FINTECH'S POWER PLAY: THE IMPACT OF BANK COMPETITION ON P2P LENDING

By

Zsófia Cuñat Lóránth

Submitted to Central European University – Private University

Undergraduate Studies

In partial fulfilment of the requirement for the Bachelor of Science in Data Science and Society

Supervisor: Tomy Lee

Vienna, Austria

2024

Copyright Notice

Abstract

This thesis examines how bank competition affects the pricing and markets served by peer-to-peer (P2P) lending platforms, focusing on the interaction between traditional banks and fintech companies. The thesis takes advantage of a large dataset from LendingClub, a formerly dominant lending platform in the United States, covering the period of 2012-2019 and combines it with state-level data about the banking sector from the FDIC and federal funds rate data. It conducts a multivariate regression analysis to uncover these interactions. The findings show that higher bank competition, measured by the Herfindahl-Hirschman Index (HHI), is associated with higher risk premiums and lower default probabilities on P2P lending platforms. These findings suggest that in competitive markets where traditional banks charge higher interest rates to mitigate adverse selection risks, fintech companies can minimally undercut bank rates, while maintaining profitability and loan quality through advanced screening technologies resulting in decreased loan defaults. Smaller loan sizes, typically associated with less creditworthy borrowers, are particularly sensitive to changes in bank competition, resulting in higher interest rates. The study provides some insights into the competitive role of fintech in the financial landscape, highlighting how P2P platforms can thrive even in highly competitive environments by utilizing innovative data sources and technologies.

Table of Contents

List of Figures

List of Tables

Introduction

The financial landscape has undergone a profound transformation with the rise of financial technology companies, challenging the longstanding dominance of banks in the financial sector. Traditionally, banks had a monopoly in payment services and lending to small and medium-sized businesses that constitute the backbone of the economy. However, advancements in technology, particularly in big data, machine learning, and AI, have paved the way for new financial institutions specializing in payment and lending services.

These services encompass the emergence of peer-to-peer (P2P) lending platforms, which connect investors with small and medium-sized companies as well as individual clients, engage in crowdfunding, and facilitate cross-border payments, among others. P2P lending platforms, which emerged around the wake of the financial crisis, often cater to underserved clients, such as small businesses, subprime clients, individuals with insufficient credit history, or those with lower job security. Such platforms often excel in obtaining new data sources, and innovative business models. This can create pressure for banks to optimize their loan underwriting procedures and employ better and faster data analytics. So far, this competitive pressure has been limited in most established markets, with Fintech credit still making up only a small proportion of overall credit in most countries. Lending platforms have also started catering to segments where there is little to no competition from banks, such as unbanked clients and other underserved segments (FSB 2019). The emergence of these new players and their interaction with traditional banks raise interesting questions as to the potential benefits of fintech firms, their role in the financial industry, the target groups they cater to and the possible competition they may create for traditional banks.

This thesis specifically focuses on P2P lending services and addresses the question: How does bank competition affect the prices and market served by peer-to-peer lending platforms? This question is answered through regression analysis, as well as an in-depth review of the existing literature. The significance of this topic lies in the crucial roles that banks and, more recently, fintech companies play in the lending market. Understanding how they interact with each other is essential, as their interactions can lead to improvements in financial services or potentially hurt the supply of credit, with consequences on the real economy.

The literature on P2P lending is small and centres on the debate on whether P2P lending is a substitute for or a complement to traditional bank lending. Some studies emphasize the financial inclusion provided by fintech (Gopal and Schnabl 2022), while others highlight its role as a substitute in the context of infra-marginal borrowing (Tang 2018). Some researchers suggest that instead of crowding out banks, the presence of fintech, particularly through peer-to-peer lending platforms, reduces moral hazard issues (Balyuk 2023) and encourages more credit lending from traditional banks (Beaumont, Tang, and Vansteenberghe, 2022). Banks, more constrained by regulations, typically require collateral for lending to small companies. With the emergence of fintech, such small firms can now obtain loans from fintech platforms and use those funds to purchase assets, which they can then pledge as collateral to banks.

This thesis undertakes a cross-state analysis utilizing a dataset from LendingClub (Nigmonov 2021), which was one of the leading peer-to-peer lending platforms in the US. The dataset consists of over two million observations and is aggregated with US state-level data, spanning from the interval 2012 to 2019. The choice of data is restricted by the little data availability, given that most bank and fintech data is private. This study also utilizes the FDIC state-level data on bank competition (FDIC 2011, FDIC 2015) as well as data on the federal funds effective rate (FRED 2024) for the same time window. The choice of performing a cross state analysis is justified by several factors, including the potential heterogeneity across different regions in terms of local GDP, unemployment, bankruptcy, institutions and the level of bank competition. At the same time, it is important that all states share the same regulatory framework, for example in terms of consumer-protection or market-conduct regulation, and monetary policy is conducted at the federal level. Focusing on counties in a specific state instead of conducting a cross-state analysis would have raised the question about the potential generalizability of findings to other regions within the USA.

The regression analysis reveals a number of findings regarding the impact of bank competition on risk premiums, defined as the difference between the interest rate charged to a specific borrower and the effective fed-fund rate, and default likelihood in P2P lending platforms. Firstly, the results indicate that higher market competition, measured by the Herfindahl-Hirschman Index, is associated with slightly higher risk premiums, aligning with economic theory (Yannelis and Zhang 2023) and previous banking research (Petersen and Rajan 1994). This suggests that as competition among banks increases, P2P lending platforms may adjust their pricing strategies to mitigate the effects of adverse selection. As higher competition reduces banks' incentives to screen, it leads to higher prices in order to compensate for the higher adverse selection. This in turn means that a P2P lending platform can also increase prices as competition increases: charging just a little bit less than banks, and a can still attract clients. Additionally, the analyses show that smaller loan sizes, usually attributed to less creditworthy borrowers, are more sensitive to changes in bank competition. Lastly, the data suggests that higher market competition correlates with decreased default probabilities. This is also in line with theory. As higher bank competition means that banks screen less and therefore end up with worse average quality clients, P2P platforms can leverage advanced screening technology and alternative data sources to offer competitive rates while maintaining strong risk assessment, giving them a competitive edge in the lending market.

The thesis starts with a literature review exploring the substitution versus complementarity debate within the context of LendingClub data and examining how competition in the credit market impacts loan prices. Next, it provides a detailed description of the data used in the analysis. This is followed by a methodology chapter that presents descriptive statistics, explains data preprocessing steps, and describes the model setup. Subsequently, the thesis discusses the results and reflects on the generalizability of the quantitative analysis in the context of the literature. Finally, it highlights the limitations of the data and analysis.

1 Literature Review

The first two papers in this literature review are key for understanding the data utilized in this thesis and defining the methodology. The following two papers provide insights into banks operations in competitive markets, facilitating the drawing of parallels to the fintech industry and interpretation of the results presented in the quantitative analysis.

My thesis is closely related to Tang's (2018) paper "Peer-to-Peer Lenders Versus Banks: Substitutes or Complements," in which she studies whether P2P lending platforms serve as bank substitutes or as complements. Similarly to this thesis, she utilizes data from LendingClub, however, for an earlier time window, namely for the years 2009 to 2012. Contrary to the approach taken in this thesis, she opts for using a diff-in-diff analysis, taking advantage of a regulatory change mandated by the Financial Accounting Standards Board in the first quarter of 2011. The regulatory change tightened banks' capital constraint and led to a reduction in bank credit supply. Tang (2018) uses this exogenous shock to bank credit supply to examine whether P2P lending is a substitute or complement for bank lending. She finds that P2P lending is a substitute to bank loans. Borrowers who apply to P2P after the regulator shock apply to much larger loans than P2P existing clients. At the same time, she finds that P2P lending is complementary with respect to small loans. Her data also differs in the sense that she has access to information on all loan requests, not just those clients whose loan request is approved. Her paper also includes a valuable insight into the loan distribution process of LendingClub, which is no longer publicly available on the LendingClub website.

In terms of methodology, the closest paper to mine is Degryse and Ongena's (2005) "Distance, Lending Relationships, and Competition". The paper does not address the impact of Fintech companies. Instead, it focuses on bank competition and the influence of geographical distance between firms and banks on loan conditions. The authors find that loan rates decrease in distance between the firm and the lending bank. Although this paper does not directly address the research question, it was instrumental in shaping the regression analysis on loan prices. The methodology used, with loan prices as the target variable and distance as a key independent variable, parallels the approach taken in this thesis. In this case, however, the key independent variable is bank competition in different states as opposed to distance between the firm and the bank serving the firm. Furthermore, the classification of loans into different size buckets in their paper inspired a similar pairwise regression analysis in this thesis, which examines how the risk premium on different loan amounts is influenced by market competition.

