0.0006
revol_util	1.0297***	0.0043			1.0268***	0.0043
chargeoff_within_12_mths	1.0502***	0.0010			1.0492***	0.0010
Term (maturity 36 months)	0.5944***	0.0042			0.5979***	0.0042
Credit grade A (Y/N)	0.1939***	0.0061			0.1898***	0.0060
Credit grade B (Y/N)	0.3294***	0.0097			0.3241***	0.0094
Credit grade C (Y/N)	0.4910***	0.0140			0.4843***	0.0138
Credit grade D (Y/N)	0.6406***	0.0183			0.6324***	0.0180
Credit grade E (Y/N)	0.7893***	0.0229			0.7751***	0.0225
Credit grade F (Y/N)	0.9123***	0.0286			0.9068***	0.0285
Employment length over 10 years (Y/N)	0.8993***	0.0107			0.9007***	0.0107
HS:ANY	1.1023	0.1522			1.1367	0.1297
HS: NONE	1.3155	0.4561			1.3339	0.6733
HS: OTHER	0.5426	0.2864			0.5351	0.4009
HS: OWN	0.5426***	0.1114			0.5336***	0.0108
HS: RENT	1.4337***	0.0088			1.4330***	0.0091
Annual income	0.8065***	0.0144			0.8061***	0.0143
LP: car	0.8537***	0.0275			0.8541***	0.0275
LP: credit_card	0.9704***	0.0127			0.9694***	0.0127
LP: debt_consolidation	1.0231**	0.0120			1.0220**	0.0119
LP: home_improvement	1.0367***	0.0156			1.0328***	0.0120
LP: house	1.0323	0.0357			10.0296	0.0356
LP: major_purchase	0.0394*	0.0227			1.0382*	0.0227
LP: medical	1.1206***	0.0297			1.1198***	0.0296
LP: moving	1.1210***	0.0365			1.1198***	0.0365
LP: renewable_energy	1.1292	0.1143			1.1179	0.1159
LP: small_business	1.4581***	0.0369			1.4569***	0.0379
LP: vacation	0.9704	0.0343			0.9687	0.0344
LP: wedding	0.8567***	0.0146			0.9213***	0.1125
open_acc	0.9999***	0.0001			0.9990***	0.0001
GDP			1.1242***	0.0059	1.1273***	0.0111
HPI			0.9872***	0.0017	0.9472***	0.0049
UR			1.1915***	0.0025	1.1835***	0.0102
S&P500			0.9981***	0.0002	0.9949***	0.0003
CSI			0.9928***	0.0005	0.9880***	0.0009
Observations	1,812,335		1,863,491		1,812,335	
Pseudo R2	0.0773		0.0031		0.1815	
Year effect	Yes		Yes		Yes	
US States effect	Yes		Yes		Yes	
Log Likelihood	−418817.16		−917037.65		−413800.34	
Akaike's information criterion (AIC)	1,635,466		1,834,087		1617.106	
Hosmer and Lemeshow's test (*p*-value)	0.034		0.000		0.736	

Note(s): Coefficients and standard errors are reported as odds ratios. All models specifications employ robust standard errors in order to handle potential heteroscedasticity or models misspecification. Year and US States effects are incorporated in each regression model to address the problems of period effects and state-level effects respectively
The base value of model for credit grade is Credit Grade G (Grade G is the loan category with the highest assigned credit risk). HS stands for Home Status and the base value for homeownership is Mortgage. LP stands for Loan Purpose and the educational purposes was taken as the base group. The base value for loan maturity (term) is 60 months and for employment length is less than 10 years. The definitions of the variables are in Table A1 and A2. The levels of significance is noted by * for 10%, ** for 5% and *** for 1%
Source(s): Author's own creation/work

Table 5.
Binary logistic regression results of loan default

The loan amount is positively correlated with loan default. The higher the loan amount the higher the probability of default on a P2P loan. However, the influence of this factor seems to be quite low. This finding is similar to results in Polena and Regner (2018) study.

