J. Int. Financ. Markets Inst. Money 86 (2023) 101801

Contents lists available at ScienceDirect

Journal of International Financial Markets, Institutions & Money

journal homepage: www.elsevier.com/locate/intfin

Peer-to-peer lending: Legal loan sharking or altruistic investment? Analyzing platform investments from a credit risk perspective☆

Barbara Dömötör*, Ferenc Illés, Tímea Ölvedi

Corvinus University of Budapest, Department of Finance, 1093, Fővám tér 8, Hungary

ARTICLE INFO

JEL codes:
G21
G23
G32

Keywords:
Alternative finance
Peer-to-peer lending
Information asymmetry
Credit rating
Scoring model

ABSTRACT

This paper analyzes the performance of peer-to-peer investments, the potential benefits of their information processing and the investor returns, based on the entire portfolio of the Estonian platform Bondora. We found that the platform's scoring model relies on different default probabilities across countries and is weak at predicting default within countries. Alternative information could improve the models, but our analysis could not confirm this benefit of the platform. The average internal rate of return on closed transactions was –4.17%, and 42% of the loans end with a negative IRR. We concluded that P2P borrowers in the European market are mainly high-risk, bank ineligible clients, accepting even loan-sharking level interest rates, which excludes altruistic motives of investors. Even so, investors are not compensated for the credit risk.

1. Introduction

Rapid technological changes in recent decades have contributed to financial disintermediation trends. Alternative financing providers have emerged or started to increase their market share by promising cheaper services and better financial inclusion through the exclusion of intermediary institutions (Polasik et al., 2020).

There are still many questions about the relevance, potential benefits and future of direct lending through an online platform. While the removal of the intermediary layer—and, hence, the regulatory obligations on banks—may be cost-effective and help absorb underbanked customers, it is not evident whether investors are compensated for the high credit and liquidity risk or whether their altruistic motives are needed to sustain the business. On the borrower side, it is also unclear whether better access to finance for subprime customers is in their interest or an access contributes to higher default rates and greater difficulties for the segment (Gosztonyi and Havran, 2021).

This paper is linked to two strands of research in the literature. The first is the role of alternative information in the risk assessment of platforms, which has so far been studied mainly in the US market. The second is whether online platforms substitute or complement banks and serve to improve financial inclusion.

We analysed the loan level data and all cash-flows of Bondora (https://www.bondora.com/en), the sixth largest peer-to-peer platform in Europe (p2pmarketdata.com). Bondora allows to their clients investing in personal debt starting at €1, and enables borrowers to receive funding directly from investors. The investments are denominated in EUR and are available for retail investors

☆ This research was supported by National Research, Development and Innovation Office - NKFIH, K-138826.

* Corresponding author.

E-mail addresses: barbara.domotor@uni-corvinus.hu (B. Dömötör), ferenc.illes@stud.uni-corvinus.hu (F. Illés), timea.olvedi@stud.uni-corvinus.hu (T. Ölvedi).

https://doi.org/10.1016/j.intfin.2023.101801
Received 10 August 2022; Accepted 2 July 2023
Available online 7 July 2023

mainly from the EU. At the time of our analysis, loans were available to borrowers in 4 countries: Estonia, Finland Slovakia and Spain. The platform itself provides the online marketplace and some services, such as credit rating of the applicants, without taking any credit risk.

In this study we address the following research questions:

1. Do P2P lending platforms have an advantage in information processing compared to traditional banks due to the incorporation of alternative information?
2. What is the performance profile of P2P investments for the lenders?

Our contribution to the literature is twofold. First, while the majority of studies on P2P lending focus on the US market, we analyze the European market in Bondora's very detailed database, which includes loan-level portfolio tables and all cash flows since the platform was launched. Second, in order to assess the performance of the loans, we conduct an ex-post cash-flow analysis, which to our knowledge has never been done before. By analyzing the cash-flow data of the platform, we can investigate not only the default defined by the platforms, which ignores all further possible payments, but also the losses realized ex-post.

We found that the platform's scoring model could not outperform our model, although we used the standard, publicly available data of the platform. Both our model and the platform's model rely on differences in default by country of origin, but they are poor at predicting default within countries.

Although alternative information could improve the predicting power of the models, there is no sign of benefits from using them. We could not confirm that the platform was able to reduce information asymmetry better than traditional financial intermediaries. This contradiction between our findings and the results of the literature could be due to the difference between the US and the continental European markets.

The ex-post analysis of the cash-flows shows that the average internal rate of return (IRR) is negative, and more than 40% of all transactions end with a negative IRR; thus, there is a net loss of investment. From a regulatory perspective, our results suggest that platform investors bear an uncompensated, maybe unforeseeable credit risk. The high default rate of non-Estonian borrowers reflects a decrease in willingness to pay with distance, that can be due to lower crossborder collection efficiency. This inefficiency combined with loan-sharking level interest rates (77.5% on average in the worst rating segment) may lead to adverse selection of the borrowers.

While harmonisation of regulation and improvement of investor protection is underway (Regulation (EU) 2020/1503 on European crowdfunding service providers for business is applied from November 2021), further requirements and specifications would be needed to improve the efficiency of cross-border collection and the transparency of platform investment performance. In addition to ex-ante data published by the platform, ex-post performance disclosure both in terms of returns and the accuracy of models could reduce information asymmetry.

The structure of the paper is as follows. The next section presents a summary of the literature, and then in Section 3, we present the analysis of the portfolio and cash-flow tables of Bondora. Finally, in Section 4, the conclusions are derived.

2. Literature review

The literature on platform lending is relatively new but quite extensive. The emergence of platform lending is frequently explained by the rapid technological changes of the last decades, which contributed to the disruption of many services in the economy (Goldstein et al, 2019). However, there is no evidence that the potential benefits of P2P lending platforms disqualify the relevance of banks.

The existence and the rationale of banks are consequences of market imperfections, such as transaction costs, liquidity shocks, and information asymmetry, which make perfect diversification impossible (Freixas & Rochet, 2008). Although P2P platforms are free of the considerable fixed costs of branch networks or employers, their cost efficiency is unsupportable if we compare their average cost of 3–4% of the intermediated amount (Morse, 2015) with the long-term stable intermediation cost of 2% that is present for banks (Philippon, 2016; Bazot, 2018). Platforms primarily only match investors and borrowers; thus, they also do not protect against liquidity shocks.

While P2P lending underperforms, compared with banks, with regard to transaction costs and liquidity insurance, these platforms may be more advantageous in that they reduce information asymmetry to a greater extent, as they use also soft and sensitive data given voluntarily and apply big data and artificial intelligence more flexibly than banks (Liu et al. 2020). Another advantage of these platforms is that they are still free of regulatory restrictions; thus, they can offer high-risk investment possibilities, without risking their own capital (Davis, 2016).

An often-emphasized argument for fintech companies such as peer-to-peer (P2P) platforms is their flexibility to apply the latest and most advanced data analysis methods, which ensures them a competitive advantage over traditional financial institutions (Duarte et al., 2012; Lin et al., 2013; Jagtiani and Lemieux, 2019; Feyen et al., 2021), however, this advantage can disappear, as banks are incentivized to improve their digital services to compete with the new challengers.

As our study relates to the importance and role of P2P lending, we focus on the strands of research that investigate the substitutive or complementary nature of this kind of alternative financing and the impact of alternative information in the lending process.

2.1. The role of platforms: substitutes or complements?

Thakor (2020) summarizes the literature on fintech around four main questions, one of which is the role of marketplace lending in financial intermediation. He concludes that as banks are unique in providing deposit insurance, P2P lending can complement banks, mainly if banks are more capital constrained, serving clientele unable to pose collateral.

The empirical literature on the role of platform lending is quite widespread, investigating the evolution of the market and the characteristics of P2P loans and comparing them to bank loans. However, only a few theoretical models aiming to situate P2P lending in financial intermediation are present in the literature. In Merton and Thakor's (2019) model of financial intermediation, financial institutions are financed by two types of partners: investors and customers. Investors are willing to take on risk in exchange for an appropriate risk-adjusted return, while customers demand financial services free of credit risk. The optimal contractual design is determined by the cost of insulating customers from the credit risk of the intermediary. According to this concept, the distinction between banks and the market (direct lending) disappears once the above costs, called customer contract fulfillment costs, become sufficiently large. In the case of financing frictions, direct lending can be an attractive alternative for at least a section of the customers. Liu et al. (2019) incorporate both social collateral and soft information into their model and show that both can reduce information asymmetry, making financing available even to small borrowers with limited assets. In contrast to the traditional lending market, low-risk borrowers can crowd out high-risk borrowers, and P2P platforms complement traditional banks by serving those who are not targeted for bank lending.

The majority of empirical evidence shows that platform borrowers are mainly underbanked customers with limited access to bank finance (Das, 2019; Maskara et al., 2021); the unavailability of other financing options encourages borrowers to turn to the platform. De Roure et al. (2016), examining the German market, also conclude that P2P lending platforms serve an underbanked segment of low-credit customers, which is out of the scope for conventional banks. This concept of collaboration is supported by Milne and Parboteeah (2016), who propose that P2P platforms have a complementary function in lending activity. Specifically, they supplement traditional banks because banks possess a few comparative advantages, which precludes platforms from competing with them. The two financing forms may cooperate in the future.

