Factors affecting the interest rate in peer-to-peer lending in Bulgaria

Author: Viktor Draganov
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT,

In this paper we investigate borrower characteristics, such as expense income ratio, age, gender, and education and their effect on interest rate in the context of Peer-to-peer lending in Bulgaria. This study was conducted by analyzing publicly available information provided by the biggest P2P lending platform in Bulgaria, Klear. We found that higher education has a significant negative impact on interest rate. Conversely, female and older borrowers receive higher interest rates. Similarly, a high expense-to-income ratio led to higher interest rates. The results of the study are in line with the signaling theory reinforcing its reliability with regards to reducing information asymmetry, which continues to be one of the greatest obstacles in alternative finance. This study emphasizes that it is imperative to include more borrower characteristics in the risk evaluation of a given loan, as lenders would be able to make more informed lending decisions. Finally, P2P lending platforms should ensure higher integrity of income verification tools as this would enable more fluent borrower-lender interactions, potentially reducing perceived risk.

Graduation Committee members:
1st Supervisor: Xiaohong Huang
2nd Supervisor: Lingbo Shen

Keywords

Peer-to-peer lending; Interest rate; Lender; Borrower; Bulgaria; Factors; Alternative finance; Signaling theory; Information asymmetry

1. INTRODUCTION

Peer-to-peer (P2P) lending is a form of alternative finance that revolves around the direct lending of money through online platforms. The platforms inquire fees that are a percentage of the deal made between the lender and the borrower (Julia Kagan, 2024). The online platform eliminates the need for institutional intervention, such as banks. It is a simplistic process, where the platform receives an application by a borrower, after which it makes an assessment of the application and determines an appropriate interest rate, according to the credit rating the borrower received. After that the borrower is obligated to pay monthly installments and repay the principal amount upon maturity (see Figure 1.).
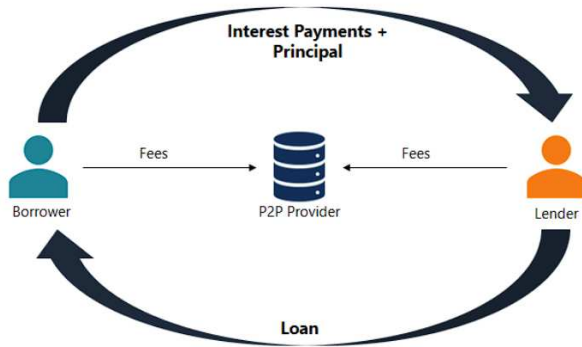

Figure 1. Peer-to-peer lending

P2P lending is a tool most commonly used by individual borrowers, SMEs, and individual investors for short and long term debt financing. P2P lending is attractive for both borrowers and lenders. It provides easily accessible capital to SMEs and individual borrowers, compared to traditional bank loans (Rainer Lenz, 2016). Additionally, it provides higher return to lenders, consequently creating demand and supply (Tomi Hietala et al. 2016). Since 2015, the total market volume of alternative finance has seen a slow but steady rise, 44$ billion in 2015 and reaching 113$ billion in 2020 (Tania Ziegler et al. 2021), excluding China. Alternative finance has also been on the rise in Bulgaria, based on statistics provided by the most dominant online platform operating there: *Klear,* going from approximately 2.75BGN (1.40€) million (Bulgarian currency) in 2017 to almost 11BGN (5,61€) million in 2024.

The importance of this paper escalates, in terms of academic relevance when specifically targeting the Bulgarian financial market. Industry reports examine the global development of P2P lending (Tania Ziegler et al. 2021), while existing literature focuses on various variables (e.g., Credit Grade, Household income, borrower's income, indebtedness ratios, etc.) relevant to different aspects of the P2P lending process (Tomi Hietala, 2016; Ming Yang et al., 2017; Duarte, J. et al. 2012; Serrano-Cinca, C. 2015). However, close to no research has been done on the Bulgarian P2P lending market. This paper contributes to the existing knowledge gap by examining the relationship between borrower's characteristics and their influence on interest rates.

Identification and analysis of those factors can have significant implications on several entities in the Bulgarian financial market, such as policy makers, borrowers and lenders. Understanding these determinants can help investors make well-informed decisions, benefit borrowers in reaching favorable terms, and assist policymakers in creating effective regulations with regards to not only P2P lending but the alternative finance sector as a whole. Another practical relevance of this research is improved forecasting of market trends in alternative finance.

1.1 Research objective

The goal of this research is to investigate the strength of borrower's characteristics in relation to interest rate in P2P lending in Bulgaria. Before discussing the factors, we note that this review focuses solely on loans with fixed interest rates. This narrower scope enhances the paper's practicality and allows for direct comparative analysis with other research. Only loans with fixed interest rates will be reviewed in this paper, Additional reason for that is the public dataset taken from the online P2P lending platform *Klear,* which includes loans with fixed interest rate only. *Klear* is the biggest company in the P2P lending sector in Bulgaria and is the only one that has gathered sufficient data to study the relationship between the different factors. The selection of these variables was based on a comprehensive review of the literature, which scarcely investigates the relationship between interest rate and borrower characteristics (Andreas Dietricha, 2019; Herzenstein et al. 2008; Qi Liu et al., 2017), hence lacking more precise and concrete findings. The factors that are going to be studied in this paper include: *age, education, verifiable net income, total expenses* and *gender.* These factors would allow for a sophisticated analysis of the signaling theory in the Bulgarian P2P lending market and whether the different assumptions about their effects prove to be significant. Additionally, the results would shed light on the perspective of lenders with regard to borrower's characteristics.