The annual income positively and significantly affects the probability of a loan success. Borrowers with high annual income are less likely to default. The odds ratio is 0.8061, suggesting that for one unit increase in the annual income we expect to see a decrease of about 20% in the odds of defaults.

Debt-to-income (dti) ratio and number of 30+ days past-due incidences of delinquency in the last two years are positive determinants of the likelihood of borrowers' default. These results are consistent with Croux *et al.* (2020), who also find that borrowers with higher dti or with a higher number of credit inquiries are more likely to default.

One more open credit line in the borrower's credit file reduces the loan default probability, with odds ratio being 0.99 (1% percent reduction in the odds ratio). The coefficient of the number of inquiries in six months are interpreted on the same basis. This is in accordance with the study of Möllenkamp (2017), although in the conclusion of his analysis the influence of the two variables were even stronger.

The factor number of charge offs during last year has a positive and statistically significant impact on default. The higher the number of charge offs within 12 months, the higher the likelihood of a P2P loan default.

The housing situation is a significant predictor for borrowers' default. The indicator of whether a borrower is a homeowner shows high economic significance and decreases the odds of delinquent behavior. On the contrary, borrowers who rent their home have a higher risk of default. This result agrees with Croux *et al.* (2020) and Polena and Regner (2018) findings.

Study's results also show that short-term loans (36 months) are associated with a lower likelihood of default. This finding is in congruence with the study of Durović (2017) who also found positive relation between loan maturity and default risk. In contrast, as the employment length of the borrower increases the odds of loan success.

The macroeconomic variables, that this research mainly focuses on, are statistically important and strongly impact the likelihood of loan being default.

The most important macroeconomic determinant for increasing the odds of loan default is unemployment rate. A positive increase of about 18% in the odds of becoming a loan defaulted is expected when the proportion of the population in unemployment increases one percent. The finding of a strong negative impact of unemployment rate on P2P loan quality is consistent with the existing financial literature (e.g. Carmichael, 2014; Fallanca *et al.*, 2020).

The coefficient of GDP growth rate is statistically significant but, surprisingly, it indicates positive relationship with loan defaults, implying that an improvement in the growth of GDP results in a higher likelihood of delinquency. Our findings contradicts, apparently, several empirical results, such as those in the study of Croux *et al.* (2020), who indicates a negative relationship between GDP growth rate and delinquency of P2P loans. Louzis *et al.* (2012), also, show that consumer and mortgage NPLs are negatively related to the GDP growth rate. The finding of our study in the light of the above could be attributed to the explanation that the high economy growth may be increase the demand for P2P loans and an increasing debt in the economy may lead to high default rates in the long run.

Our study is the first in the marketplace lending market studies that examine the impact of HPI, CSI and S&P500 Index on default probability of P2P loans.

Changes in HPI seems to have impact on default. The higher the growth rate of HPI at the time a loan is originated the lower the likelihood of a borrower to default. A 1% increase in HPI is associated with a 0.06% points reduction in the odds of loan default. This magnitude is comparable to a previous result from Bofondi and Ropele (2011). They explore the macro factors affecting household and business loans in Italy and find that an increase in HPI by 1% is associated with a decrease of 0.27% in new bad loans ratio.

Consistent with previous studies in traditional finance, the results of model 3 show that there is a negative relationship between CSI and likelihood of default. Mateus (2020) examining the factors impact the delinquency rates of auto loans, credit cards, mortgages and student loans in the United States, finds that consumer optimism has significant and negative effect on loan quality. Wadud *et al.* (2020), find that consumers' sentiment in the American states reduces mortgages and automobile delinquencies, whereas raises credit card loan defaults. Fuinhas *et al.* (2019) claims, however, that there is a positive relationship between consumers' sentiment and the proportion of student borrowers in delinquency or default.