In contrast to the above results, Tang (2019) finds, when investigating the unsecured consumer loan market in the US, that a negative shock in the bank's credit supply lowers the quality of P2P platforms' credits. He concludes that the results confirm the role of platforms as a substitute for bank lending in serving infra-marginal bank borrowers while also complementing bank lending for small loans. These results also suggest that the credit expansion of P2P lending was based on borrowers who already have access to bank credit. Cornaggia et al. (2018) also highlight the substitutive role of P2P platforms, stressing that smaller banks suffer losses due to the decline in loan volumes, while large commercial banks are not affected. Following the restrictions of the COVID-19 pandemic, Najaf et al. (2022) argue, fintech P2P lending has become the most viable alternative credit option available to borrowers. Moreover, online services have the potential to augment or replace lending provided by traditional or conventional banking institutions.

From the investor's perspective, however, platform funding has higher risk and less transparency, and risk management is underdeveloped (Milne & Parboteeah, 2016). Although platform investing provides higher interest rates, risk-adjusted interest rates are comparable (De Roure et al., 2016), and it is questionable why unsecured P2P lending, which is not even covered by deposit insurance, is beneficial for lenders. The existence of the market under unfavorable lending conditions can be explained by investor preferences, as in the model of Berentsen and Markheim (2021), where altruistic investors are willing to finance even projects generating negative expected cash-flow.

2.2. Platforms' risk assessment: role of alternative information

Information asymmetry in financial intermediation relates to the problem of the lender having constrained knowledge of the borrower's creditworthiness. The consequences of information asymmetry can be considered to be specific forms of transactional costs, which can be reduced by monitoring and, hence, improving the efficiency of lending. Financial intermediaries create value by economizing monitoring costs (Diamond, 1984) and by having better access to borrowers' credit and account history or other public sources, such as bad debtor registry or legal processes. Banks, using their own capital to finance borrowers, also provide signals about the quality of the debtor. P2P platforms have little to no access to the previous financial history of the borrowers, and the verification of this information is also costly and, sometimes, even impossible. As P2P lending platforms do not offer credit from their own sources, the signaling effect is also less significant than in the case of banks. Additionally, these platforms have the advantage of applying big data analysis techniques and obtaining "soft information," which banks are not allowed to gather (Havrylchyk & Verdier, 2018). Earlier, in the case of crowdfunding campaigns, borrowers and lenders were aware of one another, and the social relationship facilitated the screening of borrowers. In the present day, this kind of proximity is atypical among borrowers and lenders; however, the narratives submitted voluntarily by borrowers can contain sensitive information and may have a high impact on investors' decisions.

The fact that platform lending is the most intense in consumer lending, where informational asymmetry is the highest, also confirms the importance of the information that banks are prohibited from collecting (Havrylchyk & Verdier, 2018). Information asymmetry and capital requirements make it unfeasible for banks to finance high-risk customers, even if they are willing to pay higher costs. By reducing information asymmetry, P2P platforms are seemingly able to reduce credit rationing. Although financial institutions have also started to use digital techniques to handle low-quality big data, the fact that machine learning tools work as a black box constrains their applicability for regulatory purposes (van Liebergen, 2017).

The information P2P platforms collect is mainly based on hard data on the borrower and the credit itself, as well as other local economic information, such as the location's criminal statistics or employment rates (Jagtiani & Lemieux, 2018). Borrowers may provide other data sources, such as an account of the utilities availed, public reports, and alternative lending payment history. The narratives applicants provide regarding their goals and credit purposes are also a source of soft information; however, their usefulness is not confirmed (Herzenstein et al, 2008). Emekter et al. (2015) examine the credit risk and performance models of platforms based on the data of the biggest P2P lending platform Lending Club. They find that besides the platform's credit grade, a few other variables, such as debt-to-income ratio, FICO score (credit score created by Fair Isaac Corporation), and revolving line utilization, also have significant explanatory power on loan default. The credit grading reflects the riskiness of the loan, but the higher interest rate charged is not enough to compensate the investors in the worst clientele.

Jagtiani and Lemieux (2019), in their analysis of the loans of Lending Club, find that the correlation between the assigned grade of the platform and the borrowers' FICO score declined from 80% to 35% from 2007 to 2014, and the platform's grades proved to perform better while predicting loan default. Their results confirm the benefits of alternative data used by fintech lenders. Having examined an extended sample (2007–2018) of the Lending Club consumer platform, Croux et al. (2020) also confirm the importance of alternative data on loan default. Das (2019) highlights the importance of alternative data in developing better credit models, which allows lenders to select creditworthy borrowers from the lower FICO score bucket, who are otherwise excluded from traditional financing. Cumming et al. (2020) also highlight the role of soft information, proving that the higher the risk the borrower faces (large amount needed or all or nothing financing form), the higher the length and readability of soft information provided. Hughes et al. (2022) compared the lending efficiency of LendingClub's with that of traditional financial institutions and they find that the platfom's credit evaluation to be more accurate, which they explain by the use of alternative data and complex modeling capability of fintech credit providers.

3. Performance analysis of Peer-to-Peer investments

Based on the theory and the empirical evidence on P2P platforms presented in the previous sections, we first examined the information processing of the P2P platform Bondora, an Estonian marketplace launched in 2009, to find evidence on the role of alternative information. Then, to answer our second research question, we analyzed the performance of the platform's investments.

Bondora provides different datasets, which are updated daily. We used two types of datasets: first, the raw data of the loan book containing all loans with different applicant-related and other variables (Dataset 1 and 2, detailed description in the next subsection) and, second, the historical payment table that includes all cash-flow series of each loan (Dataset 3). Table 1 contains the description of them.

The loan book data were downloaded on October 5, 2020, and on March 15, 2022. The data of the first period were used to build our benchmark model, while the data of the second period were used for the out-of-sample testing. Historical payments covering the principal and the interest amount paid in each month, were downloaded on May 24, 2022. These data were used to examine the performance of platform investments.

3.1. Risk assessment of the platforms

First, we analyzed the loan book containing all loans taken between December 10, 2012, and October 5, 2020. The original dataset covered 151,866 transactions and 112 variables, including different types of information: data regarding the characteristics, financial background and payment history of the applicant (e.g. the number and the amount of the pervious loans the borrower had before this application), standard information about the loan request (e.g. interest rate, loan amount, tenor) and technical data (e.g. loan ID, the form of bidding). The full list of the variables included in the original dataset is presented in the Appendix (Table 11).

During data cleaning framework, inconsistent records were eliminated (when the default date is missing but the loan is more than 90 days past due). It impacted 3,029 rows. Additionally, missing and invalid observations were excluded from the dataset (we considered a record invalid in case its value was not defined in the Bondore data description). Loans issued within a 12-month period (from October 6, 2019) were also eliminated, as to our definition of default 12 months after issuance is relevant. After data cleaning, 107,588 observations (loan transactions) remained in the sample.

To support the decision of investors, the platform performs a credit risk assessment for each applicant and assigns a rating based on its internal evaluation. The methodology is not publicly available; only the final rating is shared with investors. According to the literature, the rating of the platform is one of the main factors that impact investors' decisions. Better rating generally results in greater success in funding (Herzenstein et al. 2008, Emekter et al. 2015, Gavurova et al. 2018). Therefore, it is crucial that the rating appropriately reflects the risk of the potential borrower. Furthermore, according to the platforms, one of the main advantages of P2P

Table 1
Description of the datasets of bondora used for the analysis.

	Dataset 1	Dataset 2	Dataset 3
Downloaded	October 5, 2020	March 15, 2022	May 24, 2022
Data type	Loan level data (112 variables)	Loan level data (112 variables)	Cash-flow data (4,343,194 payments)
Number of raw data	151,866 loans	222,978 loans	243,453 loans
Data used for	Investigation and in-sample testing of the scoring model	Out-of-sample testing of the scoring model	Performance from investor perspective (based on IRR)
Data cleaning	Inconsistent, missing and invalid observation were excluded. Origination date between December 10, 2012 and October 5, 2019	Inconsistent, missing and invalid observation were excluded. Origination date between October 6, 2019 and March 15, 2021	Missing values, loans with current or NA status were excluded.
Number of loans used for the analysis	107,588	50,251	62,537 + 42,598

lending is their credit risk assessment process, as the platform applies alternative information for their assessment, besides the standard variables used by banks (among others Popescu, 2016; Croux et al., 2020; Hughes et al. 2022).

According to the platform's webpage, the current credit risk rating methodology was introduced in December 2014 to improve the previous rating practice and support risk-based pricing. As stated, the method is in line with the industry's best practices, commonly applied by the banking sector. Overall, eight rating bucket grades are determined on a scale from AA (lowest risk) to HR (highest risk). The rating classification is based on the expected loss (EL) intervals, calculated by the platform with the following formula:

$$EL(\%) = PD^{*}LGD^{*}EAD \tag{1}$$

where *PD* is the probability of default, *LGD* is the loss given default, and *EAD* is the exposure at default. The data is derived from three main sources: information provided by the applicant, an external credit bureau database, and behavioral information collected through the application process on the webpage (Bondora, 2021). It must be noted that this practice differs from banks' methodology, where the basis of rating assignment is the probability of default and not the expected loss.