1.2 Research problem

Having clarified the different variables to be investigated in this paper we can deduct the following research question: *How strong is the relationship between borrower's characteristic and the interest rate applied on the loan in Peer-to-peer lending in Bulgaria?*

2. LITERATURE REVIEW AND RESEARCH FRAMEWORK

2.1. Peer-to-peer lending

The relevant literature regarding peer-to-peer lending proportionally expands in response to the growth of the alternative financing method. Hulem and Wright (2006) researched the first online P2P lending platform Zopa, originating in the UK. They suggest that the emergence of social lending is a response to the growing demand for financial services that combine authentic value-driven relationships with higher returns and savings, a business model unattainable by traditional financial institutions such as banks due to various regulatory constraints regarding data privacy and safeguards rules. Wang et al. (2009) argue that P2P lending marketplaces can revolutionize the intrinsic operations of a loan because of the ability to provide both middlemen functionality and a marketplace to take place, potentially leading to a more efficient market. Additionally, Kaleemullah Abbasi et al. (2021) find that P2P lending marketplaces enhance the availability of loans to SMEs who may otherwise struggle to gather capital. Studies predominantly focus on predictive factors influencing loan default rate and interest rate (e.g., Serrano-Cinca et al. 2015; Tomi Hietala et al. 2016; J. Huang et al. 2010; Duarte, J., et al. 2012; Tao et al. 2017). Alexander Bachmann et al. (2011) and Qizhi TAO (2017) both classify these factors in two broad categories: hard information (e.g., borrowers' credit scores) and soft information (e.g., age, race, gender, etc.). As per Petersen's (2004) description soft information is hard to summarize in a

numeric score, dependent on context, and is inherently difficult to distinguish between the reason for collection and its application due to judgements and subjective interpretations.

2.3 Information asymmetry problem

In the finance literature, information asymmetry revolves around a discrepancy between stakeholders, in this case borrowers and lenders. Naturally, the borrowers possess more information about their borrowing capabilities than the investors, which may result in adverse selection (Emekter et al., 2014). In the traditional landscape of lending, banks act as a professional evaluator of borrowers creditworthiness, since they have the financial knowledge and expertise to properly determine an appropriate capital distribution. In P2P lending, however, by highlighting the imbalance between the two stakeholders, challenges regarding the accurate assessment of a borrower's creditworthiness are revealed (Mohammed m Elgammal et al., 2021) precisely because of the removal of institutional intermediation. This may lead to adverse selection, consequently leading to lenders charging higher interest rates to compensate for the additionally incurred risk. Overall, in the P2P lending literature information asymmetry is a widely discussed issue and is seen as one of the biggest downsides to that alternative finance method (Tao, et al., 2017; Bachmann, 2011; Emekter et al., 2014). It is worth mentioning that every piece of information provided to the lender reduces the information gap between the two stakeholders, including the borrower's soft information.

2.4 Dual-process model for lending decisions

In the 1980s Shelly Chaiken developed the dual-process theory of information processing and persuasion HSM (i.e., heuristic-systematic model). It posits two distinct ways of processing information with the goal of forming an opinion or making a decision: heuristic processing and systematic processing. Heuristic processing is characterized by simplistic interpretations of given information (Zhang, K. et al. 2014) and mental shortcuts (heuristics). Chaiken argues that cognitive overload, low motivation or time constraints are likely factors for not engaging in more thorough evaluation of information, hence employing the heuristic processing method. On the other hand, the systematic processing involves comprehensive, analytic, and detailed observation of information, hence demanding a heavy cognitive effort. It is more likely to occur when the lender is highly motivated to grasp the information and is cognitively prepared and able to do so (Chen, S., & Chaiken, S. 1999). The two distinct ways of HSM can be used simultaneously, where heuristics can provide initial judgements, which are then refined through systematic processing. However, in the heuristic processing biases may occur, potentially skewing final evaluations. In the context of this study the systematic process occurs when examining hard information (e.g., borrower's credit scores), while heuristic processing occurs when examining the soft information of a borrower (race, age, gender, education). The importance of HSM as a theoretical foundation in this study stems from its direct connection to signaling theory since understanding the fundamental determinants of how lenders make decisions can lead to better interpretation of signals as a consequence.

2.5 Signaling theory

Signaling theory addresses the problem of information asymmetry described earlier by allowing borrowers to convey credible information in the form of signals to lenders, in order to bridge the gap between the asymmetric information and as a result foster decision-making capabilities and reduce uncertainty. Signaling theory in pair with the fundamental

assumption of heuristic processing of investors we argue for the effects of the different variables explored in this study.

2.5.1 Education

Education can serve as an identifier for potential, financial literacy, higher levels of income, lower default risk, and more stable financial income. Additionally, understanding loan terms as well as managing financial resources can be more comprehensively achieved by persons with higher education. Overall, higher education serves as a positive signal to lenders, which may lead to lower interest rates (Junhui Xu et al., 2018). Hence, combining literature and signaling theory the following hypothesis can be deduced:

Null H1: Higher degree of *Education* does not have a negative effect on interest rate.

H1: Higher degree of *Education* has a negative effect on interest rate.

2.5.2 Verifiable net income and total expenses

Verifiable net income can be used as an identifier of ability to repay loan at maturity by the borrower, hence affecting the creditworthiness rating of the borrower. If a person has a higher *verifiable net income* he may be assigned a lower interest rate as the risk that the investor takes on is less. Inversely, a high interest rate serves as compensation in the case of default, which becomes more likely if the borrower has low net income. Carlos Serrano-Cinca concluded that households with higher annual income received a higher credit score grading, affecting the interest rate negatively. By considering the literature and signaling theory, the following hypothesis can be deduced:

Null H2: Higher *verifiable net income* does not lead to lower interest rate

H2: Higher *verifiable net income* leads to lower interest rate

Proper management of *total expenses* in proportion to net income may signal to a lender well-developed financial and management skills, hence the ability of the borrower to handle debt. Conversely, high *total expenses* can increase the borrower's risk of default, hence being perceived as more risky and compared to other borrowers with less *total expenses*. There is little to no literature that explores the relationship between *total expenses* of a household and *interest rate*. However, implications from different studies such as the Martin Flodén (2021), can help us suggest that higher *total expenses* may signal financial stability and borrowing behavior, potentially leading to higher *interest rates*, due to the increased perceived risk by lenders. Considering these observations and the signaling theory discussed earlier we can deduce the following null and alternative hypothesis:

Null H3: *Total expenses* does not have a significant effect on interest rate

H3: *Total expenses* has a positive relationship with interest rate

2.5.3 Age and gender

Lower likelihood of discrimination based on race and gender is present on P2P lending platforms (Herzenstein et al. 2008) compared to traditional financial institutions, mainly because of the direct connection between borrowers and lenders, which results in more fair interactions. Several different findings regarding funding success have been established. A research in the Chinese lending market depicts a higher likelihood for a woman to receive a loan, but with a higher interest rate, although a lower default risk (Chen et al. 2017). Studies by Kuwabara and Thébaud (2017) and Li et al. (2020) have shown that female entrepreneurs who apply for business loans are

more likely to be turned down, especially if they are viewed as attractive. Additionally, studies reveal that older and female borrowers are more likely to fail on their loans, however Tao et al. (2017) found that age and gender had no effect on interest rates. On the other hand, research by Duarte et al. (2012) shows that borrowers who demonstrate their reliability have a higher chance of obtaining finance. Conversely, Ravina (2019) finds that older borrowers pay slightly higher interest rates.

Gender allows us to study whether there are systematic differences in interest rates between male and female borrowers. Various social, cultural and economic factors allow us to hypothesize about the impact of gender on interest rate. Furthermore, gender will provide an insight on the Bulgarian P2P lending marketplace and an overall grasp on the egalitarian development of the country.

Borrower's age can be another significant factor concerning interest rate and assessing creditworthiness. Variations in age lead to different circumstantial living situations, for example a younger borrower may signal that he has less experience with debt, hence receive a higher interest rate. On the other hand, older borrowers may appear more creditworthy since age strongly correlates with net income and as previously mentioned experience.

The reviewed literature in coordination with signaling theory lead us to the following hypothesis about the effect of age and gender:

Null H4: Age and gender have no significant effect on the interest rate in P2P lending in Bulgaria.

H4: Age and gender have a positive effect on the interest rate in P2P lending in Bulgaria.

2.6 Conceptual framework

In order to visualize the theoretical foundation of this study the following conceptual framework has been designed:
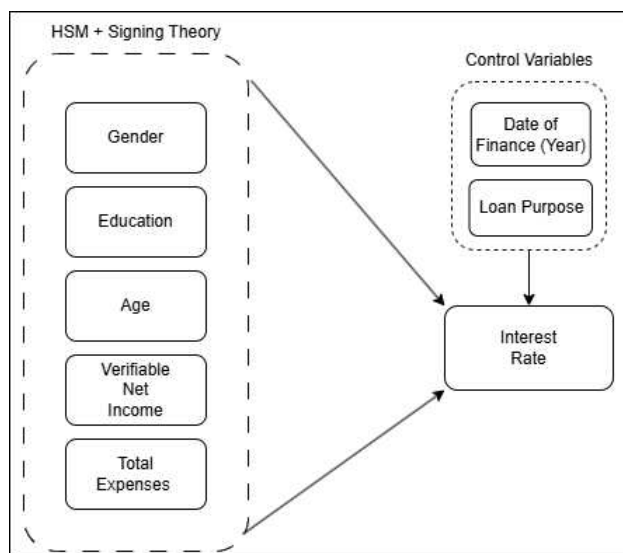

Figure 2. Conceptual framework

3. METHODOLOGY

This paper adopts quantitative research design based on a singular dataset. As previously mentioned, to investigate the research question, data from the most dominant platform on the Bulgarian P2P lending market, *Klear,* will be analyzed. They offer fixed interest rate loans, including personal loans, debt consolidation loans, car loans, home renovation loans, business loans and others. The company targets individual borrowers in

Bulgaria who need personal financing for various purposes. The application process has six stages: online application, credit assessment, loan offer, acceptance and funding, disbursement, repayment and investor returns. It offers loan sizes ranging from 100BGN (50€) to larger sums depending on the creditworthiness of the borrower. *Klear* provides a complete dataset with all the loans it has offered from 2016 up until 2024, containing all the information about the researched variables. The reason behind selecting *Klear* as a data source is because of the three online platforms operating in Bulgaria, it is the largest one, and more importantly is the only one that provides a public dataset to work with. *Klear* has funded a total of 41BGN million (approx. 21€ million; 22.5$ million) since 2016. In order to investigate the relationship between the dependent variable, *interest rate,* and the independent variables: *gender, age, education, verifiable net income* and *total expenses,* a regression analysis will be conducted. The regression analysis will then be interpreted in the *results* section of this paper. The following multiple linear regression equation can be deduced:

Interest rate = $\beta 0$ (Intercept) + $\beta 1$ (Age) + $\beta 2$ (education) + $\beta 3$ (verifiable net income) + $\beta 4$ (Total expenses) + $\beta 5$ (Gender) + $\beta 6$ (Loan purpose) + $\beta 7$ (Year of finance) + ε (error term)

Where "ε" represents the error term, capturing the unexplained variability in the dependent variable.

	Definition	Measure
Interest rate	Cost of borrowing money	Percentage (%) of loan amount
Gender	Either of the two sexes	Female / Male
Age	The length of time that a person has lived or a thing has existed	Chronological age
Education	The process of receiving or giving systematic instruction, especially at a school or university	One of the following: College; Primary education; Basic education; Vocational high school; Secondary education; University
Verifiable net Income	The income that an individual earns after taxes and other deductions	Bulgarian Lev (BGN)
Total expenses	The costs incurred by an individual or household to maintain a certain standard of living	Bulgarian Lev (BGN)
Date of finance	The moment at which the loan was funded	Year
Loan purpose	The borrower's	One of the

	reason for the loan	following: Other; Loan Refinancing / Pooling; Purchase of property; Medical needs; Vacation; Business; Car purchase; Unexpected expenses

Table 1. Definitions and measures of variables

3.1 Control variables

The literature on P2P lending depicts the date of finance (year) and loan purpose as significant factors affecting the process of lending (Serrano-Cinca, C., et. al, 2015; Matt N. Williams et al., 2013), hence they will be used as control variables in order to receive more accurate results. Date of finance covers loans from 2016 to 2024, mentioning the day, month and year when the loan was made. By analyzing the control variable in depth a depiction of external forces, such as economic conditions, regulatory changes, market dynamics, lender preferences, etc., influencing the interest rate can be observed and accounted for through the date of financing (Matt N. Williams et al., 2013), which leads to greater market predictability and easier identification of future trends that may affect the interest rate and as a result the cost of borrowing money.