Finally, the S&P500 Index is negatively associated with defaults. Although the magnitude is not strong this finding indicates that the positive percentage change in the annual returns of the Index lowers the likelihood the P2P borrower will default. This finding falls in line with the study of Mateus (2020) who argue that the S&P500 is statistically significant and has a negative effect on delinquency of credit cards. A similar negative relation for the real estate loans has been found in the empirical work of Fallanca *et al.* (2020).

5. Concluding remarks

The probability of default, as a cornerstone of credit risk, is the most central issue when the performance of P2P loans is assessed. Information asymmetry between lenders and borrowers remains an important problem existing in marketplace lending. Analysis of the determinants of loan default can significantly help investors make more accurate assessment for borrowers' credit riskiness and may resolve the issues of adverse selection and moral hazard.

This research examines the default determinants of P2P loans using an extended dataset of almost 2 million loans with clear ending resolution issued through the LendingClub from 2007 to 2020Q3, the longest analyzed period compared to previous corresponding studies. We investigate the impact of loan and borrower characteristics together with macroeconomic factors on P2P loan delinquencies utilizing logistic regression analysis. A binary logistic model of a total of 21 explanatory variables is proposed to predict loan default.

Our results, consistent with previous studies, confirm that loan and borrower information can indeed predict the likelihood of loan default.

Except from loan and borrower's characteristics, research interest during the last few years has begun to turn to exogenous economic conditions. Recent studies include macroeconomic factors to explain the reasons that lead P2P borrowers to default. Empirical evidence show that default is influenced by factors such as inflation, interest rate, GDP growth, Unemployment rate and VIX Index. Our findings, in line with existing studies, show that the Unemployment rate positively and significantly influence loan default. Higher unemployment rates is linked to higher default rates.

However, the novelty of our study is that we introduced HPI, S&P Index and CSI as drivers that explain the defaults in P2P lending market. Consistent with theoretical predictions of traditional financial market literature, our findings reveal significant impact of HPI, CSI and S&P500 Index on likelihood of P2P default. Higher percentage change in HPI, CSI and S&P500 Index in the year of a loan issued seems to be associated with a lower probability of delinquency.

To sum up, this study contributes to the growing literature by providing a deeper understanding of the predictors of loan default. The empirical findings reveal that alternative data should be utilized to identify borrowers' creditworthiness. Macroeconomic factors play an important role in borrowers' delinquency and should be taken into consideration when assessing credit risk. There are important implications of our findings for researchers, lenders and policy makers.

When evaluating borrowers' riskiness to default, data on the country's economic situation should also be used so that the lender can identify good borrowers from subprime ones and make profitable credit decisions. This study can encourage further research in the field of P2P lending credit risk regarding the investigation of additional macroeconomic variables as possible determinants, not only in the USA but also in other countries with developed online lending market.

Notes

1. Yash. 2020. LendingClub 2007-2020Q3. Kaggle. https://www.kaggle.com/ethon0426/lending-club-20072020q1?select = Loan_status_2007-2020Q3.gzip. Date accessed: 2022-5-13.
2. The authors use the FHFA House Price Index (FHFA HPI), which is a broad measure of the movement of single-family house prices based on data from all 50 states and over 400 American cities that extend back to the mid-1970s.
3. The S&P 500 index measures the value of the stocks of the 500 largest corporations by market capitalization listed on the New York Stock Exchange or Nasdaq.
4. In this study the authors use the Michigan Consumer Sentiment Index (MCSI). MCSI is a monthly survey of consumer confidence levels in the United States conducted by the University of Michigan. The survey is based on telephone interviews that gather information on consumer expectations for the economy.

References

Altman, E., Brady, B., Resti, A. and Sironi, A. (2005), "The link between default and recovery rates: theory, empirical evidence and implications", *The Journal of Business*, Vol. 78 No. 6, pp. 2203-2228.

Balyuk, T. and Davydenko, S. (2018), "Reintermediation in FinTech: evidence from online lending", *Working Paper*, University of Toronto.