The portfolio table provides information on the performance of the transaction by disclosing the number of days past due (DPD) and the date of default if the loan is in default. However, these default indicators are not fit our purpose, as industry practice considers a default to be a delay in payment of more than 90 days in the first 12 months after the loan is issued. Therefore, we created a default flag accordingly and considered a borrower to be in default if they were more than 90 days in arrears in the 12 months following the origination of the loan. In the following, we use this default definition for modeling and also when presenting ex-post default frequency. Since the portfolio table contains no information on the delayed amount, the materiality threshold is not included in the default definition. It is important to note that the default defined above does not represent the ultimate loss considered in subsection 3.4, as loans that are 90 days past due in the first year after issuance may recover and meet their obligations later. Similarly, loans that perform well in the first year may default later, leading to losses.

Table 2 below presents the distribution of the loans examined for the first period (2012–2019) based on the rating category provided by the platform at origination and the days past due (DPD).

The platform's credit risk assessment seems to be mostly reasonable. Claims with worse ratings usually showed worse performance. As we had no external data on the borrowers, such as the FICO score in the US, we could not compare the platform's ratings with banking models, as Jagtiani and Lemieux (2019) do, but our data support their findings.

For a high-level portfolio overview, the loans' main characteristics were examined. Table 3 provides a summary of the main statistics by the platform's internal rating.

The loan amounts are low, with an average of €2,531 and a maximum of €10,632. Loans are concentrated in the lower rating categories, with more than 80% of the loans falling into the C-rated or even lower one. Risk-based pricing is reflected in the average interest rate. Even for the AA grade, the interest rate is higher (11.50%) than the average bank interest rate in the Euro area, which is around 6.9% for the examined period in the retail consumer loan segment (ECB, 2020). Investors' expected returns are growing monotonously with the lowering of the rating, except for the HR rating, where the higher risk is not compensated with a higher expected yield. In comparison, the average bank deposit in the Euro area for the same period was around 0.9% for maturity over 2 years (Euro Area Statistics, 2020). Regarding the default rate (based on the above-described default definition), the portion of defaulted loans also increases with the worsening of the rating. The average default rate on the whole portfolio is 31%, which is extremely high, compared to commercial banks' retail portfolios.

3.2. Analysis of the scoring model

To investigate the performance of the platform's scoring model, we built a benchmark model using publicly available standard variables, usually included in the credit risk assessment process of a commercial bank. Then, we compared our results with the platform's estimation to find evidence of the role of potential alternative data used by the platform.

For our benchmark model, we selected 12 standard variables related to the financial position of the borrower, their previous loan history, and a few social features. The variables were selected taking into account the relevance of the data for scoring and their availability. We performed sanity checks to confirm that the dataset of each variable was complete and valid. The examined variables

Table 2
The Distribution of the Portfolio by Rating and DPD.

DPD	AA	A	B	C	D	E	F	HR
0	71%	70%	65%	59%	50%	44%	33%	27%
1–7	3%	3%	3%	3%	3%	3%	2%	1%
8–15	3%	3%	4%	4%	4%	3%	1%	1%
16–30	5%	7%	7%	7%	8%	7%	6%	3%
31–60	3%	3%	3%	4%	5%	5%	6%	3%
61–90	2%	2%	2%	2%	3%	3%	3%	2%
91–120	0%	1%	1%	1%	1%	1%	1%	1%
121–150	1%	1%	1%	1%	2%	2%	1%	1%
151–180	1%	1%	1%	1%	1%	2%	2%	1%
180+	10%	10%	13%	18%	24%	30%	43%	61%

Note. Data obtained from Bondora, as of October 2020.

Table 3
Descriptive Statistics of Bondora's Portfolio by Rating as of 2020 October.

Rating	Number of loans	Average loan amount (EUR)	Standard deviation of loan amount	Average interest	Standard deviation of interest rate	Average expected return	Number of defaulted loans	Average default rate
AA	2,686	1,390	1,452	11.50%	4.50%	9.58%	255	9%
A	5,381	1,575	1,661	13.56%	4.68%	10.46%	566	11%
B	12,986	2,003	1,969	16.15%	3.91%	10.82%	1,740	13%
C	17,332	2,481	2,304	21.81%	3.94%	12.27%	3,073	18%
D	18,079	2,761	2,305	28.51%	3.98%	13.50%	4,623	26%
E	17,624	2,879	2,298	35.14%	4.19%	14.37%	5,727	32%
F	19,701	3,367	2,259	53.17%	11.30%	17.87%	9,710	49%
HR	13,799	1,750	1,516	77.24%	50.90%	15.52%	8,113	59%
Total	107,588	2,531	2,201	36.62%	27.72%	13.93%	33,807	31%

Note. Data obtained from Bondora, as of October 2020.

are listed in Table 4. For descriptive statistics, see Table 12 in Appendix. We examined the value set and the distribution of each variable and performed a few transformations where we found them reasonable.

Even if in a predictive model the variable X has a high explanatory power in the sense that the conditional distributions P(Y|X) (where Y is the dependent variable) vary with the values of X, it may not be suitable for logit regression, for two main reasons. The values of categorical variables are usually arbitrarily chosen integers that have no specific meaning, so their weighted sum would be meaningless in the model. Because the logit function has low values for low PD and high values for high PD, or vice versa, if PD is not a monotonic function of a variable, it performs poorly in the model. In such cases, some transformation must be applied to the variable to make it fit the model. The values of the variable are grouped into 'similar' categories (or split if the distribution is continuous), which

Table 4
Description of the variables of the scoring model.

Variable	Description	Transformation
Age	Age of the borrower	No transformation was applied.
Country	Country of the borrower. A category type variable, it can take the following values: EE (Estonia), ES (Spain), FI (Finland), and SK (Slovakia).	Weight of evidence (WOE) transformation was performed. The WOE was calculated using the following equation: $WOE = ln\left(\frac{Portionofgoods}{Portionofbads}\right)$ where good refers to the portion of borrowers who paid the claims, and bad is the portion of debtors who defaulted according to our definition of a default. The estimated WOE by buckets are assigned to each observation in the portfolio table, and they are used to perform binary logistic regression instead of the original values.
IncomeTotal	The sum of the debtor's total income	The natural logarithm of the total income was taken.
ExistingLiabilities	The number of current liabilities of the debtor	Above the first five categories, the other categories were merged and considered as one category (existing liabilities above 4).
LiabilitiesTotal	The total monthly liabilities of the debtor	We created a dummy variable. We calculated the deciles and merged the first two categories and assigned a 0 for them, while the rest got 1.
DebtToIncome	The debtor's monthly loan installments divided by the monthly gross income. Expressed in percentage (%)	No transformation was applied.
NoOfPreviousLoansBeforeLoan	The number of previous loans taken before this loan was issued	Above the first five categories, the other categories were merged and considered as one category (the number of previous loans above 4).
AmountOfPreviousLoansBeforeLoan	The amount of previous loans taken before this loan was issued	We created a dummy variable. In case the borrower has zero amount of previous loans we assigned 0, otherwise 1.
PreviousRepaymentsBeforeLoan	The amount of previous loans repaid by the borrower	We created a dummy variable. In case the borrower has zero amount of previous loans we assigned 0, otherwise 1.
Employmentduration	Time spent with the current employer	No transformation was applied.
Education	The education level of the debtor. An ordinal variable, it can take the following values: 1 (primary education), 2 (basic education), 3 (vocational education), 4 (secondary education), and 5 (higher education).	No transformation was applied.
Homeownershiptype	The type of the debtor's home ownership. A categorical variable, it can take the following values: 0 (homeless) 1, (owner), 2 (living with parents), 3 (tenant, pre-furnished property), 4 (tenant, unfurnished property), 5 (council house), 6 (joint tenant), 7 (joint ownership), 8 (mortgage), 9 (owner with encumbrance), and 10 (other).	We created a dummy variable. In case the borrower is an owner we assigned 0, otherwise 1.

Note. Data obtained from Bondora's website, as of October 2020.

means that the PD is more or less constant within each category, but varies widely between categories. If an intuitive high and low risk category (as in our case all liabilities) can be identified, a dummy variable is usually created. If there seem to be more than two homogeneous categories (in our case, the borrower's country), a WOE transformation is usually applied. The average PD of each category is calculated and plugged into the inverse logit function $WOE_i = ln((1-p_i)/p_i)$, so that a linear relationship between the values of the variable and the log-odds is established, which is ideal for the logit model. Finally, if a variable has a "wild" distribution, a smooth transformation (e.g. logarithm) is usually used to handle extreme values and outliers. This is the case for the total income variable. The transformation used is detailed in Table 4.

As a next step, we estimated the GINI coefficients, which reflect the explanatory power of the variables, results are presented in Table 5.

By examining the default explanatory power of each single variable country proved to be the strongest, but income and indebtness characteristics are also impacting the default frequency. Although theoretically possible, it is not realistic to build strong models from weak variables, so we only included variables with sufficiently high explanatory power. There is no exact rule for when a variable is strong enough, but GINI above 10% is a good rule of thumb. There are two other variables close to this threshold, total liabilities and education, which are generally considered very important for retail lendings. We included these in the model and dropped the weakest three. Based on the correlation of the variables, we eliminated one pair of variables (with lower GINI) whose correlation coefficient was above 0.5. To build a scoring model, we ran different logistic regressions. We had a group of variables, measuring the indebtedness of the borrower (AmountOfPreviousLoansBeforeLoan, NoOfPreviousLoansBeforeLoan, PreviousRepaymentsBeforeLoan) and we tried each of them together and also separated along with other variables in the regression, however the GINI of the model was the same in the different versions. Therefore we selected the final model based on intutition and economic interpretation of the betas. The final variables to be used for our scoring model are the country of the borrower; the type of home ownership; total income; the number of loans taken before the loan; and the total liabilities the borrower has. The results of the regression are presented in Table 6 below.