In his study on "Determinants of Default in P2P Lending" Carlos Serrano-Cinca (2015) directly connects loan purpose with perceived risk, since taking a loan for a business venture is considered riskier than taking a loan for the purchase of a car. In his study he concluded that loan purpose was statistically significant when accounting for loan defaults, subsequently a case for its effect on the interest rate can be made. Since loan purpose is not a borrower's characteristic yet has an effect on the interest rate we will use it as a control variable. In the statistical model loan purpose will have to be transformed to a numerical variable with the following subsequent variables: other, loan refinancing / pooling, purchase of property, medical needs, vacation, business, car purchase and unexpected expenses.

3.2 Sample

The dataset contains 3517 anonymous loan entries taken from the Bulgarian population. Albeit the variables researched are numerical, the dataset also contains categorical data for each of the entries. One limitation of the data is that it may contain information on international borrowers and investors as well, however, being based in Bulgaria the mass of the users are Bulgarian. The public dataset that *Klear* has provided does not need additional cleaning and formatting, there are no missing values, where the variables investigated in the paper are concerned.

3.3 Data analysis

The data will be imported in the statistical programming tool RStudio to perform the following tests and analysis. The following sections are ordered in the same manner that the tests were performed. We test for these assumptions to avoid Type I and Type II errors as well as over- or under- estimation of significance (Jason W. Osborne,. 2019).

3.3.1 Multicollinearity

A multicollinearity check between the independent variables will be made, in order to ensure accurate results and help make inference on a large sample data (Jamal I. Daoud 2017). Multicollinearity is a statistical phenomenon, where two or more independent variables from a multiple regression analysis are highly correlated (Jamal I. Daoud 2017). Several methods can be employed to detect multicollinearity, such as correlation matrix, variance inflation factor, tolerance, etc. (Theodoros Kyriazos1, Mary Poga,. 2023). For most accurate results two of the above mentioned tests will be performed: correlation matrix and variance inflation factor. If multicollinearity appears as a problem several solutions can be instated: removing one or more of the variables that cause the phenomenon; principal component analysis (Mostafa Khalil., 2023); etc. If high collinearity of given variables appears to be present, then taking their ratio would be a permanent solution to the problem. After performing the regression an analysis of the results will be performed, in order to examine, interpret and make sense of them so that we draw relevant conclusions.

3.3.2 Linearity

The assumption of linearity checks for an existing linear relationship between the dependent variable *interest rate* and each of the independent variables (Zach Bobbit., 2020). The employed way in this research, for checking this assumption is mapping the variables on a scatter plot and visually observing linearity. The assumption is violated if a non-linear or a curvilinearity relationship is observed. Two solutions are present for each of the violations. A nonlinearity transformation (Zach Bobbit,. 2020), which includes taking the log, the square root, or the reciprocal of either variable can be used. In extreme cases we can use a nonlinear statistical model like spline.

3.3.4 Homoscedasticity

Homoscedasticity deals with the variance of the residuals. Specifically, it means that their variance is constant for every level of the dependent variable (Jason W. Osborne,. 2019). If this assumption is not met then heteroscedasticity is present, which means that the results become inaccurate and may result in declaring a term statistically significant. Plotting the standardized residuals versus the regression standardized predicted value will allow for visual observation of whether heteroscedasticity is present. A solution to this problem would be to take the log of the problem variables, which will stabilize the variance and make the relationship between them more linear.

3.3.5 Normality

Using a Q-Q plot we can visually check whether the residuals are normally distributed. Additionally, the use of the Shapiro-Wilk test will be used to either confirm or reject the observations from the Q-Q plot. Again a possible solution is transformation of either the independent variables or the dependent variable, like taking the log, the square root, or the reciprocal.

4. RESULTS

4.1 Descriptive analysis

In table 2. descriptive statistics of the numerical independent variables and the dependent variable can be observed. The mean and the median for *verifiable net income* are somewhat close, which could suggest a relatively symmetrical distribution, however, the high standard deviation indicates variability in net incomes. The mean of *total expenses* is slightly higher than the

median, suggesting a possible right skew of the distribution. Similarly to *verifiable net income* the standard deviation indicates moderate variability in *total expenses*. As for *age* we can deduct that there is a symmetrical age distribution as the mean and median are relatively close. However, the mode being lower depicts a clustering of younger borrowers. The standard deviation shows a wide variability in ages. Finally, the dependent variable *interest rate* has a higher mean than median indicating a right skew in the distribution. The low mode signals that most of the *interest rates* are on the lower end. Moderate variability can be observed as the standard deviation is high. The distribution of *date of finance* (*year*) and *loan purpose* can be observed in appendix 9.4.

Independent variables	Mean	Median	Min	Max	SD
Verifiable net income	2940	2267	0	28869	2.545
Living expenses	2067	1760	400.36	14767	1.266
Dependent variable	Mean	Median	Min	Max	SD
Interest rate (%)	7.44	6.59	3.49	15.69	2.61

Table 2. Preliminary descriptive statistics (excluding *education* and *gender*)

4.1.2 Description of dummy variables

The three categorical variables *education, gender,* and *year* were preprocessed, creating six new variables for *education,* two new variables for *gender,* and eight new variables for *year.* The newly created variables for *education* are the following: College, Primary education, Basic education, Vocational high school, Secondary education and University. The new variable University is set as the reference as it is most recurring. *Gender* was divided into Male and Female variables, with Male being the reference. The control variable *loan purpose* was also preprocessed, consequently creating eleven new variables, with Home renovation the most recurring, hence being the reference. Additionally, *Date of finance (year)* was dummified into eight variables depicting the *year* when the loan was financed and 2016 as the reference category. The dispersion of *education, gender,* the control variable *loan purpose,* and *year* with regard to the newly created dummy variables can be observed in table 3.