In interpreting my results, I found the paper by Yannelis and Zhang (2023) particularly useful. They argue that in more competitive markets, because lenders have lower market shares, they have reduced incentives to invest in screening. As a consequence, banks end up with a riskier pool of borrowers and charge higher interest rates in order to protect themselves for higher expected defaults. They find evidence consistent with their model by using administrative credit panel data in the auto loan market. This finding aligns with the conclusions of the quantitative analysis conducted in this thesis, which indicates that when market share concentration is lower, i.e., bank competition is higher and therefore banks charge higher interest rates to possibly account for adverse selection, fintech companies can also increase their prices without losing (too many) clients. Although competition has only a slight effect on loan prices in this thesis, it significantly impacts the likelihood of loan defaults, with fewer defaults occurring when competition is higher. This result again can be understood in an adverse selection context. When banks' pool of clients is worse, the benefit of using a superior screening technology, that a P2P platform can have access to, is higher. Thus, a P2P platform is more likely to end up with better clients.

In their 1994 paper "The Effect of Credit Market Competition on Lending Relationships," Petersen and Rajan highlight how competition in the credit market affects loan prices and amounts. While this study concentrates on banks rather than fintech companies, the comparable operations of fintech firms and banks render its findings relevant to this analysis. The paper finds that in concentrated markets, characterized by higher HHI, lenders are more likely to offer lower initial interest rates to young or credit-constrained borrowers in anticipation of being able to service these firms also when they get older and uncertainty about them reduces (intertemporal cross-subsidization). Credit market competition allows the firm to switch lender when it gets older. This in turn destroys banks' incentive to offer lower rates when the firm is young and credit constrained. Although the paper focuses on small businesses rather than individual borrowers, it highlights the potential for loan prices to increase in competitive environments.

2 Data

2.1 LendingClub Data

2.1.1 History

LendingClub, initially launched on Facebook, developed into a full-scale peer-to-peer lending company in August 2007. Headquartered in San Francisco, LendingClub became the first peer-to-peer lender to offer loan trading on a secondary market. At its peak, it was the world's largest peer-to-peer lending platform. LendingClub enabled borrowers to create unsecured personal loans ranging from $1,000 to $40,000, with investors earning interest from these loans. However, like other peer-to-peer lenders, LendingClub faced difficulties attracting investors in early 2016, leading to increased interest rates and a significant drop in its share price. In December 2020, LendingClub ceased its peer-to-peer lending operations.

2.1.2 Business Model

For this analysis, it is important to understand the business model of LendingClub to address the potential endogeneity issues related to how loans and interest rates are determined. To apply for a loan on LendingClub, applicants provide their name, address, loan purpose, and amount requested. The platform then acquires their credit report and deems ineligible any applicant with a debt-to-income (DTI) ratio above 0.35 or a FICO score below 660. Eligible applicants receive a menu of loans with different amounts, maturities (36 or 60 months), and interest rates. Once an applicant selects a loan, it is listed on the website for investors to fund on a first-come, first-served basis. Investors see loan characteristics and some borrower credit report information. Applicants also report their income, profession, and employment length, with LendingClub verifying this information for some applicants. LendingClub's national pricing policy ensures uniform rules across locations. The interest rate is set by assigning a loan grade and adding an adjustment to the platform's base rate for that grade (Tang 2018). Potential endogeneity concerns related to simultaneous determination of loan amount and interest rate by LendingClub are mitigated as interest rates are determined independently after borrowers submit loan size requests.

2.1.3 Data

Unlike the data used in Tang's paper (Tang 2018), the dataset utilized in this thesis (Nigmonov 2021), retrieved from the Mendeley Data repository, only contains successful loan applications introducing both selection bias and lack of comprehensive risk assessment which will be further discussed in a later chapter. The dataset contains over 2.7 million observations, with diverse loan, borrower and state-specific features, with each observation accounting for a specific loan.

Relevant loan characteristics to this thesis include, the interest rate on the loan, loan amount, the duration of the loan, which is either 18 or 36 months, the loan grade and the purpose of the loan. Borrower characteristics are the self-reported annual income of the borrower, whether they own a home or have a mortgage, their address, their profession and years of employment, the debt to income ratio (DTI), whether they have a history of bankruptcy and the number of open accounts they have. The dataset also discloses whether the borrower defaulted on the loan. Exact descriptions of each variable can be found in the appendix (Table A1-A3).

2.2 FDIC Data

The Herfindahl-Hirschman Index (HHI), which is the key causal variable in this analysis, is a commonly accepted measure of market concentration. The HHI is calculated by squaring the market share of each firm in a market and summing the results. It measures market concentration, approaching zero when there are many firms of similar size. Higher HHI values indicate fewer firms or larger disparities in firm sizes. The values for each state can be obtained from the Pro Forma (HHI) Report provided by the FDIC (FDIC 2011, FDIC 2015). This report includes a summary of all FDIC-insured institutions within the specified geographic area, in this case per state, as well as the number of branches, the amount of deposits, the market share and the Herfindahl-Hirschman index.

Unfortunately, the index for each state and year must be obtained individually, requiring 400 separate searches to retrieve the relevant data from the LendingClub dataset. To accommodate time constraints, this thesis utilizes HHI data from two specific years, 2011 and 2015. Loans issued between 2012 and 2015 utilize the 2011 index values, whereas loans issued between 2016 and 2019 utilize the 2015 HHI data. The two data points are motivated by the observation that there has been a downward trend in the HHI for the period of 2012-2019 covered in this thesis.

2.3 Federal Funds Effective Rate Data

Given that interest rates vary from year to year, the federal funds effective rate serves as a key benchmark for other interest rates in the economy, influencing borrowing costs, inflation, and economic growth. The loan price regression has as its target variable the risk premium that LendingClub charges, which is determined by subtracting the federal funds effective rate from a specific month and year from the interest rate of the loan. This data was exported from the Federal Reserve Economic Data database, maintained by the Federal Reserve Bank of St. Louis (FRED 2024).

3 Methodology

This analysis employs a structured approach to investigate the impact of bank competition on the prices and market dynamics of peer-to-peer lending platforms. The primary variables of interest are the risk premium and the quality of LendingClub clients (‘badloan’), and the Herfindahl-Hirschman Index (HHI). The risk premium is defined as the price of the loan minus the federal funds rate. The 'badloan' variable is an indicator that denotes whether an individual defaulted on their Lending Club loan. The HHI serves as a measure of market competition, where a higher HHI indicates higher market concentration and thus lower competition. Other variables include loan and borrower characteristics.

The central research question guiding this analysis is: How does bank competition affect the prices and market served by peer-to-peer lending platforms? The analysis to answer this question is conducted in two main steps.

Firstly, an Ordinary Least Squares (OLS) regression is performed with the risk premium as the dependent variable. The independent variables include borrower characteristics, loan characteristics, state fixed effects, year fixed effects, and the HHI for the specific state and time window. This step aims to determine whether higher market competition, indicated by a lower HHI, influences the risk premium offered by LendingClub, while controlling for borrower and loan characteristics.

Secondly, a logit regression is conducted with 'badloan' as the dependent variable, incorporating the same control variables as the OLS regression. This step examines whether bank competition affects the likelihood of loan defaults, thereby indicating the market served by fintech companies. The underlying hypothesis is based on the adverse selection problem, suggesting that higher market competition (lower HHI) correlates with lower default rates, as proposed by Yannelis and Zhang.

This two-step approach allows for a comprehensive examination of both the pricing and market dynamics influenced by bank competition in the context of peer-to-peer lending platforms.

3.1 Descriptive Statistics

Table 1 presents summary statistics for two variables: the risk premium and the Herfindahl-Hirschman Index. In the LendingClub dataset, the mean loan premium is 12%, ranging from 3% to 31%.

Table 1: Summary Statistics for HHI and Risk Premium

	HHI	Risk Premium
count	2703430	2703430
mean	0.11	0.12
std	0.07	0.05
min	0.02	0.03
25%	0.08	0.08
50%	0.09	0.12
75%	0.12	0.15
max	0.63	0.31

Figure 1 displays box plots illustrating the distribution of risk premiums from 2012 to 2019, indicating a slight decreasing trend in the median risk premium over the years. Outliers start appearing in 2015 and become more prominent in subsequent years, suggesting a growing number of loans with significantly higher risk premiums. The interquartile range remains relatively stable until 2017 but expands notably in 2018 and 2019, implying increased variability in risk premiums during these years. This pattern may reflect changing market conditions or adjustments in LendingClub's lending criteria and risk assessment strategies over time.