Based on the regression results, all variables proved to be significant, all of them at the 99% significance level. The impact of income and total liabilities is in line with our previous expectations, with higher income and lower liabilities reducing PD. Interestingly, if borrowers had more credit before the current one, their creditworthiness improves. Home ownership has a value of zero if the borrower owns (wholly, jointly owned, encumbered or unencumbered) the house and a value of one if the house is not owned by the borrower (e.g. living with parents or renting). Being an owner reduces PD as expected.

Using the beta coefficients of the final model, the score for each observation was estimated. This was transformed into the probability of default with the following formula:

$$f(y) = \frac{1}{1+e^{-y}} \tag{3}$$

where

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + .. + \beta_n x_n \tag{4}$$

The average probability of default for each rating category based on our model (Benchmark PD), the PDs estimated by the platform at origination (ex-ante), and the real default rates using our default definition are presented in Fig. 1.

Based on our data, the platform underestimates the probability of default for all rating grades. Our model's in-sample estimation is a little closer to the observed default rates for the worse rating cathegories than the platform's results, however, we overestimated the probability of default for the best rating category.

We estimated the ROC curve to check the classification power of the models (presented in Fig. 2). Our benchmark model's in-sample performance resulted in a GINI of 44.10%, while the platform's classification achieved a GINI of 41.08%. The goodness of these GINI values is hard to judge. In the case of a commercial bank, a retail scoring model is expected to achieve a GINI higher than 80%, but for special, high-risk portfolios, significantly lower GINIs may also be acceptable. On the other hand, Jagtiani and Lemieux's (2018) model obtained a GINI of 38% (Area under the ROC curve 69%), even for the best-performing variable set.

The in-sample GINI we could achieve was only slightly higher than the GINI of the platform's model for the same period. Therefore,

Table 5
The GINI value of the variables used for the scoring model.

Variable	GINI
Age	0.035
country	0.387
IncomeTotal	0.153
ExistingLiabilities	0.182
LiabilitiesTotal	0.076
DebtToIncome	0.114
NoOfPreviousLoansBeforeLoan	0.214
AmountOfPreviousLoansBeforeLoan	0.186
PreviousRepaymentsBeforeLoan	0.002
Employmentduration	0.005
Education	0.075
Homeownershiptype	0.149

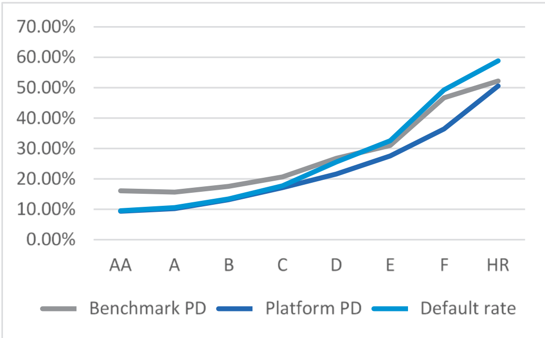
Note. Data obtained from Bondora, as of October 2020.

Table 6
Results of binary logistic regression.

Variable	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	−0.312080	0.095584	−3.265	0.00109 **
country	−0.934619	0.010481	−89.171	< 2e-16 ***
IncomeTotal	−0.085385	0.013042	−6.547	5.86e-11 ***
LiabilitiesTotal	0.166665	0.017715	9.408	< 2e-16 ***
NoOfPreviousLoansBeforeLoan	−0.086781	0.005162	−16.811	< 2e-16 ***
HomeOwnershipType	0.397527	0.014481	27.451	< 2e-16 ***

Note. Data obtained from Bondora, as of October 2020.
**p <.1, **p <.05, ***p <.01.*

Rating	Benchmark PD (ex-ante)	Platform PD (ex-ante)	Default rate (ex-post)
AA	16.02%	9.32%	9.49%
A	15.66%	10.22%	10.52%
B	17.54%	13.15%	13.40%
C	20.58%	17.10%	17.73%
D	26.75%	21.56%	25.57%
E	30.97%	27.52%	32.49%
F	46.69%	36.37%	49.29%
HR	52.16%	50.51%	58.80%



Data obtained from Bondora, as of October 2020.

Fig. 1. *Comparison of the ex-ante Model PDs and the Observed Default Rate. Note.* Data obtained from Bondora, as of October 2020.

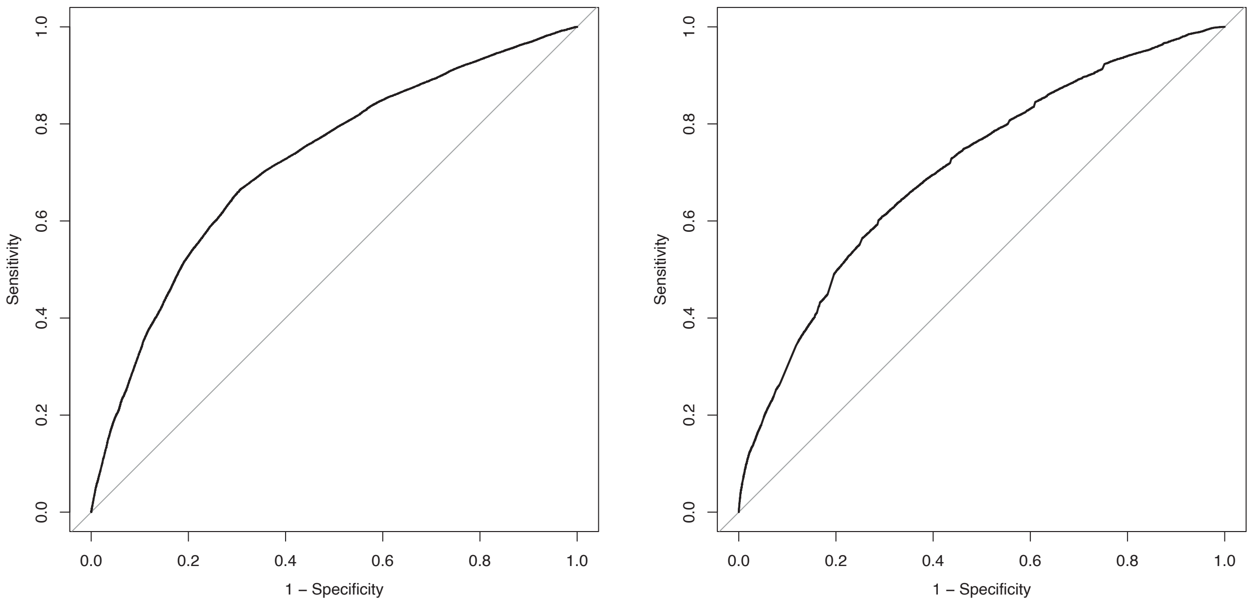


Fig. 2. *Comparison of the ROC (Receiver operating characterstic) curves in-Sample. Note.* The figure on the left shows our model's curve, and the one on the right is the curve of the platform's model. Data obtained from Bondora, as of October 2020.

we can confirm that the platform's model performs appropriately. However, the performance of our benchmark model suggests that a similar result can be achieved based on a classic "banking-like" information dataset. Hence, in our investigation, we could not detect any sign of the benefits of using alternative data sources or information processing of fintech lenders.

3.3. Out-of-sample results

We also performed out-of-sample testing to check the default prediction performance on the independent data horizon. We performed the same data cleaning steps for the portfolio table as of March 15, 2022, as mentioned for the benchmark model. As the time horizon of the benchmark model ends on October 5, 2019 (we filtered out loans issued within 12 months), the validation sample consists of 50,251 loans starting from October 6, 2019. We used the parameters (WOE binning for the country variable and coefficients) of the original model and re-estimated the PDs for the new database. The results are presented in Fig. 3, along with the platform estimation and the observed default rates.

The platform PDs underestimated the real default rates in the best and the worst rating categories, while our model overestimated them, in a manner similar to the previous period. The GINI of our model is 43.28%, only slightly below the in-sample value (44.10%). The analysis of the platform's PD resulted in a GINI of 37.92%, compared to the previous period's 41.08% The ROC curves are presented in Fig. 4. Interestingly, the drop in GINI was higher for the platform's estimation, even though their rating is given at the origination. Thus, this change is not due to the out-of-sample testing, as in the case of our benchmark model, but the changing market conditions.

As the highest explanatory value—GINI—was classified by the country, it is worth examining the loans on a country level. Fig. 5 presents the country-level default rates of the rating categories, separately for the first and the second periods. Ratings and risk assessment is adequate for Estonia, but for the other countries the default rate is not a monotonic function of rating. Loans outside Estonia (foreign origin) are much riskier, reflected in high default rates. There are no loans provided for Slovakian borrowers in the second period.

Loans out of Estonia not only have lower ratings, but also higher default rates in all rating categories (especially in the first period). This suggests that the payment discipline of borrowers of foreign origin is much lower, which can be due to the weaker cross border credit collection.