Bellotti, T. and Crook, J. (2009), "Credit scoring with macroeconomic variables using survival analysis", *Journal of the Operational Research Society*, Vol. 60 No. 12, pp. 1699-1707.

Bofondi, M. and Ropele, T. (2011), "Macroeconomic determinants of bad loans: evidence from italian banks", *Bank of Italy Occasional Paper*, No. 89.

Carmichael, D. (2014), "Modeling default for peer-to-peer loans", *SSRN Electronic Journal*, pp. 713-743, doi: 10.2139/ssrn.2529240. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2529240 (accessed 05 June 2023).

Chen, P., Lo, S. and Tang, H. (2022), "What if borrowers stop paying their loans? Investors' rates of return on a peer-to-peer lending platform", *International Review of Economics and Finance*, Vol. 77, pp. 359-377.

Crook, J. and Banasik, J. (2012), "Forecasting and explaining aggregate consumer credit delinquency behavior", *International Journal of Forecasting*, Vol. 28 No. 1, pp. 145-160.

Croux, C., Jagtiani, J., Korivi, T. and Vulanovic, M. (2020), "Important factors determining Fintech loan default: evidence from a LendingClub consumer platform", *Journal of Economic Behavior and Organization*, Vol. 173, pp. 270-296.

Di Maggio, M. and Yao, V. (2021), "Fintech borrowers: lax screening or cream-skimming?", *Review of Financial Studies*, Vol. 34 No. 10, pp. 4565-4618, 3224957.

Durović (2017), "Estimating probability of default on peer to peer market – survival analysis approach", *Journal of Central Banking Theory and Practice*, Vol. 6 No. 2, pp. 149-167.

Emekter, R., Tu, Y., Jirasakuldech, B. and Lu, M. (2015), "Evaluating credit risk and loan performance in online peer-to-peer (P2P) lending", *Applied Economic*, Vol. 47 No. 1, pp. 54-70.

Fallanca, M., Forgione, A. and Otranto, E. (2020), "Forecasting the macro determinants of bank credit quality: a non-linear perspective", *The Journal of Risk Finance*, Vol. 21 No. 4, pp. 423-443, doi: 10.1108/JRF-10-2019-0202.

Foglia, M. (2022), "Non-performing loans and macroeconomics factors: the Italian case", *Risks*, Vol. 10 No. 1, 21, doi: 10.3390/risks10010021.

Fuinhas, J.A., Moutinho, V. and Silva, E. (2019), "Delinquency and default in USA student debt as a proportional response to unemployment and average debt per borrower", *Economies*, Vol. 7 No. 4, pp. 1-16.

Ghosh, A. (2015), "Banking-industry specific and regional economic determinants of non-performing loans: evidence from US states", *Journal of Financial Stability*, Vol. 20, pp. 93-104.

Ghosh, A. (2017), "Sector-specific analysis of non-performing loans in the US banking system and their macroeconomic impact", *Journal of Economics and Business*, Vol. 93, pp. 29-45.

Havrylchyk, O., Mariotto, C., Rahim, T.-U. and Verdier, M. (2019), *The Expansion of the Peer-To-Peer Lending and Barriers to Entry*, Unpublished Working Paper, University of Lille.

Jagtiani, J. and Lemieux, C. (2019), "The roles of alternative data and machine learning in fintech lending: evidence from the lending club consumer platform", *Financial Management*, Vol. 48 No. 4, pp. 1009-1029.

Käfer, B. (2016), "Peer to peer lending – a (financial stability) risk perspective", in *Joint Discussion Paper Series in Economics by the Universities of Aachen*, No. 22-2016.

Lenz, R. (2017), "Peer-to-Peer lending: opportunities and risks", *European Journal of Risk and Regulation*, Vol. 7 No. 4, pp. 688-700.

Louzis, D.P., Vouldis, A.T. and Metaxas, V.L. (2012), "Macroeconomic and bank-specific determinants of non-performing loans in Greece: a comparative study of mortgage, business and consumer loan portfolios", *Journal of Banking and Finance*, Vol. 36 No. 4, pp. 1012-1027.