Gender	Frequency	Percentage
Male	2347	67%
Female	1170	33%
Education	Frequency	Percentage
College	125	4%

Primary education	2	0%
Basic education	15	0%
Vocational high school	215	6%
Secondary education	881	25%
University	2279	65%
Loan Purpose	Frequency	Percentage
Unpredictable expenses	257	7%
Car purchase	514	15%
Home renovation	890	26%
Other	317	9%
Loan Pooling / Refinancing	555	16%
Combination loans	403	12%
Education	87	3%
Purchase of property	215	6%
Medical needs	111	3%
Vacation	88	3%
Business	38	1%
Year	Frequency	Percentage
2016	77	2.19%
2017	341	9.70%
2018	371	10.55%
2019	472	13.42%
2020	378	10.75%
2021	469	13.34%
2022	659	18.74%
2023	597	16.97%
2024	153	4.35%

Table 3. Frequency of dummy variables

As far as the independent variables are concerned (*Gender* and *Education*) the analysis reveals that the sample is predominantly male and highly educated, with most borrowers having University or Secondary education, leading to a skewed distribution towards the more educated demographic. Additionally, the males are overrepresented as compared to females, which may introduce potential bias when generalizing findings.

The frequency distribution of *loan purpose* reveals that fifty percent of the loans taken out were either for home renovation, car purchase and other necessities. Notable proportion of twenty eight percent is taken up by Loan Pooling / Refinancing and Combination loans.

The distribution of *year* reveals the steady development of the platform with an increasing amount of loans given out every year. The only downturn can be observed between 2022 and

2023, where *Klear* gave out less loans compared to the previous year.

4.2 Correlation

A correlation matrix can be observed in Appendix 9.5, showing the correlation between each of the numerical independent variables and the dependent variable. With correlation coefficients of -0.28 and -0.26, higher *verifiable net income* and *total expenses* are linked to lower interest rates, respectively. As far as *age* is concerned it has a high correlation with *total expenses* (r = 0.152), indicating that older individuals tend to have higher *total expenses*. *Verifiable net income* has a high correlation with *total expenses* (r = 0.81).

In order to fix the high correlation between *verifiable net income* and *total expenses* a new variable *Expense Income Ratio* was created, which took the ratio: *living expenses / verifiable net income.* Consequently a new correlation matrix was created (appendix 8.6), where the correlation coefficients of the variables are stable.

As the previously discussed hypothesis regarding the two variables *verifiable net income* and *living expenses* cannot be tested due to the transformation of the variables into one (*Expense Income Ratio*), a new hypothesis for *Expense Income Ratio* is needed, hence the following formulation:

Null H5: Expense Income Ratio has no significant effect on interest rate

H5: Expense Income Ratio has a significant positive effect on interest rate

4.3 Scale validation

4.2.1 Multicollinearity

Performing the multicollinearity check revealed that all variables, but two, had low multicollinearity, close to 1. Only *verifiable net income* and *total expenses* had a VIF value of 3.67 and 3.54 respectively, indicating low to moderate multicollinearity (see Appendix 6.3).

4.2.2 Linearity

Despite extensive efforts to address nonlinearity, including transforming *verifiable net income* and *living expenses* by taking their log and ratio, the model remained nonlinear (appendix 9.1). More advanced modeling techniques might be necessary to address this issue and capture the underlying relationships in the data. This will be included in the limitation section of this report.

4.2.3 Normality

According to Paula Katherine Hagedorn Diehr (2002) the normality assumption for residuals is not always crucial for the proper inference on data if the dataset that is used is large enough. She employs the central limit theorem stating that with a large sample size the distribution of the sample mean approaches normality regardless of the distribution of the variables. In the scope of this research the large sample size fixes for any violation of this assumption and allows for accurate inference of the results, without being worried about the normal distribution of residuals.

4.2.4 Homoscedasticity

The presence of heteroscedasticity was detected by performing a Breusch-Pagan test, which returned a value of 5.256e-15 indicating strong significance, hence failing to reject the null hypothesis for homoscedasticity. However, this issue was fixed after taking the log of *verifiable net income, total expenses,* and *interest rate,* altering the result of the Breuscg-Pagan test to a p-value of 0.4173, which allows us to accept the null hypothesis for homoscedasticity.

4.4 Hypotheses testing

In Appendix 9.4 the results of the regression analysis can be observed. It tests for the relationship between the independent variables *expense income ratio, age,* and the dummified categories of *gender* and *education.* Additionally, two dummified control variables are included in the model: *Date of finance (Year)* and *Loan purpose. Loan purpose* can be observed through its previously discussed numerically encoded categories, while *Date of finance (Year)* can be observed through the *years* of which the loan was made. The adjusted R^2 is 0.2319, meaning that the variables included in the model explain 23.19% of the variance.

4.4.1 Education and interest rate

For each of the numerically encoded categories of *education* a p-value of < 0.05 is present. This leads us to conclude that the test is statistically significant and we reject the null hypothesis, hence meaning that higher *education* has a negative impact on the dependent variable *interest rate.* The positive correlation further supports the alternative hypothesis as the reference category *University* is the highest degree of education present.

4.4.2 Expense income ratio effect on interest rate

With a positive coefficient of 0.9151 *expense income ratio* reveals a positive relationship between an increase in the ratio between *living expenses* and *verifiable net income* and the *interest rate.* This is further supported by the p-value (< 2e-16), which allows us to reject the null hypothesis (Null H5) and conclude that *expense income ratio* has a significant positive effect on *interest rate.*

4.4.3 Age and Gender effect on interest rate

Age is statistically significant (p-value = 2.53e-07), indicating that as age increases the interest rate also increases. Additionally, the numerically encoded variable female also has a statistically significant relationship with *interest rate* with a p-value of 0.00016. This allows us to accept the alternative hypothesis that *age* and *gender* have a positive impact on the *interest rate.*

4.5 Control variables

The regression analysis incorporates several dummy variables for *loan purpose* and *date of finance (year).* The coefficients represent how the purpose of the loan affects the interest rate compared to the reference category Home Renovation. The same goes for the date *of finance (year),* with reference category the year 2016.