We tested the performance of our model for each country separately by calibrating a scoring model to country-level data. The predictive power of the risk models is essentially reduced at the country level; a summary of the GINIs is shown in Table 7. Our country-level models, which rely on the variables used for the benchmark model above except for the country (total liabilities, the type of home ownership; total income; the number of loans taken before the loan) performed poorly with GINIs of around 20%. The platform model performed slightly better in 5 out of 7 cases, but its discriminatory power is also very low.

The decline in the performance of the scoring model at the country level is due to the fact that the rating of borrowers differs significantly across countries. Consequently, the score of each loan is determined by the country of the borrower and the other variables have much less explanatory power. It seems that in the high-risk segment, where platform lending is active, individual defaults are much less predictable. Since the platform model could not significantly outperform our naive models, and both performed rather poorly, it seems that additional - alternative - information could help to build a better model. However, based on the data analysed, there is no evidence that platforms use and exploit alternative information.It is also important to note that the high default frequency suggests that P2P loans differ significantly from the loans acceptable for traditional financial institutions. Thus, using traditional techniques for their risk assessment is not appropriate.

To better understand the characteristics of P2P loans and the motives of P2P lending from the lenders' perspective, in the following section, we analyze the ex-post performance of platform investments.

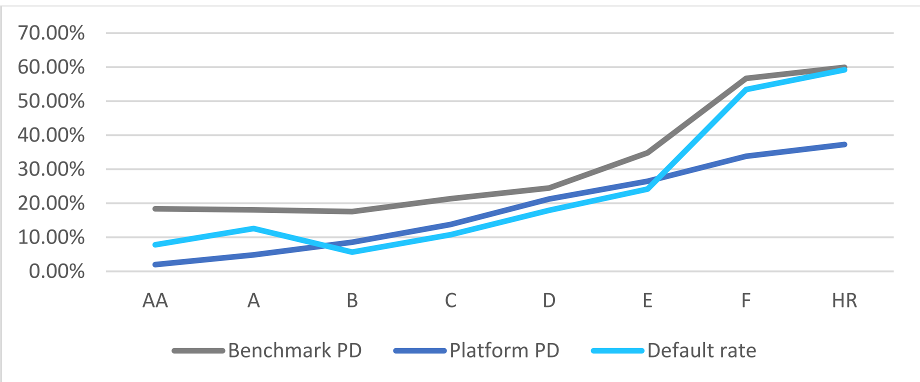


Fig. 3. *Comparison of the Out-of-Sample PDs and the Real Default Rate. Notes.* Benchmark model Data obtained from Bondora, as of March 2022.

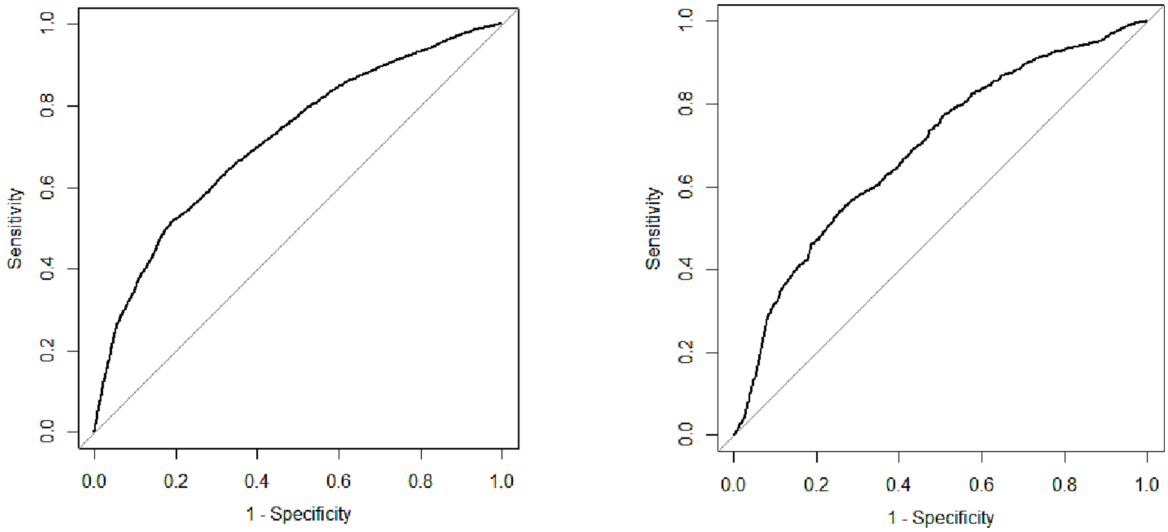


Fig. 4. *Comparison of the ROC Curves out-of Sample. Note.* The figure on the left shows our model's curve, and the one on the right is the curve of the platform's model. Data obtained from Bondora, as of March 2022.

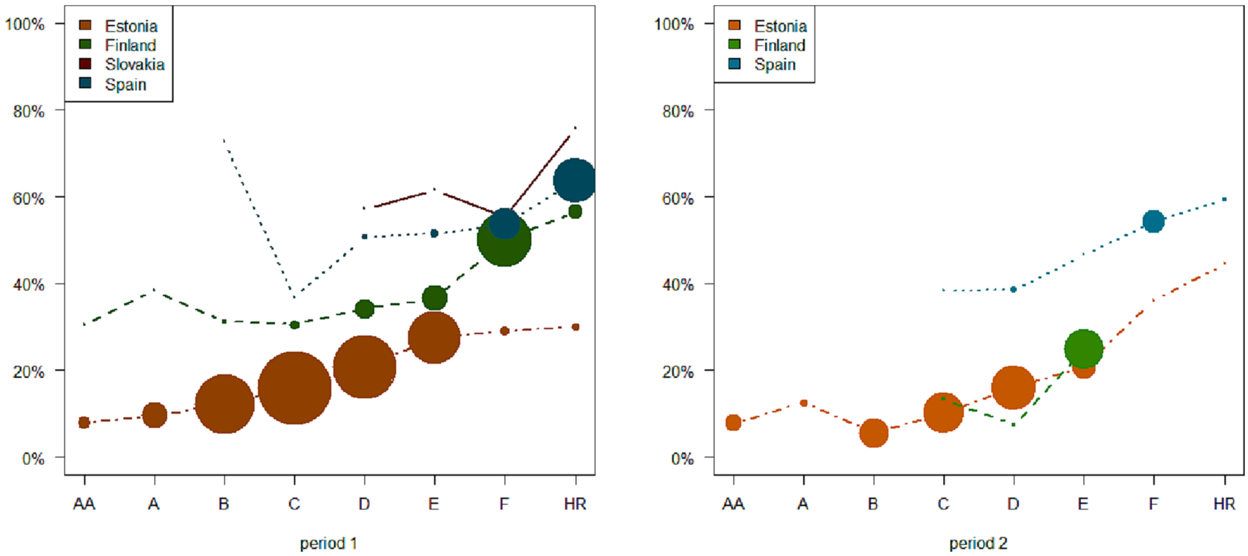


Fig. 5. *Default Rates by Country for the Two Periods. Note.* The size of the bubbles reflects the number of loans. Period 1 contains loans issued between December 2012 and October 2019, Period 2 lasts from October 2019 to March 2021. Data obtained from Bondora, as of October 2020 (period 1) and March 2022 (period 2).

Table 7
Default Rates and Performance (GINI) of the scoring models by country.

	Default rate		Number of loans		Platform model GINI		Benchmark model GINI	
	2012–2019	2019–2022	2012–2019	2019–2022	2012–2019	2019–2022	2012–2019(In-sample)	2019–2022(Out-of-sample)
Estonia	18.12%	12.54%	62,392	33,407	28.80%	25.13%	15.12%	22.35%
Finland	43.66%	22.54%	26,317	9,661	21.98%	16.87%	16.12%	17.13%
Spain	58.16%	51.34%	18,585	7,183	19.51%	14.70%	22.62%	11.99%
Slovakia	70.49%	NA	288	0	29.89%	NA	15.88%	NA

Note. Data obtained from Bondora, as of October 2020 and March 2022.

3.4. Return of Peer-to-Peer investments

P2P lending represents high-risk, bank-ineligible loans. Due to the strict regulatory requirements and reputational risks, banks are unable to provide financing for this segment, but individual or institutional investors may benefit if the risk is compensated by high interest rates.

To understand the effective performance of P2P loans, we examined the dataset of historical payments, with the principal and the interest amount paid each month by loan ID. The cash-flow table of Bondora downloaded on May 24, 2022, contains 4,343,194 rows, representing all payments during the lifetime of the platform. First, we ordered the payments according to loan ID. Consequently, we got the detailed cash-flow of 243,453 loan transactions. Then, we ordered the other details (rating, PD, status, etc.) based on lon ID from the portfolio table according to each loan.

Bondora assigns a status to each loan: current (transactions in progress), closed or unavailable (NA). For the latter category, other data is missing and therefore not suitable for analysis. As ex-post analysis is feasible only for closed transactions, so, first, we investigated the loans with a closed status. We calculated the IRR for each loan based on the cash-flow and the historical payment schedule. 62 537 transactions issued between 28 February 2009 and 14 March 2022 were analysed, excluding loans for which no payment was made other than at the time of disbursement.

Table 8. shows the ex-post performance of closed transactions according to rating categories. The rating was missing for 2,556 transactions. Thus, they are shown separately in the table. Average loan term refers to the difference between the last and the first payment, Sum CF is calculated by simply summing up all cash-flows.