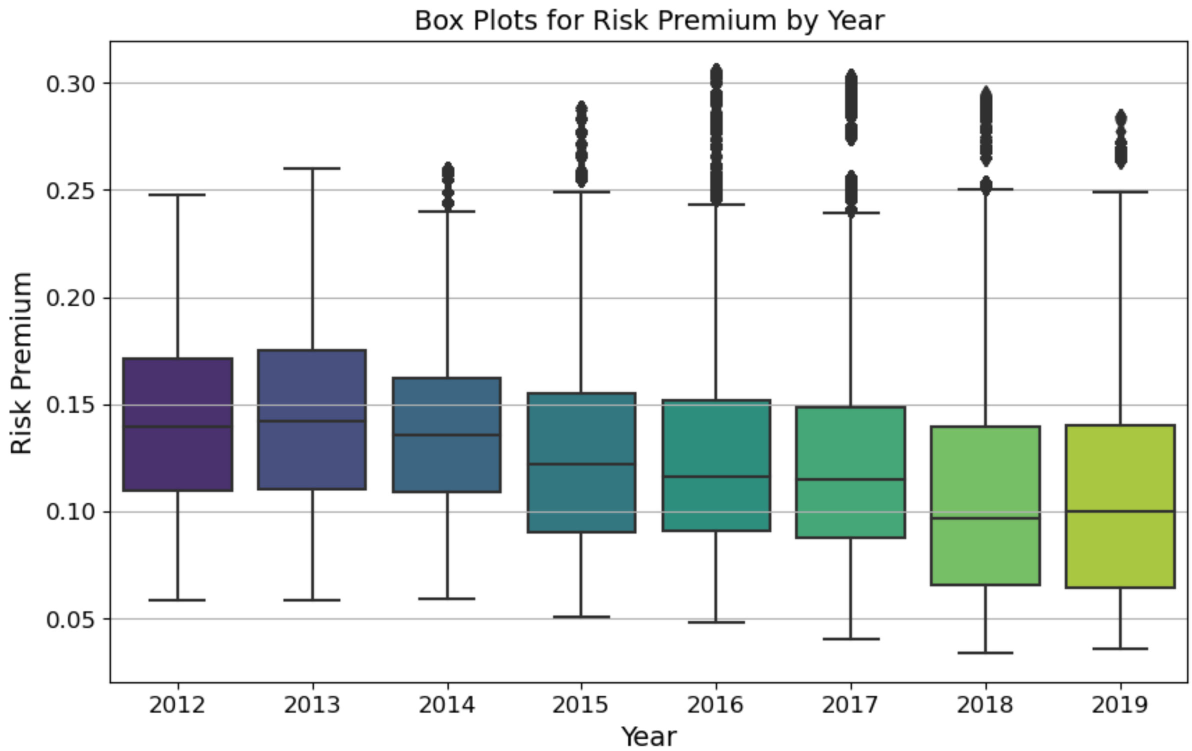

Figure 1: Box Plots for Risk Premium by Year

Defaults comprise around 8% of loans in the dataset. Figure 2 presents this comparison on a yearly basis, revealing a higher default rate in the first four years, particularly in 2015, compared to the latter half of the time interval. This difference could potentially be attributed to an enhancement in Lending Club’s screening process.

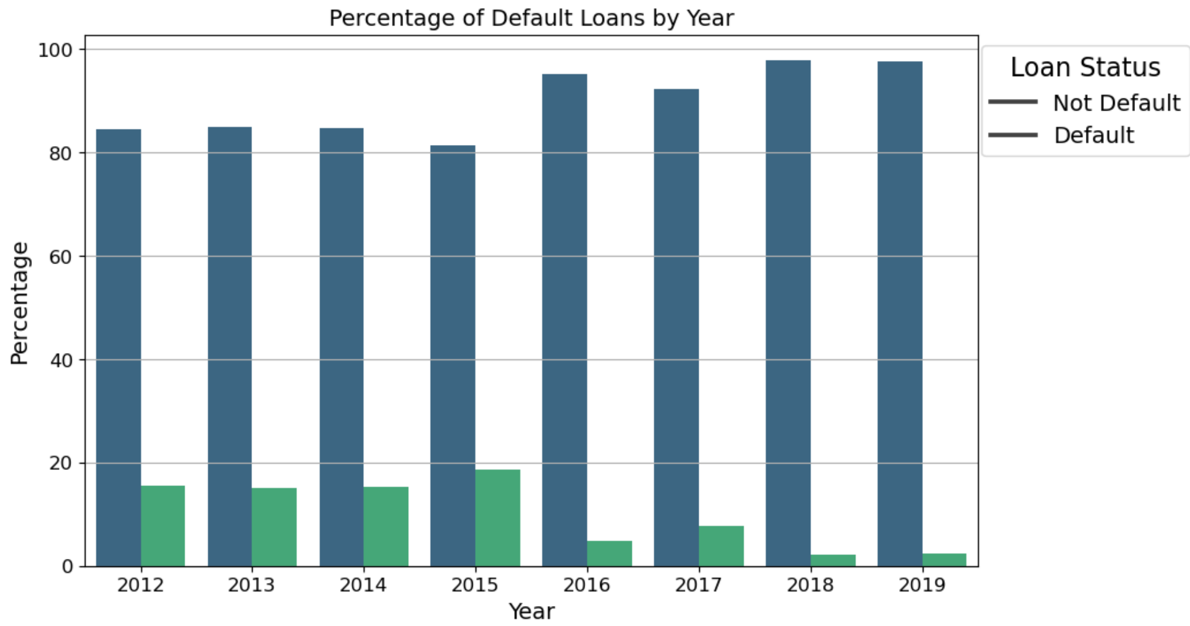

Figure 2: Percentage of Defaulted Loans by Year of Issue

Table 2 offers summary statistics for the most significant continuous independent variables. As previously discussed, LendingClub provides loan amounts ranging from $1,000 to $40,000, with the mean loan amount slightly exceeding $15,000.

Table 2: Summary Statistics on Continuous Independent Variables

	Loan Amount	Annual Income	DTI	Open Account	Bankruptcies
count	2703430	2703430	2703430	2703430	2703430
mean	15396.9	79762	19.35	11.7	0.13
std	9454.19	112786	22.21	5.73	0.36
min	1000	1	0.01	1	0
25%	8000	47000	12.12	8	0
50%	13125	66000	18.14	11	0
75%	20150	95000	24.95	15	0
max	40000	1.1e+08	9999	104	12

Annual income spans from $1 to $95,000, although it should be noted that an annual income of $1 is not very realistic, suggesting a potential bias in the data. On average, loan borrowers have an annual income of $80,000. Public bankruptcies range from zero to twelve, with the interquartile range indicating that most LendingClub clients have zero bankruptcies on record. The bar charts in Figure 3 illustrate the most important categorical predictor variables.