Certain *loan purposes,* such as car purchase, combination loans, and vacations, are associated with lower *interest rates* with coefficients -0.047, -0.053 and -0.076, respectively. Conversely, other *loan purposes* do not significantly differ from the reference category, home renovation.

As far as *date of finance (year)* is concerned only 2021 and 2022 show a significant decrease in interest rates with coefficients -0.255 and -0.242, respectively. Other years have no significant impact on the *interest rate,* with relation to the reference category.

4.6 Summary of the results

The first step of the data processing in this research was to check for multicollinearity issues. The high correlation between *verifiable net income* and *living expenses* was resolved by taking the ratio and creating a new variable called *Expense Income Ratio.* After that a linearity check of the variables was made, revealing nonlinear relationships between the independent and dependent variables. After this, it was deduced that the normality assumption was met affecting the inference of the data, as the distribution of the sample mean approaches

normality regardless of the distribution of the variables. Heteroscedasticity was detected through the Breusch-Pagan test, which returned a value of 5.256e-15 indicating strong significance. However, after the transformation of the numeric independent and dependent variables, this no longer was a problem. As far as the hypothesis testing was concerned the alternative hypothesis was accepted by: *age, expense income ratio, education,* and *gender.*

5. DISCUSSION

The goal of this research was to assess the relationship between borrower's characteristics and interest rate in the context of P2P lending in Bulgaria. To achieve this, four distinct hypotheses were formulated stemming from the research question, and one additional hypothesis was formulated due to data requirements previously discussed.

The variable *education* was split into five subsequent dummy variables, which all were found to have a significant effect on *interest rate,* leading us to reject the null hypothesis and accept the alternative hypothesis "Education has a negative effect on interest rate". This is further supported by their positive coefficients with regards to university, which were the reference variables, leading to the conclusion that borrowers with higher degree of education receive lower interest rates. The results are also in line with the literature (Junhui Xu et al., 2018).

The significant positive effect of *Expense Income Ratio,* allows us to reject the null hypothesis (Null H5) expense Income Ratio has no significant effect on interest rate, hence suggesting that borrowers with higher *living expenses* relative to their income are perceived as a higher risk by lenders, leading to higher *interest rates*. This result further promotes the signaling theory as a viable theoretical standpoint when dealing with P2P lending determinants of default or interest rate.

Significant results were found for the first hypothesis "Age and gender have a positive effect on the interest rate in P2P lending in Bulgaria" from the performed regression analysis with regards to *age* and *gender*. From the performed tests we can observe that being female as well as being an older borrower leads to higher interest rates due to a positive coefficient compared to the reference category, male. This creates a nuanced picture with regards to the literature as Tao et al. (2017) found no significant relationship between both independent variables *age* and *gender* and the dependent variable *interest rate*. However, these results are in line with findings from Ravina (2019), who similarly indicated a positive relationship between *age* and *interest rate.*

6. CONCLUSION

The fundamental question of this research was, "How strong is the relationship between borrower's characteristic and the interest rate applied on the loan in Peer-to-peer lending in Bulgaria?". The results of the regression analysis show that in the context of P2P lending in Bulgaria, the *interest rate* that is set on a given loan is influenced by the lenders' perception of the borrower's *education, age, expense income ratio* and *gender.* Higher levels of *education* reduce the perceived risk from the lender's perspective confirming assumptions deriving from the signaling theory. Similarly, higher *expense income ratio* signals to the lender financial instability and increases the possibility of a higher *interest rate*. Additionally, confirming the literature proposed by Ravina (2019) older borrowers are offered a higher *interest rate.*

In 6.1 theoretical and practical implications will be discussed, while in 6.2 limitations of the study as well as future research will be mentioned.

6.1 Implications

6.1.2 Theoretical implications
In terms of theoretical implications this study has explored the extent to which the signaling theory addresses an underlying problem in the alternative finance sector as a whole, information asymmetry. The findings in this paper are in line with the suggestion made by Carlos Serrano-Cinca (2015) that borrower's characteristics have the potential to reduce information asymmetry.

The dual-process model for lending decisions has been employed in this paper helping to frame the underlying decision making process of lenders. By employing this theory in the context of P2P lending it helps us understand lender's behavior with regards to determination of interest rate on a loan. Even though borrower's characteristics fall under the category of heuristic processing, which is characterized by simplistic interpretations of given information, they are still considered valuable as per the significance of the regression model.

6.1.3 Practical implications
Two valuable practical implications can be derived from the results of this study. Firstly, the significant relationship between *expense income ratio* and lower interest rate, underscores the importance of accurate income verification in P2P lending. This will potentially allow for lower perceived risk and more fluent lender - borrower interactions, which will consequently attract more creditworthy borrowers.

Secondly, the results of the regression reveal that borrower's characteristics play a significant role in the determination of risk. Consequently, they should include more borrower's characteristics in the data for a more comprehensive analysis to ensure maximum reduction of information asymmetry and subsequently perceived risk.

6.2 Limitations and future research
Three limitations can be recognized in this study. Firstly, we are investigating only fixed *interest rates* on loans, as that falls under Klear's policy. The results of the regression cannot be generalized to variable interest rate loans as the lender's reluctance to engage in such an investment may increase or decrease if variable interest rate was present. However, even though that narrows the scope of the research, a direct comparative analysis becomes feasible with regards to other potential research.

Secondly, a more diverse sample with regards to the variable *education* would be preferred as 65% of borrower's have graduated university and 25% have finished secondary school. This heavy skewness towards higher education can destabilize the results of the regression analysis and is considered not representative, hence hardly generalizable.

Thirdly, the fitted model remained nonlinear despite extensive preprocessing of data. This issue should be further investigated, potentially by employing more sophisticated data preprocessing measures such as Generalized Additive Models (GAMS) or other machine learning approaches.