The worst two categories offer really a high return with high volatility, but surprisingly, the effective loss (negative IRR) in those segments is not significantly higher than in the much better rating categories. Because of the definition of default, transactions with the defaulted flag do not necessarily represent an actual loss, and defaulting transactions that are past due for more than 90 days in the first year after origination may subsequently be recovered and all their obligations settled. Additionally, even non-defaulting transactions can result in losses if the default occurs outside the first annual period after the contract is signed. The average ex-post default rate denoted by the platform was 14.11% (see Table 8), much lower than the proportion of loans having non-performance status in both periods (see Figs. 1 and 3.). Investments with an effective loss—negative IRR—make up even less, specifically, 12.96% of the loans. On average, investors realizing negative IRR lost 23% of their initial investment. We can conclude that the majority of the defaulted closed transactions were recovered, ensuring an average IRR of 27.72%.

The picture is not that bright if a closer look is provided at transactions with a "current" status. Although we do not have the original repayment schedule of the loans, which could be indicative of non-payment and default problems, we also included in the analysis loans where the original maturity (calculated from the date of issuance and the original loan duration) is exceeded at the time of data collection or where no payment was made in the last one-year period. We assume here that these transactions can also be considered closed as there will be no further related payments. This represents 42,598 additional loans that are more problematic, 48% of which have received a default flag from the platform. In total, including closed status transactions, Bondora's total closed portfolio consists of 105,135 deals.

The distribution of actual loan term is presented on the histogram in Fig. 6. The P2P loans are mainly short-term transactions, half of them were paid back below one year, while the average original loan term is around 3.5 years. We also found that the majority of the debtors (69%) prepaid the claim before the maturity date, as there is no prepayment charge to be paid by the borrower (Bondora, 2021).

The relationship between the annual IRR of the loan and the interest rate priced initially by the platform is shown in Fig. 7. The IRR distribution is wider for riskier, higher-level interest rates, and the correlation coefficient between interest rate and IRR is low, below 0.09. Although a substantial section of the high-risk borrowers performs well, offering above 100% return for the investors, a maximal, 100% loss can be realized at all interest levels.

The ex-post performance of the extended closed category is shown in Table 9. The average IRR of the portfolio is negative, which means that investors on average not only do not receive compensation for the risk, but also make a loss on their original investment.. The average IRR is negative even in the best rating categories, with only rating C, HR, and the unrated transactions resulting in a positive IRR. The IRR dispersion is high, but overall 41.63% of all transactions have a negative IRR and the realised loss is 55% of the amount invested (for transactions with a negative IRR, the nominal amount of payments received is on average 55% less than the initial investment). So, despite an average initial expected return of 9.58% − 15.52%, the average realised return is negative in most rating categories.

The negative IRR of platform loans is particularly striking when considering that the period 2012–2022 was a period of economic recovery and boom following the 2007–2009 crisis. The non-performing loan ratio ranged between 7.48% and 1.79% (Statista.com), with a monotonic downward trend after 2015. To assess the performance of P2P lending from an investor perspective, the 10-year performance of different asset classes is presented in Table 10.

The IRR is not comparable to the ex-post returns of indices or individual assets, but it is clear that all asset classes performed well over the period and generated a positive risk premium. Thus, the negative performance of platform investments is not due to economic factors.

Breaking down the results by the country of the borrower, we found that loans for Estonian borrowers had a slightly positive average IRR of 2.93%, while the average IRR is negative for all other countries. Spain had the lowest IRR with an average of –22.57%, followed by Slovakia with –15% and Finland with –7.6%%. The distributions are shown in Fig. 8. It seems that the willingness to pay reduces with physical distance.

The IRR distribution of the analyzed loans is shown in Fig. 9. The green boxes show the distribution of the loans with a non-default

Table 8
Main characteristics and ex-post performance of closed transactions.

Rating	Number of loans	Default rate	Average loan amount	Average loan term (in days)	Average Sum CF	IRR mean	IRR St. Dev.	P (IRR < 0)
AA	2,672	5.01%	1,922	620	176	9.01%	10.54%	9.81%
A	3,166	6.16%	1,796	744	315	12.24%	12.32%	11.37%
B	8,205	6.79%	2,057	683	368	14.41%	11.67%	10.65%
C	11,246	8.94%	2,281	661	492	19.24%	14.52%	11.33%
D	11,099	13.86%	2,409	629	528	23.41%	19.46%	14.33%
E	1,0431	16.12%	2,597	557	533	25.61%	22.30%	15.31%
F	7,471	20.28%	3,025	501	728	44.57%	38.18%	15.13%
HR	5,691	31.15%	1,756	761	516	70.93%	127.33%	16.87%
NA	2,556	16.63%	642	712	166	28.36%	18.86%	2.27%
ALL	62,537	14.11%	2,261	635	484	27.72%	46.40%	12.96%

Note. Data obtained from Bondora, as of May 2022.

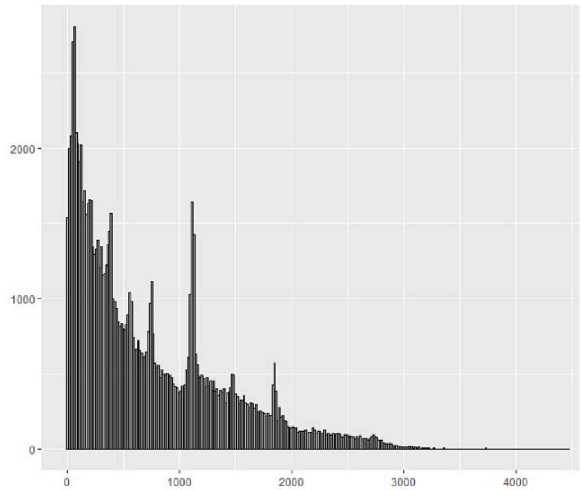


Fig. 6. *Distribution of Actual Loan Term (in Days). Note.* Based on the extended closed dataset of Bondora. Data obtained from Bondora, as of May 2022.

status, while the orange boxes are the defaulted loans. The standard deviations increase as the rating worsens, but there is no significant difference in the dispersion of the distributions within a given category. The IRR is positive for about 75% of the non-defaulted deals and negative for the majority of the loans with a default flag, except for the unrated category, in which even the loans with a default flag realized a positive IRR in 75% of the cases.

According to the analysis of the realized cash-flows and the IRR, investors do not seem to be compensated for the high risk undertaken by P2P loans. The ex-post return is negative on average, confirming the findings of Emekter et. al (2015). The negative IRR also casts doubts on the economic rationale of platform lending and raises the potential role of other motives such as altruism, as discussed in Berentsen and Markheim (2021). However, interest rates of up to 260% depending on creditworthiness do not reflect altruistic motives. It is also worth noting that, in addition to the transactions analysed, there where 4 997 loan IDs that have only one cash-flow, the initial loan disbursement. We can consider them to be credit fraud, where the borrower had no intention to pay anything back. Fraud transactions make up more than 2% of all loan IDs.

4. Conclusion

This paper investigated the possible explanations for the rapid growth in P2P platforms' market share in the credit market. From a theoretical perspective, the strongest argument for the relevance of P2P lending is the reduction in information asymmetry through the use of alternative data and P2P platforms' information processing. We analyzed the loan-level data of an Estonian platform, Bondora, to find evidence on the benefit of using alternative information based on the performance of the credit risk model of the platform. We found that the grades assigned by the platform are in line with the default risk of the borrower, but the assigned default probabilities underestimated the real default rate in all segments. Our benchmark scoring model performed slightly better than the platform's model both on the in-sample and out-of-sample data; however, we used traditional explanatory variables: age, debt-to-income, home ownership, employment duration, country of origin, and existing liabilities. When looking at loans by country of origin, the performance of the models decreases significantly, indicating that both our model and the platform's model primarily capture different

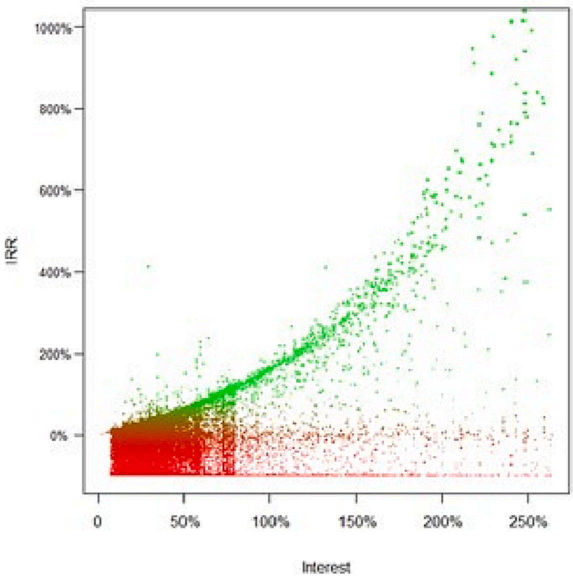


Fig. 7. *Realized IRR as a Function of the Interest Rate. Note.* Based on the extended closed dataset of Bondora. Data obtained from Bondora, as of May 2022.

Table 9
Main characteristics and ex-post performance of the extended closed dataset.