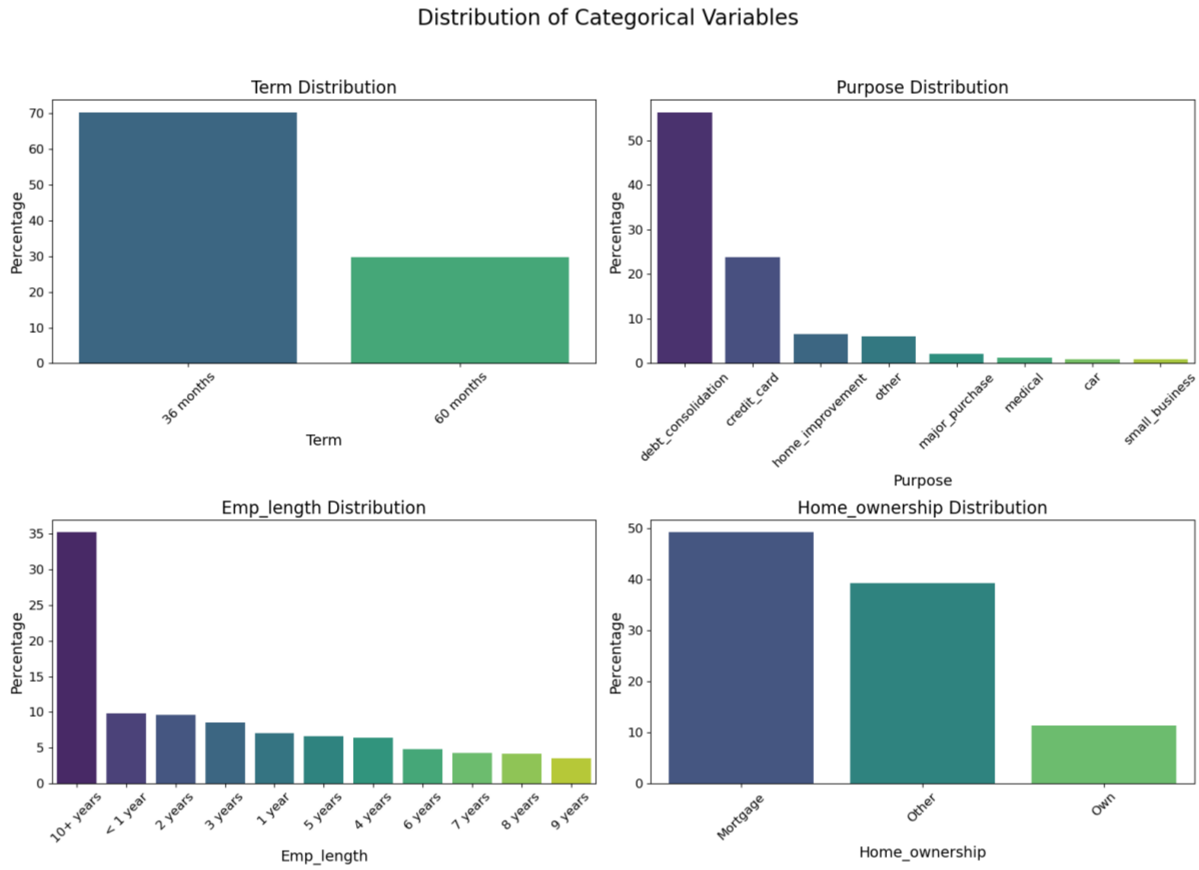

Figure 3: Distributions of Important Predicor Variables

More than two-thirds of loans have a 3-year duration period. Debt consolidation constitutes the primary loan purpose (~60%) in the dataset, while small business lending only represents around 25,000 observations. Nearly 50% of borrowers have a mortgage, with just over 10% owning a house. This measure can be comparable to collateral or a signal of the borrower's trustworthiness, even though LendingClub's loans are unsecured.

Finally, it is worth examining whether there is heterogeneity in the number of loans distributed per state and year. Figure 4 reveals that California has the largest number of loans, totalling around 14%, which aligns with LendingClub's headquarters being located in San Francisco, California. Texas, New York, and Florida follow suit.

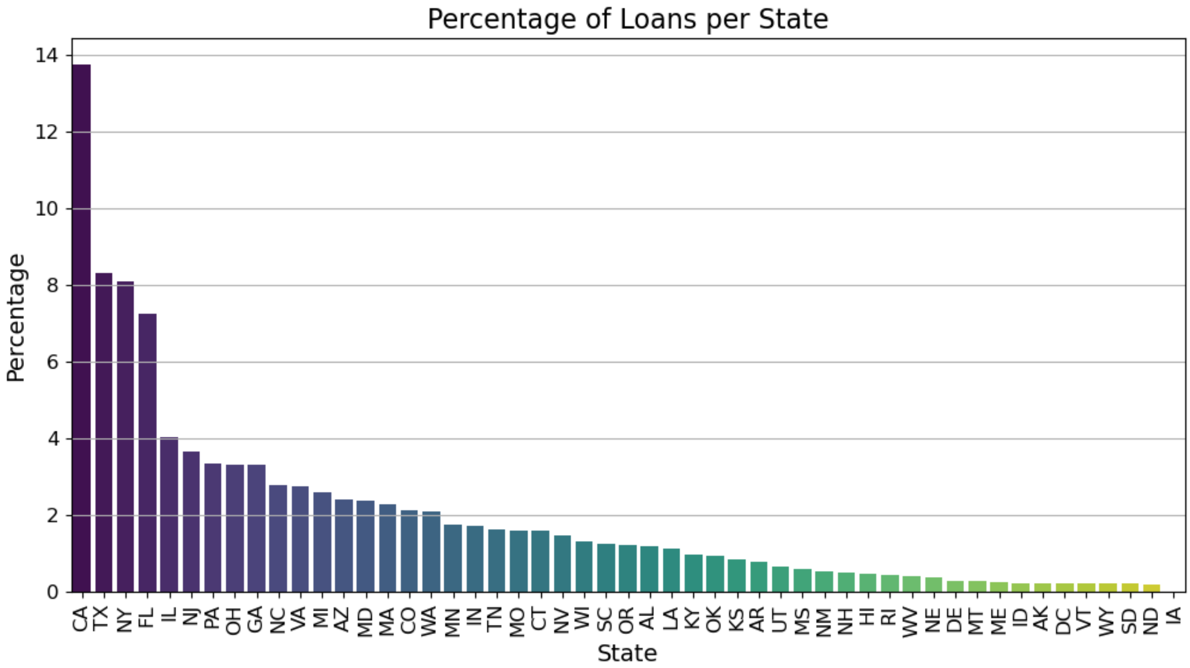

Figure 4: Percentage of Loans per State

There is a noticeable increase in loan amounts over the time interval (Figure 5), with the most significant increase occurring between 2012 and 2015, followed by a more moderate increase in the second time window.

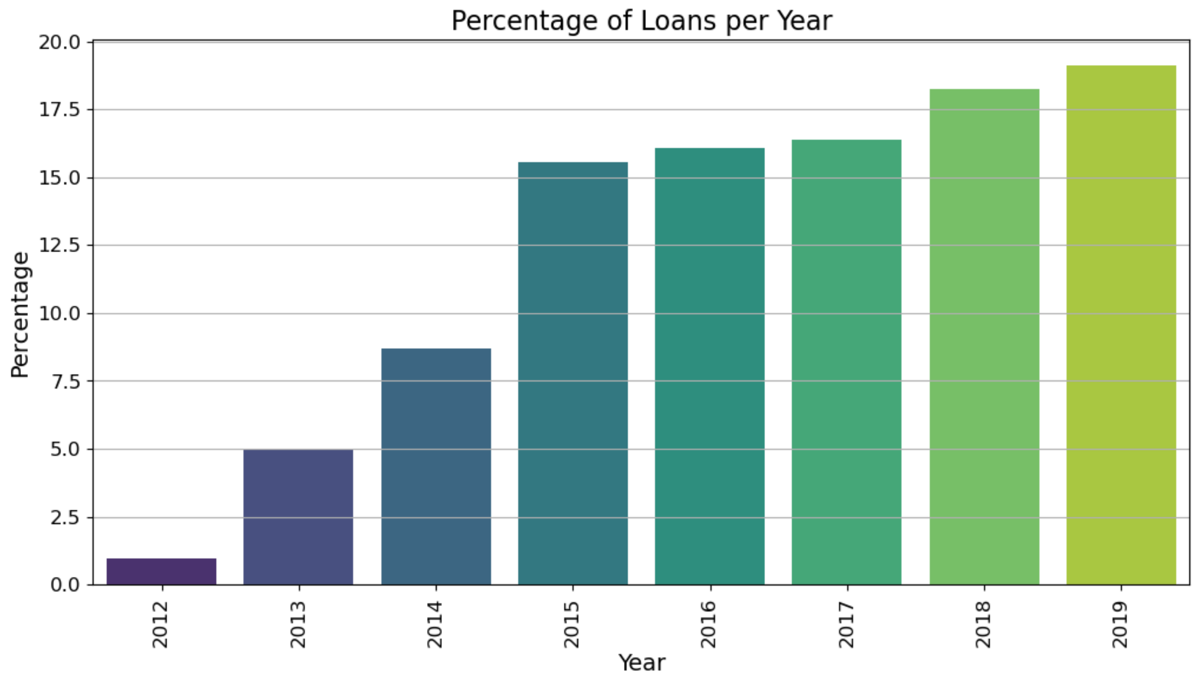

Figure 5: Percentage of Loans per Year

3.2 Data Preprocessing

All independent variables in this analysis are standardized to enhance the interpretability of their influence, and logarithmically transformed when necessary to address skewness and heteroscedasticity. Additionally, categorical variables such as state, loan term, homeownership, application type, loan purpose, employment length, issue year, bankruptcy records, and income verification status, were converted into dummy variables for both regressions.

3.3 Correlation Maps

In preparation for regression analysis, it is crucial to assess potential multicollinearity among independent variables, as high correlations can adversely affect the reliability of the model. Therefore, three distinct correlation metrics were employed for different variable types.

Firstly, Spearman's rank correlation coefficient was utilized for continuous variables. The use of Spearman's rank correlation coefficient was particularly relevant for continuous variables that had been log-transformed. In analysing the correlation heatmap (Figure 6), it is evident that there are no instances of exceptionally high correlations among variables, except for the correlation score of 0.46 between loan amount ('logloan_amnt') and annual income ('logannual_inc').