Mateus, J.B.L. (2020), *The Impact of Unemployment and Income on Delinquency and Default in the USA Universidade da Beira Interior*, ProQuest Dissertations Publishing, Portugal, 28762605.

Möllenkamp, N. (2017), *Determinants of Loan Performance in P2P Lending*, Thesis, University of Twente.

Nigmonov, A., Shams, S. and Alam, K. (2022), "Macroeconomic determinants of loan defaults: evidence from the U.S. peer-to-peer lending market", *Research in International Business and Finance*, Vol. 59, 101516.

Oh, E.Y. and Rosenkranz, P. (2020), "Determinants of peer-to-peer lending expansion: the roles of financial development and financial literacy", *Asian Development Bank Economics Working Paper Series*, Vol. 613.

Polena, M. and Regner, T. (2018), "Determinants of borrowers' default in P2P lending under consideration of the loan risk class", *Games*, Vol. 9 No. 4, p. 82.

Rau, P.R. (2020), "Law, trust, and the development of crowdfunding", *SSRN Working Paper*, available at: https://ssrn.com/abstract=2989056

Serrano-Cinca, C., Gutierrez-Nieto, B. and Lopez-Palacios, L. (2015), "Determinants of default in P2P lending", *PLoS One*, Vol. 10 No. 10, e0139427.

Setyaningsih, T., Murti, N.W. and Nugrahaningsih, P. (2019), "Fintech based peer to peer lending: an opportunity or a threat?", *Riset Akuntansi dan Keuangan Indonesia*, Vol. 4 No. 3, pp. 122-133.

Skarica, B. (2014), "Determinants of non-performing loans in Central and Eastern European countries", *Financial Theory and Practice*, Vol. 38 No. 1, pp. 37-59.

Suryono, R.R., Purwandari, B. and Budi, I. (2019), "Peer to peer (P2P) lending problems and potential solutions: a systematic literature review", *Procedia Computer Science*, Vol. 161, pp. 204-214.

Tobback, E. M., Gestel, D.V. and Bart, T.B. (2014), "Forecasting Loss Given Default models: impact of account characteristics and the macroeconomic state", *Journal of the Operational Research Society*, Vol. 65, doi: 10.1057/jors.2013.158.

Wadud, M., Ahmed, A., Joher, H. and Xueli, T. (2020), "Factors affecting delinquency of household credit in the U.S: does consumer sentiment play a role?", *The North American Journal of Economics and Finance*, Vol. 52, 101132.

Zhao, C., Li, M., Wang, J. and Ma, S. (2021), "The mechanism of credit risk contagion among internet P2P lending platforms based on a SEIR model with time-lag", *Research in International Business and Finance*, Vol. 57, 101407.

Appendix

Variable name	Description	Value
loan_amnt	The listed amount of the loan applied for by the borrower	Amount (US $)
Term	The number of payments on the loan. Values are in months and can be either 36 or 60	Dummy variable
Grade	LC assigned loan grade	Dummy variable
emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years	Dummy variable
home_ownership	The home ownership status provided by the borrower during registration or obtained from the credit report. Our values are: RENT, OWN, MORTGAGE, OTHER	Dummy variable
annual_inc	The self-reported annual income provided by the borrower during registration	Amount (US $)
Purpose	A category provided by the borrower for the loan request	Dummy variable
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, divided by the borrower's self-reported monthly income	Ratio
delinq_2 yrs	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past two years	Number
inq_last_6 mths	The number of inquiries in past six months (excluding auto and mortgage inquiries)	Number
mths_since_last_delinq	The number of months since the borrower's last delinquency	Number
mths_since_last_record	The number of months since the last public record	Number
open_acc	The number of open credit lines in the borrower's credit file	Number
revol_bal	Total credit revolving balance	Number
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit	Number
chargeoff_within_12_mths	Number of charge-offs within 12 months	Number

Note(s): The description of the data is from the Data Dictionary file provided by LendingClub