Rating	Number of loans	Default rate	Average loan amount	Average loan term (in days)	Average Sum CF	IRR mean	IRR St. Dev.	P (IRR < 0)
AA	3,701	9.92%	1,843.50	764.40	−96.31	−4.00%	29.42%	30.96%
A	4,867	11.42%	1,693.45	900.82	18.30	−3.49%	31.14%	36.47%
B	12,292	12.41%	2,042.84	864.78	25.13	−1.15%	32.08%	33.25%
C	17,116	16.68%	2,349.80	834.13	9.38	0.29%	37.58%	33.46%
D	18,544	25.04%	2,574.12	783.91	−185.74	−3.86%	46.59%	39.70%
E	17,660	29.85%	2,753.38	644.83	−340.97	−9.12%	53.03%	44.51%
F	16,374	44.99%	2,943.98	529.52	−584.43	−17.29%	68.88%	56.99%
HR	11,880	53.10%	1,760.87	726.39	−228.66	4.29%	116.83%	53.21%
NA	2,701	19.29%	643.52	767.04	148.63	24.93%	25.64%	5.96%
ALL	105,135	27.98%	2,355.22	736.35	−201.15	−4.17%	60.36%	41.63%

Note. Data obtained from Bondora, as of May 2022.

Table 10
10-years Performance of different investments.

Investment type	10 yearAnnual return	Volatility
MSCI EUR High Yield Corporate Bond Index	3.21%	7.14%
MSCI Europe Large Cap Index (equity)	7.64%	13.80%
MSCI Euro Index (equity)	8.58%	16.43%
Bitcoin (EUR)	103.08%	97.26%
All transactions of Bondora	IRR: −4.17%	60%

Note. MSCI data as of 31th March 2023, Bitcoin statistics based on daily values available at investing.com.

country-level creditworthiness. Therefore, alternative data may be needed to improve the models, but our results do not confirm that the platform can incorporate them.

By analyzing the ex-post performance of “closed” transactions, we found a significantly high average IRR of 27.72%. However, if we extend the analysis to transactions with a “current” status that are not expected to generate further cash flows, we find that the IRR is negative on average, with 41.63% of all transactions ending in a net loss (negative IRR).

There are huge differences in credit performance across countries. Foreign (non-Estonian) borrowers are not just lower rated, but their credit performance is significantly worse even in the same rating category, indicating the inefficiency of crossborder collection and higher information asymmetry. We conclude that the high credit risk reflected in the extremely high default rates is associated with net ex-post loss of the investment. Although the level of the interest rates excludes the altruistic motives of the investors, they are

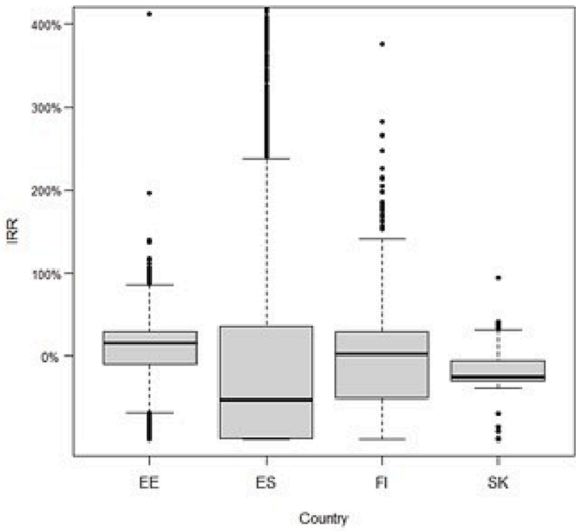


Fig. 8. *IRR Distribution According to Country. Note.* Based on the extended closed dataset of Bondora. Data obtained from Bondora, as of May 2022.

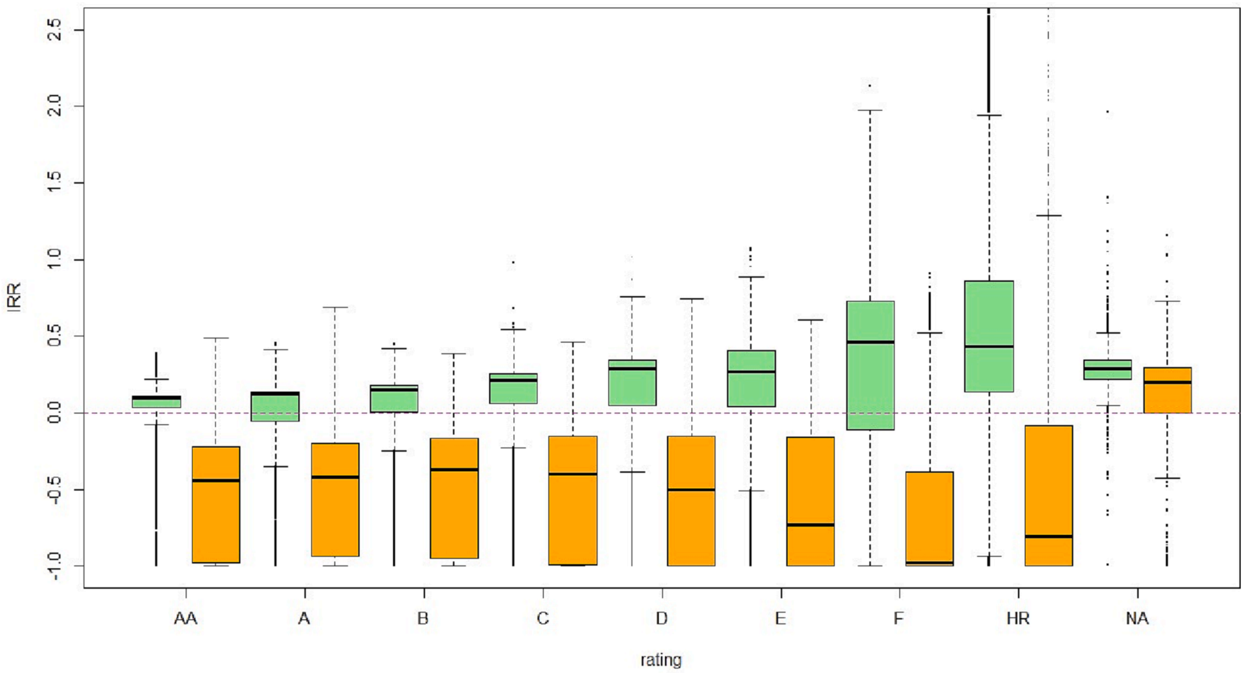


Fig. 9. *Distribution of IRR by Rating Category and Default Status. Note.* Based on the extended closed dataset of Bondora. Data obtained from Bondora, as of May 2022.

insufficient to compensate investors for the credit risk. The loan-sharking level interest rates on the other hand may lead to adverse selection on the borrower side.

Platform loans represent high-risk transactions that may not be acceptable for a traditional financial institution; thus, P2P lending complements traditional financial intermediation. However, the negative ex-post performance casts doubt on the rationale of P2P investments, even for market participants free of capital burden and reputation risk. Regulation is now better focused on the segment, but more transparency, disclosure of ex-post performance of loans and models and improvement of the crossborder collection would be needed to reduce information asymmetry and adverse selection.

CRediT authorship contribution statement

Barbara Dömötör: Conceptualization, Methodology, Validation, Formal analysis, Supervision. **Ferenc Illés:** Methodology, Software, Validation, Formal analysis, Visualization. **Tímea Ölvedi:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This research was supported by National Research, Development and Innovation Office - NKFIH, K-138826

Appendix

See Tables 11 and 12.

Table 11
List of all variables in the raw portfolio table.

Report as of EOD	Age	Home ownership type
Loan Id	Date Of Birth	Income From Principal Employer
Loan Number	Gender	Income From Pension
Listed On UTC	Country	Income From Family Allowance
Bidding Started On	Applied Amount	Income From Social Welfare
Bids Portfolio Manager	Amount	Income From LeavePay
Bids Api	Interest	Income From Child Support
Bids Manual	Loan Duration	Income Other
User Name	Monthly Payment	Income Total
New Credit Customer	County	Existing Liabilities
Loan Application Started Date	City	Liabilities Total
Loan Date	Use Of Loan	Refinance Liabilities
Contract End Date	Education	Debt To Income
First Payment Date	Marital Status	Free Cash
Maturity Date_Original	Nr Of Dependants	Monthly Payment Day
Maturity Date_Last	Employment Status	Active Schedule First Payment Reached
Application Signed Hour	Employment Duration Current Employer	Planned Principal Till Date
Application Signed Weekday	Employment Position	Planned Interest Till Date
Verification Type	Work Experience	Last Payment On
Language Code	Occupation Area	Current Debt Days Primary
Previous Repayments Before Loan	Principal Debt Servicing Cost	Next Payment Nr
Previous Early Repayments Before Loan	Interest And Penalty Debt Servicing Cost	Nr Of Scheduled Payments
Previous Early Repayments Count Before Loan	Active Late Last Payment Category	ReScheduled On
Debt Occurred On	EAD1	Rating_V0
Current Debt Days Secondary	EAD2	EL_V1
Debt Occurred On For Secondary	Principal Recovery	Rating_V1
Expected Loss	Interest Recovery	Rating_V2
Loss Given Default	Recovery Stage	Status
Expected Return	Stage Active Since	Restructured
Probability Of Default	Model Version	Active Late Category
Default Date	Rating	Worse Late Category
Principal Overdue By Schedule	EL_V0	Credit ScoreEsMicroL
Planned Principal Post Default	Grace Period Start	CreditScoreEsEquifaxRisk
Planned Interest Post Default	Grace Period End	CreditScoreFiAsiakasTietoRiskGrade
Interest And Penalty Payments Made	Next Payment Date	CreditScoreEeMini
Principal Write Offs	Principal Balance	PrincipalPaymentsMade
Interest And Penalty Write Offs	Interest And Penalty Balance	Amount Of Previous Loans Before Loan
No Of Previous Loans Before Loan		

Table 12
Descriptive statistics for the variables in the scoring model.