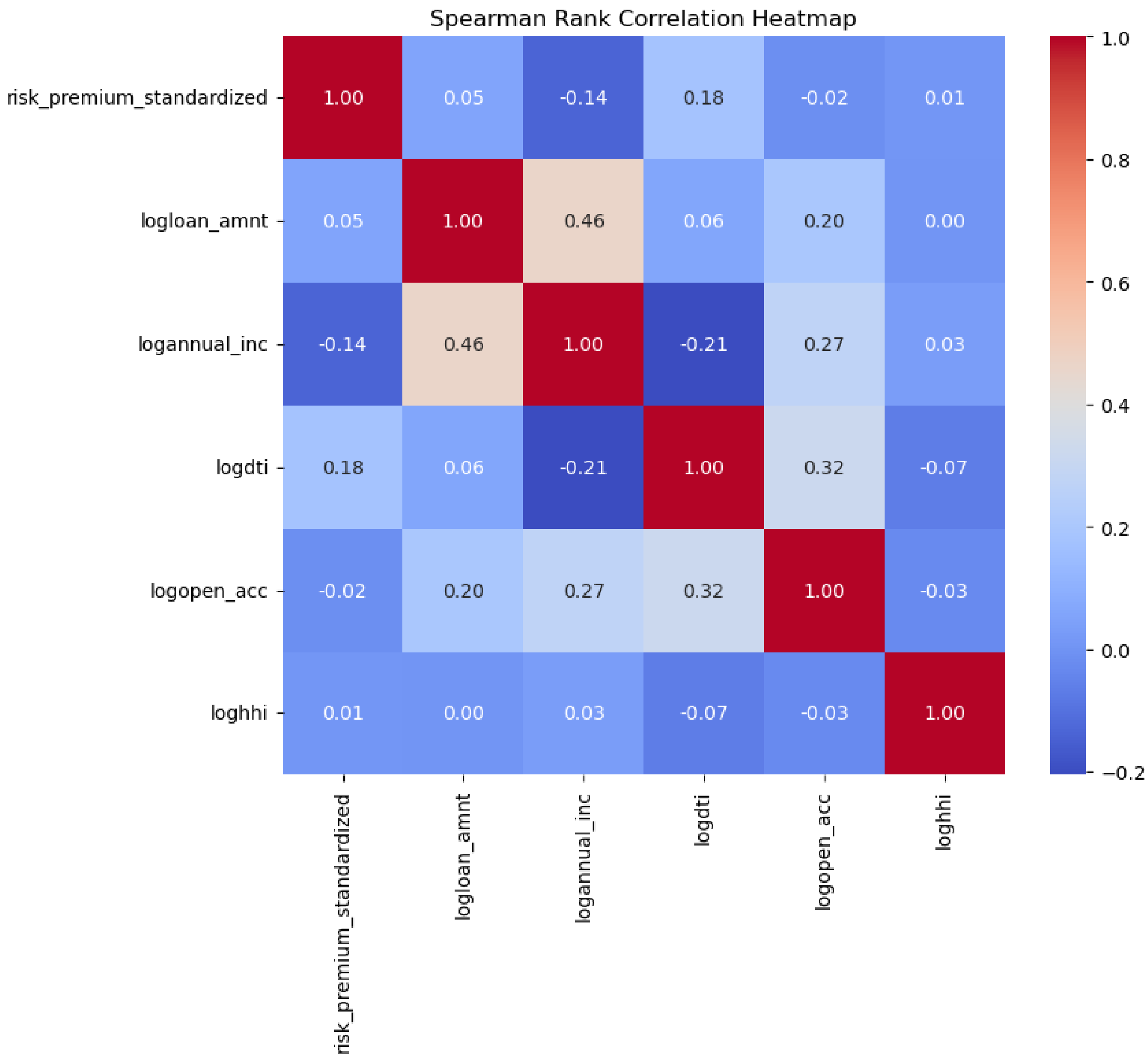

Figure 6: Spearman Rank Correlation Heatmap

Nonetheless, even this correlation is considered moderate. Despite the presence of this correlation, both variables can be included in the regression model, as their correlation does not exceed a threshold typically associated with multicollinearity concerns, which is usually around ±0.7.

Next, the point-biserial correlation coefficient was chosen to assess the strength and direction of association between binary and continuous variables. The highest correlation observable between continuous and binary variables in Figure 7 is found between term and risk premium. This correlation does not exceed 0.5, indicating relatively moderate levels of correlation.

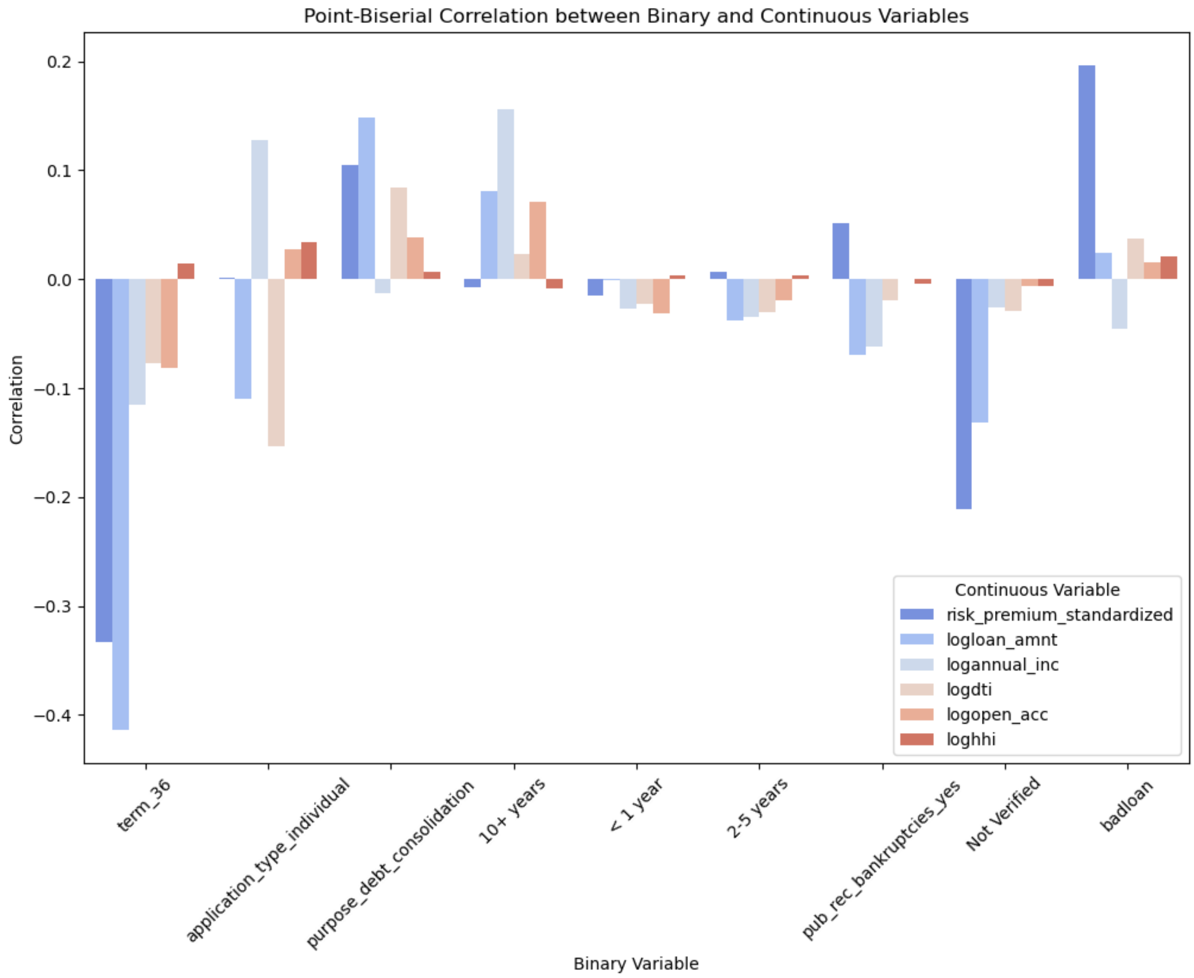

Figure 7: Point-Biserial Correlation

Lastly, Cramér's V, an extension of Pearson's chi-squared test, was employed to evaluate associations among categorical variables, as it provides a standardized measure of association suitable for categorical data. The correlation between the relevant categorical variables does not appear (Figure 8) to be notably high, except for a moderate correlation of

0.52 observed between two dummy variables representing different categories of employment length.