Table A1.
Definition of the P2P variables

Variable name	Notation	Value	Data source	Expected sign
Real GDP growth rate	GDP	annual growth rate	Bureau of Economic Analysis	–
House price index	HPI	annual growth rate	US Federal Housing Finance Agency	–
Unemployment rate	UR	Percentage (annual)	US Bureau of Labor Statistics	+
Standard and poor's 500 Index	S&P500	annual growth rate	Macrotrends.net	–
Consumer sentiment index	CSI	annual growth rate	Survey of Consumers – University of Michigan	–

Note(s): All the variables, except for unemployment rate, are in growth values at loan origination year and expand between 2007 and 2020. Real GDP growth rate represents the average annual growth rate of GDP in the U.S. economy. S&P500 annual growth rate shows the change in the annual return from the previous year

Table A2.
Macroeconomic variables used in the study

Table A3. Correlation matrix

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1)	Default	1																
(2)	loan_amnt	0.072	1															
(3)	Annual income	−0.023	0.187	1														
(4)	dti	0.065	0.038	−0.081	1													
(5)	delinq_2 yrs	0.014	−0.005	0.024	−0.010	1												
(6)	inq_last_6 mths	−0.021	0.100	0.039	−0.040	−0.177	1											
(7)	mths_since_last_delinq	−0.121	0.100	0.039	−0.040	−0.177	1	1										
(8)	mths_since_last_record	−0.065	−0.021	0.019	−0.010	0.022	−0.087	−0.087	1									
(9)	open_acc	−0.008	−0.016	−0.029	0.010	−0.553	0.099	0.099	0.013	1								
(10)	revol_bal	0.021	0.182	0.085	0.195	0.051	0.018	0.018	0.137	−0.037	1							
(11)	revol_util	0.024	−0.059	−0.003	−0.027	−0.022	−0.190	−0.190	0.061	0.078	−0.015	1						
(12)	chargeoff_within_12_mths	0.094	0.008	0.037	0.096	−0.055	−0.097	−0.097	0.275	0.117	0.495	0.092	1					
(13)	GDP	0.034	0.044	−0.001	0.029	−0.003	0.003	0.003	−0.047	0.018	0.014	−0.002	0.005	1				
(14)	HPI	−0.032	0.059	0.016	0.069	0.038	−0.063	−0.063	−0.100	−0.013	0.070	0.069	0.056	0.265	1			
(15)	UR	0.049	−0.036	−0.023	−0.069	−0.016	−0.035	−0.035	0.116	−0.009	−0.049	−0.044	−0.081	−0.312	−0.273	1		
(16)	S&P500	−0.031	−0.021	0.003	−0.018	−0.006	0.004	0.004	0.035	−0.005	−0.014	−0.016	−0.035	−0.442	−0.087	0.231	1	
(17)	CSI	0.001	0.012	−0.011	0.001	0.011	−0.065	−0.065	0.019	−0.010	0.010	0.006	−0.024	0.538	−0.059	0.168	−0.285	1

Note(s): The results are calculated by the authors

Table A4.
Non-parametric test

Variables	Defaulted loans	Fully paid loans	*p*-value
loan_amnt	Higher	Lower	0.00
Annual income	Lower	Higher	0.00
dti	Higher	Lower	0.00
grade	Higher	Lower	0.00
delinq_2 yrs	Higher	Lower	0.00
inq_last_6 mths	Lower	Higher	0.00
mths_since_last_delinq	Higher	Lower	0.00
mths_since_last_record	Higher	Lower	0.00
open_acc	Lower	Higher	0.00
revol_bal	Higher	Lower	0.00
revol_util	Higher	Lower	0.00
chargeoff_within_12_mths	Higher	Lower	0.00
Term	Higher	Lower	0.00
emp_length	Lower	Higher	0.00

Note(s): The results are calculated by the authors

Corresponding author
Evangelia Avgeri can be contacted at: e.avgeri@unipi.gr