Recommendation for potential future research would be to perform a similar study to this but for a different country. This would help for a direct comparable analysis and would further explore the effectiveness of signaling theory on the information asymmetry issue. Furthermore, a similar study with variable interest rate would provide valuable insights on lenders and borrowers behavior with regards to perceived risk. Finally, in this study were employed only several of borrower's characteristics. There is scarcity of comprehensive studies on the effect of type of employment, family situation, number of

children and other borrower's characteristics on the *interest rate*.

7. ACKNOWLEDGEMENTS

8. REFERENCE

Abbasi, K., Alam, A., Brohi, N., Brohi, I., & Nasim, S. (2021, July). P2P lending Fintechs and SMEs' access to finance. Elsevier.

Alexander Bachmann and Burkhardt Funk (2011). Online peer-to-peer lending – A literature.

Andreas Dietricha and Reto Wernlib (2015). What Drives the Interest Rates in the P2P Consumer Lending Market? Empirical Evidence from Switzerland

Andreas Dietricha and Reto Wernlia (2019). Determinants of Interest Rates in the P2P Consumer Lending Market: How Rational are Investors?

Cader HA, Leatherman JC (2011) Small business survival and sample selection bias. Small Bus Econ

Caglayan, M., Pham, T., Talavera, O., & Xiong, X. (2020). Asset mispricing in peer-to-peer loan secondary markets.

Carlos Alberto, Gomez Grajales and Dason Kurkiewicz,. (2013). Assumptions of Multiple Regression: Correcting Two Misconceptions.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion.

Chen, D., Li, X., & Lai, F. (2017). Gender discrimination in online peer-to-peer credit lending: evidence from a lending platform in China.

Chen, D., Lou, H., & Van Slyke, C. (2015). Toward an understanding of online lending intentions: Evidence from a survey in China.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.).

Culkin, N., Ekaterina M., & Andrew D. (2015). Critical innovations in the UK peer-to-peer (P2P) and equity alternative finance markets for small firm growth.

Culkin, N., Ekaterina M., & Andrew D. (2015). Critical innovations in the UK peer-to-peer (P2P) alternative finance markets for small firm growth.

Duarte, J., Siegel, S., & Young, L. (2012). Trust and Credit: The Role of Appearance in Peer-to-Peer Lending.

Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2014). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. Applied Economics.

Enrichetta Ravina (2019). Love & Loans The Effect of Beauty and Personal Characteristics in Credit Markets

Guo, G., Zhu, F., Chen, E., Liu, Q., Wu, L., & Guan, C. (2016). From footprint to evidence: An exploratory study of mining social data for credit scoring.

Haitian Lu, Bo Wang, Haizhi Wang and Tianyu Zhao (2020). Does social capital matter for peer-to-peer-lending? Empirical evidence.

Hakkon Kim, Kwangwoo Park, Byungtae Lee and Eunho Choi (2013). A Study on Determinants of Loan Repayment in Peer-to-Peer Lending

Hassan, M. (2022, September 19). Control Variable - Definition, Examples. Research Method. https://researchmethod.net/control-variable/.

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.

HONGKE ZHAO, Anhui Province Key Laboratory of Big Data Analysis and Application, University of Science and Technology of China

Hulme, M. K., & Wright, C. (2006). Internet based social lending: past, present and future.

Jason W. Osborne, Elaine Waters,. (2019, November). Four assumptions of multiple regression that researchers should always test.

Jamal I. Daoud (2017). Multicollinearity and Regression Analysis.

J. Huang, H. Maassen van den Brink And W. Groot (2010). A Meta-Analysis of the Effect of Education on Social Capital.

Kagan, J. (2024, January 11). Peer-to-Peer Lending: The Lowdown. Investopedia. https://www.investopedia.com/terms/p/peer-to-peer-lending.asp

Khalil, M. (2023, January 18). Multicollinearity: A Guide to Understanding and Managing the Problem in Regression Models. Medium. https://becominghuman.ai/multicollinearity-a-guide-to-understanding-and-managing-the-problem-in-regression-models-d953099a2bdd.

Kuwabara, K., & Thébaud, S. (2017). When beauty doesn't pay: Gender and beauty biases in a peer-to-peer loan market.

Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. Management Science.

Martin Flodén, Matilda Kilström, Jósef Sigurdsson, Roine Vestman (2021). Household Debt and Monetary Policy: Revealing the Cash-Flow Channel

Martina E. Greiner & Hui Wang (2009). The Role of Social Capital in People-to-People Lending Marketplaces.

Matt N. Williams.

Ming Yang, Haiyang Li, Zhen Shao, and Wei Shang (2017). Influencing Lenders' Repeat Investment Intention in P2P Lending Platforms in China through Signaling

Mohammed m Elgammal and Shabeen Basha,. (2021). Journal Pre-proofs Online Peer-To-Peer Lending: A Review of the Literature.

Paula Katherine Hagedorn Diehr (2002). The Importance of the Normality Assumption in Large Public Health Data Sets

Petersen, M. A. and José María Liberti (2004). Information: Hard and soft.

Professor Junhui Xu, Professor Jitka Hilliard and Professor James R. Barth (2018) On Education Level and Terms in Obtaining P2P Funding: New Evidence from China

Qi Liu, Guifeng Wang, Enhong Chen, and Hefu Zhang (2017) P2P Lending Survey: Platforms, Recent Advances and Prospects

Rabbani, R., Bashar Abu, Hawaldar T., Shaik M., and Selim M. (2022). What Do We Know about Crowdfunding and P2P Lending Research? A Bibliometric Review and Meta-Analysis.

Rainer Lenz (2016). Peer-to-Peer lending – opportunities and risks

Robert H. Shumway & David S. Stoffer (2017). Time Series Analysis and Its Applications.

Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of Default in P2P Lending. PLOS ONE, 10(10), e0139427. https://doi.org/10.1371/journal.pone.0139427.

Shiyi Chen, Yan Gu, Qingfu Liu and Yiuman Tse (2020). How do lenders evaluate borrowers in peer-to-peer lending in China?.

Steve McDonald and Christine A. Mair (2010). Social Capital Across the Life Course: Age and Gendered Patterns of Network Resources.