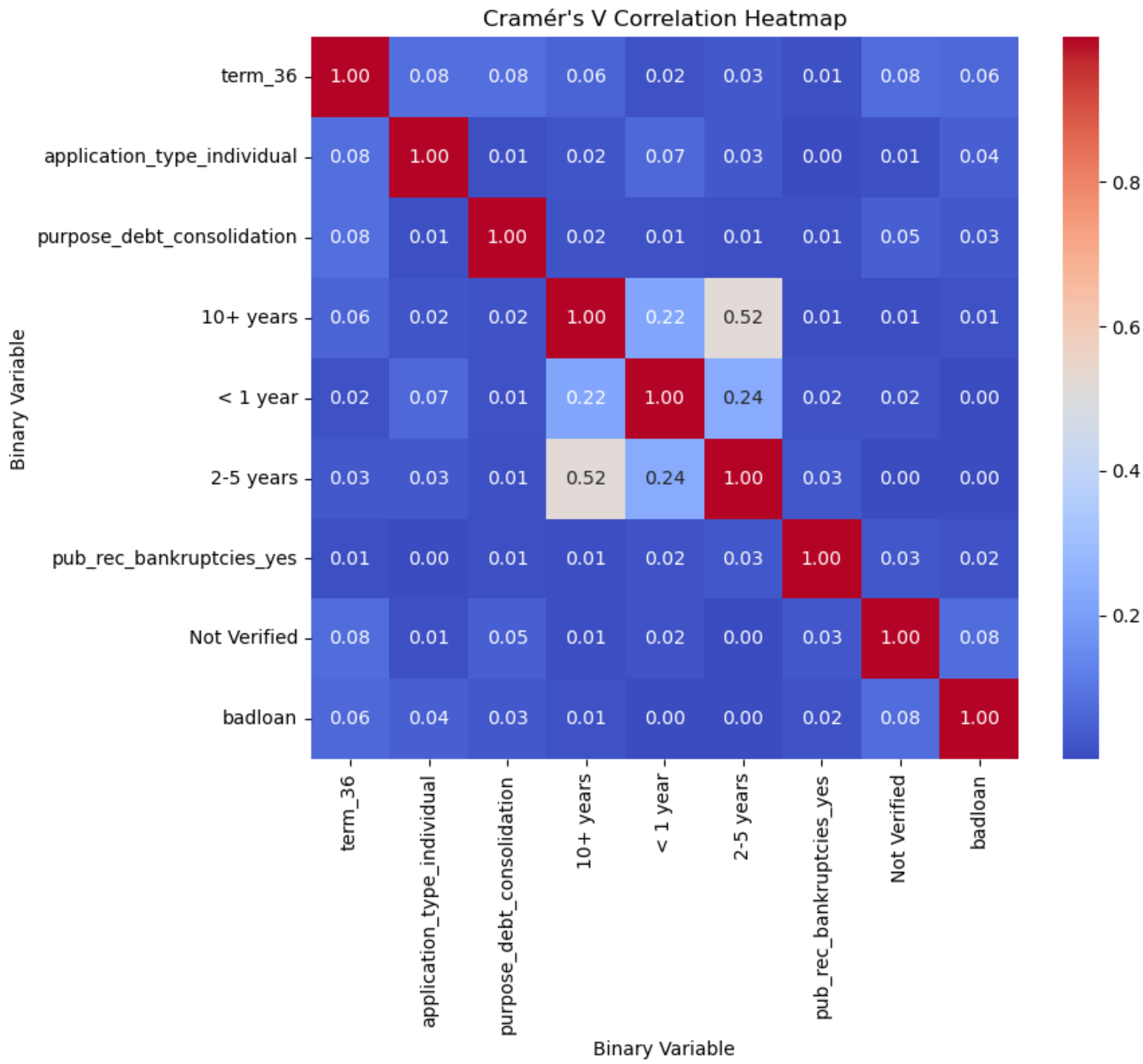

Figure 8: Cramer's V Correlation Heatmap

As previously noted, the analysis incorporates state dummy variables. While these variables did not exhibit a strong correlation with other predictor variables, there appears to be multicollinearity among certain states. This suggests that some states may be highly correlated with each other in terms of the predictor variables, which could potentially affect the reliability of the regression results.

4 Results

4.1 Ordinary Least Square Regression: Interest Rates & Competition

This regression analysis aims to understand how various factors, including bank competition measured by the Herfindahl-Hirschman Index, influence the risk premium offered by peer-to-peer lending platforms like LendingClub. The model, whose detailed information can be found in the appendix (Figure A1), exhibits a moderate level of explanatory power with an R-squared value of 0.24, suggesting that approximately 24% of the variation in the risk premium can be explained by the independent variables included in the model.

The coefficient for the standardized logged HHI ('loghhi') variable indicates a negative relationship between HHI and the risk premium. Specifically, for every one standard deviation increase in the logged HHI, the risk premium decreases by approximately 0.0008 standard deviations. This suggests that lower HHI, indicative of higher bank competition, is associated with a slightly higher risk premium offered by P2P lending platforms. In other words, as competition among banks increases, P2P lending platforms may raise their risk premium. This is in line with Yannelis and Zhang (2023), indicating that heightened competition typically prompts lenders to seek higher returns to compensate for increased risk given the adverse selection problem.

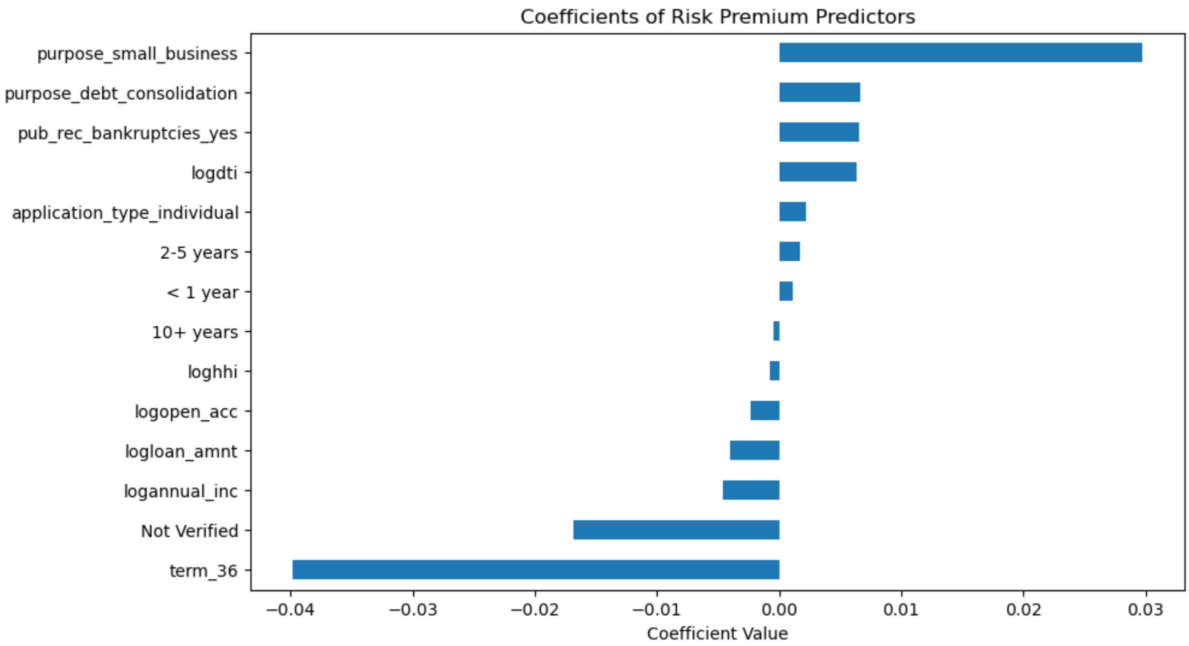

Figure 9: Coefficients of Risk Premium Predictors (OLS Regression)

Most coefficients (Figure 9) exhibit statistically significant relationships with the risk premium, with p-values close to zero. Notably, the coefficient for the HHI variable ('loghhi') is statistically significant at the 0.001 level. Other significant independent variables include term length, application type and loan purpose, with small business funding attributing to a higher risk premium.

The omnibus test yields a statistically significant result ($p < 0.05$), indicating that the model adequately explains the variance in the risk premium. Additionally, the Durbin-Watson statistic, with a value of approximately 1.979, suggests minimal autocorrelation of residuals. However, the large condition number (2.12e+03) raises concerns about potential multicollinearity, warranting further investigation. This multicollinearity arises from the accounting of state fixed effects. Including state fixed effects in the regression analysis is essential for controlling unobserved heterogeneity across states and addressing regional variation in economic conditions and regulatory frameworks. By incorporating state fixed effects, the analysis can more accurately estimate the effects of bank competition on the risk premium, ensuring the validity and reliability of the results. However, it should be kept in mind that this may influence the interpretability of coefficients.