	Age	Income total	Debt to income	Country	Employmentduration	Education	Existing liabilities	Liabilities total	No of previous loans before loan	Amount of previous loans before loan	Previous repayments before loan	Home ownership type
Average	41	1 576	0.48	–	–	–	3	501	1	2 840	950	–
Standard deviation	12	5 178	19.83	–	–	–	4	1 046	2	4 431	1 852	–
Minimum	18	0	0	–	–	–	0	0	0	0	0	–
Maximum	75	1 012 019	4 607.82	–	–	–	40	145 042	25	44 417	34 077	–
Missing values	0	0	0	0	0	0	0	0	0	0	0	0
Type of variable	Continuous	Continuous	Continuous	Category	Category	Category	Continuous	Continuous	Continuous	Continuous	Continuous	Category

References

Bazot, G., 2018. Financial consumption and the cost of finance: Measuring financial efficiency in Europe (1950–2007). J. Eur. Econ. Assoc. 16 (1), 123–160. https://doi.org/10.1093/jeea/jvx008.
Berentsen, A., Markheim, M., 2021. Peer-to-peer lending and financial inclusion with altruistic investors. Int. Rev. Financ. 21 (4), 1407–1418. https://doi.org/10.1111/irfi.12333.
Bondora (2021). What are the prepayment risks? Available: https://support.bondora.com/en/what-are-the-prepayment-risks.
Cornaggia, J., Wolfe, B., & Yoo, W. (2018). Crowding out banks: Credit substitution by peer-to-peer lending. *Available at SSRN 3000593*. 10.2139/ssrn.3000593.
Croux, C., Jagtiani, J., Korivi, T., Vulanovic, M., 2020. Important factors determining Fintech loan default: Evidence from a lendingclub consumer platform. J. Econ. Behav. Organ. 173, 270–296. https://doi.org/10.1016/j.jebo.2020.03.016.
Cumming, D.J., Leboeuf, G., Schwienbacher, A., 2020. Crowdfunding models: Keep-it-all vs. all-or-nothing. Financ. Manag. 49 (2), 331–360. https://doi.org/10.2139/ssrn.2447567.
Das, S.R., 2019. The future of fintech. Financ. Manag. 48 (4), 981–1007. https://doi.org/10.1111/fima.12297.
Davis, K., 2016. Peer-to-peer lending: structures, risks and regulation. JASSA 3, 37–44. https://search.informit.org/doi/10.3316/informit.419817919644101.
De Roure, C., Pelizzon, L., & Tasca, P. (2016). How does P2P lending fit into the consumer credit market? *Bundesbank Discussion Paper* No. 30/2016. 10.2139/ssrn.2756191.
Diamond, D.W., 1984. Financial intermediation and delegated monitoring. Rev. Econ. Stud. 51 (3), 393–414. https://doi.org/10.2307/2297430.
Duarte, J., Siegel, S., Young, L., 2012. Trust and credit: The role of appearance in peer-to-peer lending. Rev. Financ. Stud. 25 (8), 2455–2484. https://doi.org/10.1093/rfs/hhs071.
Emekter, R., Tu, Y., Jirasakuldech, B., Lu, M., 2015. Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. Appl. Econ. 47 (1), 54–70. https://doi.org/10.1080/00036846.2014.962222.
Euro Area Statistics (2020). Bank interest rates - Deposit. Downloaded: 2020.12.02. Available: https://www.euro-area-statistics.org/bank-interest-rates-deposits?cr=eur&lg=en&page=0&charts=M..B.L22.L.R.A.2250.EUR.N+M..B.L22.F.R.A.2250.EUR.N+M..B.L22.K.R.A.2250.EUR.N&template=1.
European Central Bank (2020). Statistics. Downloaded: 2020.12.02. Available: https://www.ecb.europa.eu/stats/html/index.en.html.
Feyen, E., Frost, J., Gambacorta, L., Natarajan, H., Saal, M., 2021. Fintech and the digital transformation of financial services: implications for market structure and public policy. BIS Papers.
Freixas, X., Rochet, J.C., 2008. Microeconomics of banking. MIT press.
Gavurova, B., Dujcak, M., Kovac, V., Kotásková, A., 2018. Determinants of successful loan application at peer-to-peer lending market. Economics & Sociology 11 (1), 85–99. https://doi.org/10.14254/2071-789X.2018/11-1/6.
Goldstein, I., Jiang, W., Karolyi, G.A., 2019. To FinTech and beyond. Rev. Financ. Stud. 32 (5), 1647–1661. https://doi.org/10.1093/rfs/hhz025.
Gosztonyi, M., Havran, D., 2021. Highways to Hell? Paths Towards the Formal Financial Exclusion: Empirical Lessons of the Households from Northern Hungary. Eur. J. Dev. Res. 1–34 https://doi.org/10.1057/s41287-021-00434-9.
Havrylchyk, O., Verdier, M., 2018. The financial intermediation role of the P2P lending platforms. Comp. Econ. Stud. 60 (1), 115–130. https://doi.org/10.1057/s41294-017-0045-1.
Herzenstein, M., Andrews, R.L., Dholakia, U.M., Lyandres, E., 2008. The democratization of personal consumer loans? Determinants of success in online peer-to-peer lending communities. Boston University School of Management Research Paper 14 (6), 1–36.
Hughes, J.P., Jagtiani, J., Moon, C.G., 2022. Consumer lending efficiency: commercial banks versus a fintech lender. Financial Innovation 8 (1), 1–39. https://doi.org/10.1186/s40854-021-00326-1.
Jagtiani, J., Lemieux, C., 2018. Do fintech lenders penetrate areas that are underserved by traditional banks? J. Econ. Bus. 100, 43–54. https://doi.org/10.21799/frbp.wp.2018.13.
Jagtiani, J., Lemieux, C., 2019. The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. Financ. Manag. 48 (4), 1009–1029. https://doi.org/10.21799/frbp.wp.2018.15.
Lin, M., Prabhala, N.R., Viswanathan, S., 2013. Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. Manag. Sci. 59 (1), 17–35. https://doi.org/10.1287/mnsc.1120.1560.
Liu, Z., Shang, J., Wu, S.Y., Chen, P.Y., 2020. Social collateral, soft information and online peer-to-peer lending: A theoretical model. Eur. J. Oper. Res. 281 (2), 428–438. https://doi.org/10.1016/j.ejor.2019.08.038.
Maskara, P.K., Kuvvet, E., Chen, G., 2021. The role of P2P platforms in enhancing financial inclusion in the United States: An analysis of peer-to-peer lending across the rural–urban divide. Financ. Manag. 50 (3), 747–774. https://doi.org/10.1111/fima.12341.
Merton, R.C., Thakor, R.T., 2019. Customers and investors: a framework for understanding the evolution of financial institutions. J. Financ. Intermed. 39, 4–18. https://doi.org/10.1016/j.jfi.2018.06.002.
Milne, A., Parboteeah, P., 2016. The business models and economics of peer-to-peer lending. European Credit Research Institute. https://doi.org/10.2139/ssrn.2763682.
Morse, A., 2015. Peer-to-peer crowdfunding: Information and the potential for disruption in consumer lending. Annu. Rev. Financ. Econ. 7, 463–482. https://doi.org/10.3386/w20899.
Najaf, K., Subramaniam, R.K., Atayah, O.F., 2022. Understanding the implications of FinTech Peer-to-Peer (P2P) lending during the COVID-19 pandemic. Journal of Sustainable Finance & Investment 12 (1), 87–102. https://doi.org/10.1080/20430795.2021.1917225.
Philippon, T. (2016). *The fintech opportunity* (No. w22476). National Bureau of Economic Research. 10.3386/w22476.
Polasik, M., Huterska, A., Iftikhar, R., Mikula, Š., 2020. The impact of Payment Services Directive 2 on the PayTech sector development in Europe. J. Econ. Behav. Organ. 178, 385–401. https://doi.org/10.1016/j.jebo.2020.07.010.
Popescu, G. (2016). Alternative data usage by FinTech lenders. Available: https://cointimes.info/2016/06/20/over-12000-variables-of-alternative-data-for-36-billion-in-alternative-loans/.
Tang, H., 2019. Peer-to-peer lenders versus banks: substitutes or complements? Rev. Financ. Stud. 32 (5), 1900–1938. https://doi.org/10.1093/rfs/hhy137.
Thakor, A.V., 2020. Fintech and banking: what do we know? J. Financ. Intermed. 41, 100833 https://doi.org/10.1016/j.jfi.2019.100833.
Van Liebergen, B., 2017. Machine learning: a revolution in risk management and compliance? J. Finan. Transf. 45, 60–67.

Cited websites

https://www.bondora.com/en last download 8 June, 2023.
p2pmarketdata.com last download 15 May, 2023.
Statista.com last download 10 April, 2023.