Tao, Q., Dong, Y., & Lin, Z. (2017). Who can get money? Evidence from the Chinese peer-to-peer lending platform.

Tania Ziegler et al. (2021). The 2nd Global Alternative Finance Market Benchmarking Report.

Theodoros Kyriazos1, Mary Poga,. (2023). Dealing with Multicollinearity in Factor Analysis: The Problem, Detections, and Solutions.

Thomas Lumley, Paula Diehr, Scott Emerson, and Lu Chen (2002). The Importance of the Normality Assumption in Large Public Health Data Sets

Tomi Hietala et al. (2016). Determinants of interest rate of peer-to-peer business loans.

Tristan Claridge (2018). Introduction to Social Capital Theory.

Wang, H., Chen, K., Zhu, W., & Song, Z. (2015). A process model on P2P lending. Financial Innovation, 1(1). https://doi.org/10.1186/s40854-015-0002-9.

Zach Bobbitt. (2020, January 8). The Four Assumptions of Linear Regression. Statology. https://www.statology.org/linear-regression-assumptions/.

Zach. (2021, January 21). The Durbin-Watson Test: Definition & Example. Statology. https://www.statology.org/durbin-watson-test/.

Zhang, K. Z. K., Zhao, S. J., Cheung, C. M. K., & Lee, M. K. O. (2014). Examining the influence of online reviews on consumers' decision-making: A heuristic-systematic model.

Ziegler, T., and Shneor, R. (2020). The global alternative finance market benchmarking report.
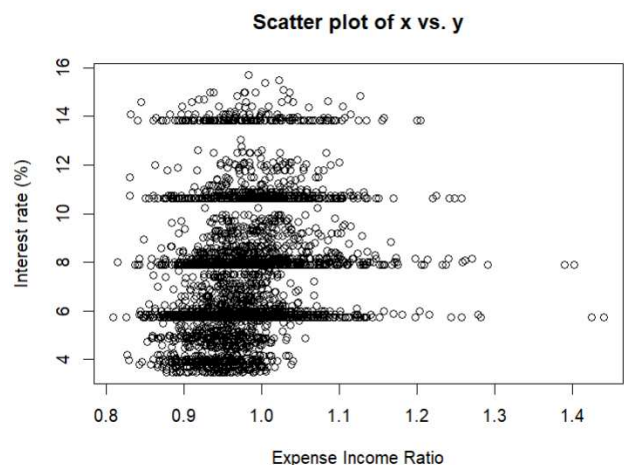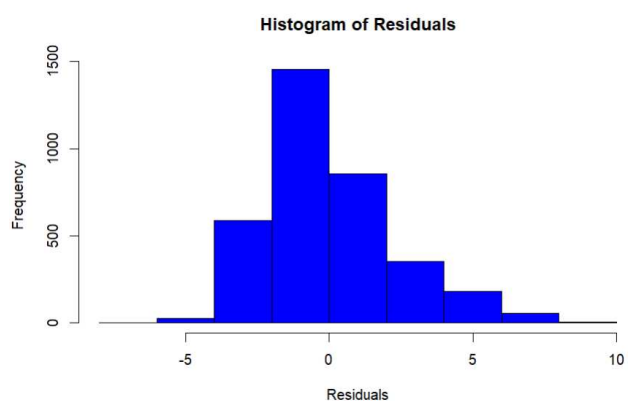
9. APPENDIX

9.1 Linearity

Figure 3. Expense Income Ratio Scatter Plot

9.2 Normality of residuals

9.3 Regression results

Coefficients:	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	1.1968304	0.0880383	13.594	< 2e-16 ***
Age	-0.0029838	0.0005777	-5.165	2.53e-07 ***
Expense Income Ratio	0.9150776	0.0835913	10.947	< 2e-16 ***
Female	0.0413851	0.0109508	3.779	0.00016 ***
Primary education	0.4319413	0.2096076	2.061	0.03940 *
Basic education	0.479921	0.0771193	6.223	5.45e-10 ***
Vocational High School	0.1843838	0.0212637	8.671	< 2e-16 ***
Secondary Education	0.2121361	0.0120508	17.603	< 2e-16 ***
College	0.1204613	0.0272469	4.421	1.01e-05 ***

Unexpected expenses	0.0352003	0.0207098	1.7	0.08928
Car purchase	-0.0473484	0.0160998	-2.941	0.00329 **
Other	0.0019877	0.0192003	0.104	0.91755
Loan Refinancing	-0.0068438	0.0157571	-0.434	0.66408
Combination	-0.0533361	0.018554	-2.875	0.00407 **
Property Purchase	-0.0092419	0.0223576	-0.413	0.67936
Medical needs	0.0397673	0.0298126	1.334	0.18232
Vacation	-0.0759303	0.0330616	-2.297	0.02170 *
Business	0.0573174	0.0490569	1.168	0.24273
Year 2017	-0.0057102	0.037569	-0.152	0.8792
Year 2018	-0.0235563	0.0373759	-0.63	0.52857
Year 2019	-0.0552014	0.0368473	-1.498	0.13419
Year 2020	-0.0683063	0.0374669	-1.823	0.06837
Year 2021	-0.2548908	0.0367699	-6.932	4.92e-12 ***
Year 2022	-0.2420231	0.0359863	-6.726	2.03e-11 ***
Year 2023	-0.0180482	0.0361521	-0.499	0.61765
Year 2024	-0.0417069	0.0416235	-1.002	0.31641

9.5 Correlation matrix (before ratio)

	Interest	Age	Verifiable	Total

	rate (%)		net income	expenses
Interest rate (%)	1	-0.06	-0.28	-0.26
Age	-0.06	1	0.06	0.15
Verifiable net income	-0.28	0.06	1	0.81
Total expenses	-0.26	0.15	0.81	1

9.6 Correlation matrix (after ratio)

	Interest rate (%)	Age	Expense Income Ratio
Interest rate (%)	1	-0.06	0.22
Age	-0.06	1	0.11
Expense Income Ratio	0.22	0.11	1

9.7 Variance of residuals

Histogram of Residuals