State and year fixed effects were included in the analysis to control for macro-effects that could potentially affect prices or defaults. The state fixed effects account for time-invariant differences at the state level that may affect rates quoted by the LendingClub and also the quality of its clients. For example, exemptions in case of bankruptcy vary by state. Some states are more generous, i.e., the bankrupt borrower can keep more of its personal asset/belonging while going through bankruptcy. Also, there might be state level differences in setting up a debt reorganization plan. A simple way to summarize, this is to say that some states are more creditor and some states are more debtor-friendly. Lenders anticipate this and may adjust the rate accordingly, charging higher rates for a given borrower quality in more debtor-friendly states or lending less ("Debt.org - America's Debt Help Organization" 2018). Year fixed effects control for national-level time-varying economic conditions. According to Tang (2018), LendingClub asserts that its rates do not vary by state, indicating that local economic conditions do not affect interest rates. This suggests that while rates can change with national economic conditions, interaction terms between state and year variables are not necessary.

4.1.1 Residuals Plot

Figure 10 visualises the residuals from the OLS regression model against the fitted values, revealing several observations. Firstly, there is no distinct non-linear pattern, indicating that the linearity assumption is reasonably met. However, a noticeable downward shifting pattern is present, suggesting potential missing variables, including relevant predictors, interaction effects, or unobserved factors, which could bias the regression model predictions. There was an attempt at employing polynomial regression, as it allows for capturing nonlinear relationships between predictors and the response variable. Ridge regression was also ventured in order to mitigate multicollinearity and overfitting issues. However, the presence of the same downward shifting pattern in both cases indicates that the problem may be more complex, possibly involving omitted variables or unaccounted-for interaction effects.

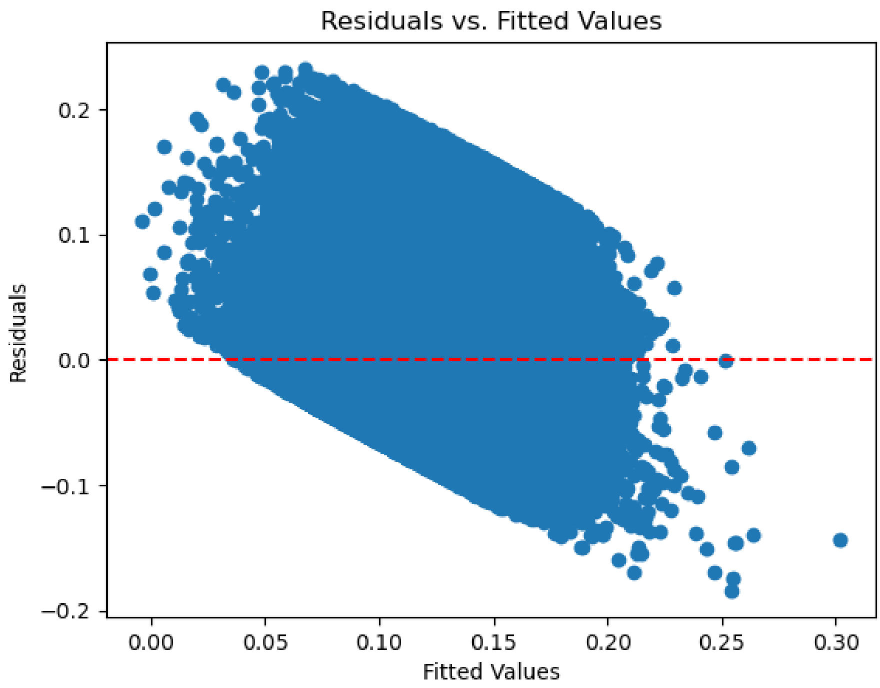

Figure 10: Plot of Residuals vs Fitted Values

There is no strong evidence of heteroscedasticity, as the spread of residuals remains relatively constant across the range of fitted values. The residuals are not perfectly symmetric around the horizontal axis, indicating some skewness. Additionally, a few outliers are present, particularly on the right side of the plot, which could influence the regression results.

4.1.2 Loan Sizes

In assessing the impact of loan size on the risk premium, the regression analysis was partitioned into different loan amount quartiles (Q1-Q4) to discern any variations in the effect of bank competition, represented by HHI, across these quartiles. The loan amount quartiles are defined based on interquartile ranges, with only the first ($1000-$8000) and fourth ($20150-$40000) quartiles exhibiting significant coefficients for the HHI. Notably, the coefficient for Q1 is larger and more negative than that of Q4, as observed in Figure 11. This implies that smaller loan sizes, corresponding to the first quartile, are more sensitive to changes in bank competition. Conversely, larger loan sizes in the fourth quartile appear to be less affected by variations in bank competition.

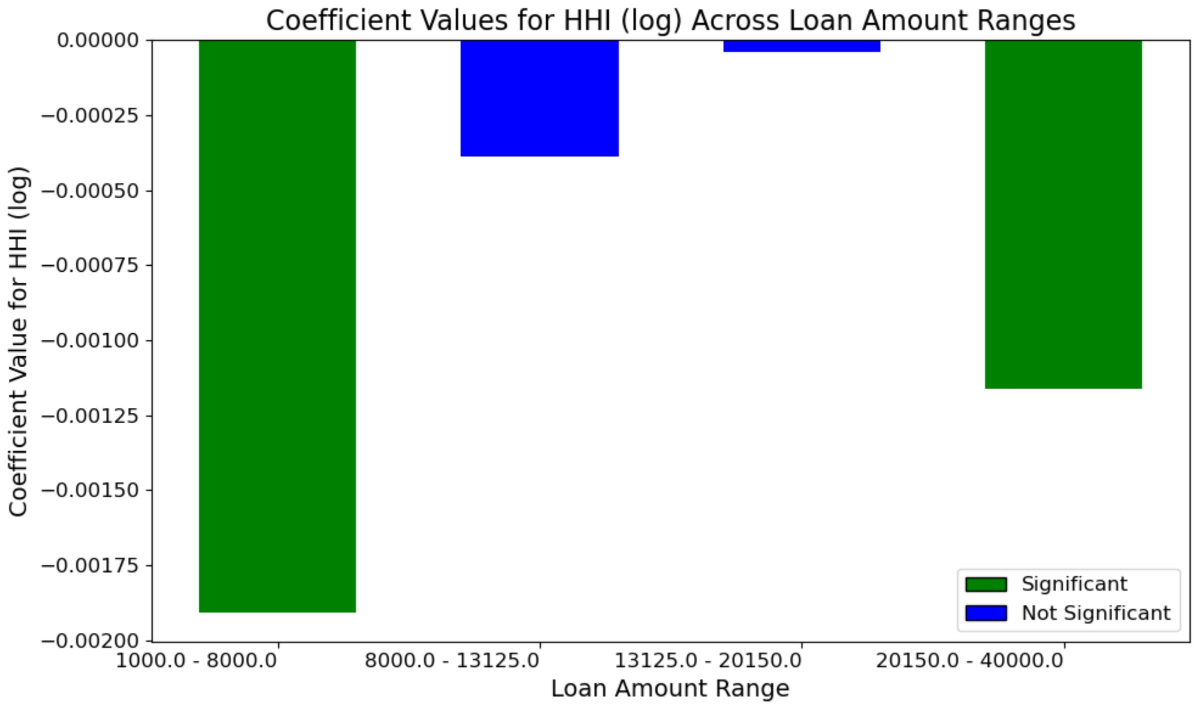

Figure 11: Coefficient Values for HHI Across Loan Amount Ranges (OLS Regression)

A reasoning for this could be that lower loan amounts are often associated with borrowers who may exhibit less financial stability or have less established credit histories. In a competitive market where adverse selection of borrowers is more prevalent, lenders have to take measures to account for the proportion of riskier individuals seeking smaller loans is higher. This scenario amplifies the importance of risk assessment, leading lenders to adjust their terms, in this case increase the loan margin, to mitigate the heightened risk associated with smaller loan amounts. Contrastingly, larger loans tend to attract more financially stable borrowers, who present lower default risks, allowing lenders to offer comparatively lower risk premiums.

4.2 Logit Regression: Default & Competition

The second regression (Figure A2) in question aims to understand how various factors, most importantly market competition, influence the default on loans given by peer-to-peer lending platforms like LendingClub. The target variable of this logit regression is ‘badloan’, with 0 being the value attributed to no default and 1 if the individual defaulted on the loan. The model exhibits a moderate level of explanatory power with a Pseudo R-sqaured value of 0.1154, suggesting that approximately 11.54% of the variation in default occurrence can be explained by the independent variables included in the model.

Market competition is a significant predictor of loan default, with a higher HHI (0.0776) associated with increased odds of bad loans. So for every one standard deviation increase in the logged HHI, the odds of a bad loan increase by approximately 8.1%, holding all other variables constant. Higher banking competition can result in more adverse selection if banks choose to invest less in screening. A way to account for this is to charge higher prices. This however also means that a potential competitor such as LendingCLub with a slightly better screening technology and/or slightly lower prices can attract better quality clients. This implies that there is a positive relationship between market competition, prices and LendingClub average quality client, measured by the fraction of loans that turn sour at a given point in time.

Borrower characteristics such as annual income, employment history, and loan characteristics such as size and purpose also influence loan default probabilities. Interestingly enough, individuals whose loan purpose is for a small business have a higher probability of default, this can be observed in Figure 12. Moreover, higher annual income and longer employment histories are associated with decreased odds of bad loans, while certain loan purposes are associated with increased odds of default.

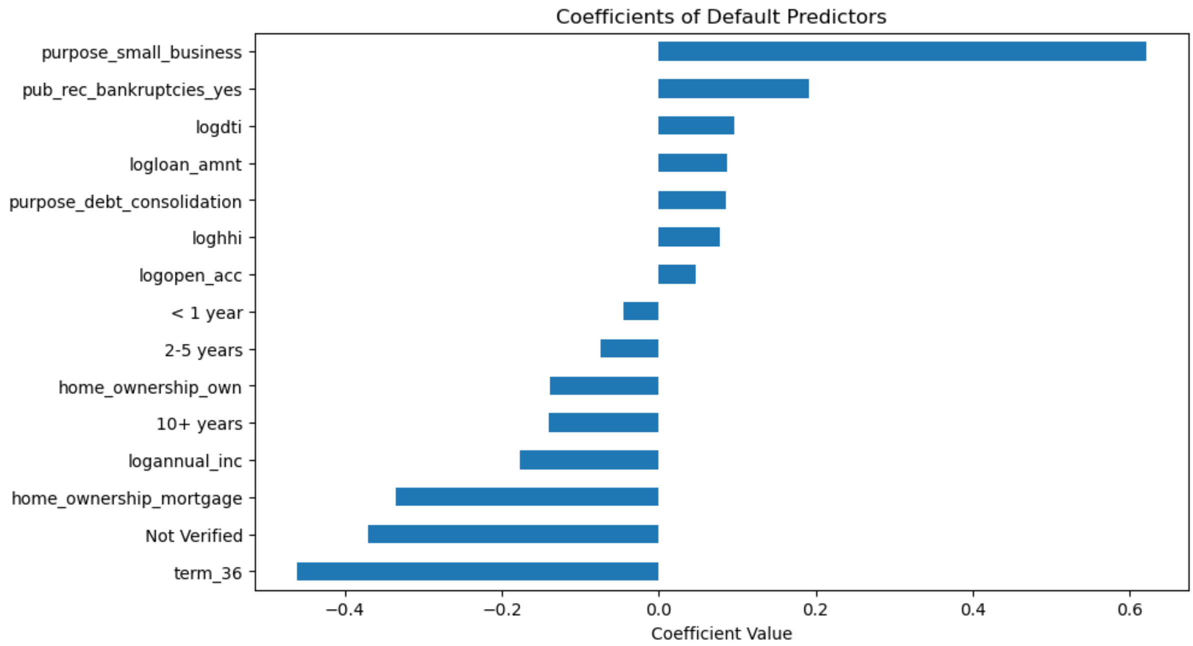

Figure 12: Coefficients of Default Predictors (Logit Regression)

In summary, higher market competition is found to correlate with increased default probabilities. Additionally, borrower characteristics and loan purposes also influence default likelihoods.

5 Reflection

5.1 Competition, Loan Prices and Default

Similarly to Petersen and Rajan (1994), where interest rates charged to young and credit-constrained firms are higher in more competitive markets, the regression analysis suggests that LendingClub charges slightly higher interest rates with an increase in bank competition. This analysis suggests the following: increased competition in the banking sector can lead traditional banks to screen less. In order to protect themselves from customer default, they charge higher rates. When traditional banks opt to charge higher interest rates to compensate for the increased risk in adverse selection, LendingClub can undercut them by offering competitive rates while still safeguarding their profitability. P2P lending platforms can leverage their advanced screening technology and access to alternative data sources not typically used or available to traditional banks (Frost 2020), allowing them to maintain a good risk assessment process. The plausibility of this explanation is further strengthened by the negative and significant effect of market competition on loan defaults, with a lower HHI (higher competition) associated with decreased odds of bad loans. This means that even in markets with heightened competition, LendingClub can uphold good screening standards, ensuring that they attract borrowers with better creditworthiness. This ability to offer attractive interest rates, in addition to their good screening capabilities, gives LendingClub a competitive edge in the lending landscape.

Petersen and Rajan's study (1994) suggests that smaller loan amounts are more influenced by competitive credit market dynamics. Smaller loans face higher costs per unit lent to the extent that some of the costs associated with providing them are fixed (for example, screening costs). Furthermore, loan size may also correlate with unobservable borrower quality, implying lower borrower quality for smaller loans. As a consequence, the effect of competition on the pricing of these loans might be even more pronounced. This can also be observed in the LendingClub data for similar reasons.

5.2 Data Limitations

One of the main limitations of the data was that the dataset only provided information on successful loan applications, omitting rejected loan requests. Thus, it was not possible to study how competition by banks affected the probability of getting a loan from LendingClub. Thus, the current analysis cannot answer whether the interaction between LendingClub and banks resulted in better credit availability for borrowers and whether this effect varied with the level of bank competition. While it is not possible to directly observe clients whose loan requests were rejected, this thesis was able to analyse successful loan requests and examine how competition influences the pricing and quality of clients selected by LendingClub.

Initially, the aim was to investigate whether LendingClub covered similar (overlapping) or complementary credit market segments to banks. However, due to the unavailability of banking data, this approach proved challenging. An alternative method, akin to Tang's study, would have involved taking advantage of the time window of the data and observing how a policy shock, that would artificially constrain bank lending, such as the implementation of the FAS 166/167 regulation in 2010 (Tang 2018), thus allowing us to discern whether fintech platforms are substituting for reduced bank lending. Nevertheless, between 2012 and 2019, no such policy shock occurred.

Furthermore, I only could use a limited set of variables to capture borrower characteristics that could have affected loan rates quoted by LendingClub for a particular borrower. For example, I was unable to collect the FICO score, included in Tang's paper. This could have been a good control variable for borrower quality and would have explained more of the variations in loan rates.

Lastly, as mentioned earlier, the analysis relies on HHI index values of only two years, corresponding to the years 2011 and 2015. This limited scope may introduce potential accuracy concerns, particularly regarding the representation of market concentration dynamics over the intermediate years.

5.3 Analysis Limitations

Regarding analysis limitations, there appears to be multicollinearity among the state dummy variables, as indicated by a condition number of 2.12e+03. These state dummy variables are used in the analysis to capture differences in institutional factors, such as the stringency of bankruptcy laws. This multicollinearity could be attributed to certain states having similar characteristics, resulting in highly correlated dummy variables.

Secondly, as previously mentioned in the history of LendingClub, the platform, like other fintech companies, increased interest rates in 2016 to attract investors. This change in business strategy is not controlled for in the quantitative analysis. Therefore, it should be noted that some of the variance in the marginal interest rate could be attributed to this factor rather than to the HHI.

Lastly, the second part of the research question posed in this thesis concerns how market competition influences borrower quality. There is no reliable ex-ante measure of a borrower's riskiness; given the data, we can only observe certain characteristics that correlate with borrower quality. The only way to measure this quality in this analysis is by observing the default rates of borrowers to whom LendingClub has lent. However, defaulting does not always directly reflect borrower quality, as it is an ex-post measure. While creditworthiness significantly affects whether a borrower defaults, other factors such as sudden death, illness in the family, or natural disasters can also lead to default, irrespective of the borrower's initial riskiness. Since defaults can occur due to factors unrelated to the borrower's initial riskiness, using defaults as a proxy for borrower quality may not accurately capture the true effect of market competition on borrower quality.

Conclusion

This thesis attempts to answer the question of how bank competition affects the prices and markets served by peer-to-peer lending platforms, aiming to understand the dynamics of bank-fintech interaction and make possible speculations on the competitive role of fintech in the financial landscape.

By analysing variables such as risk premiums, bad loans, and the Herfindahl-Hirschman Index, the study aims to uncover the dynamics between competition, pricing strategies, and loan defaults. The quantitative analysis is performed on data from LendingClub, once the dominating P2P lending platform in the United States. The regression analysis performed in this thesis indicate that higher market competition, measured by the Herfindahl-Hirschman Index, is associated with slightly higher risk premiums. This aligns with previous findings (Yannelis and Zhang 2023) that, due to a riskier pool of borrowers resulting from adverse selection in competitive markets, higher interest rates are set to mitigate this issue. Additionally, this is more relevant for smaller loan sizes, which are usually attributed to less creditworthy borrowers. Smaller loan sizes, similarly to Petersen and Rajan's (1994) findings in the bank lending market for small companies, are more sensitive to changes in bank competition, resulting in higher interest rates. Lastly, the data suggests that higher market competition correlates with decreased default probabilities. This could be specific to fintech companies, highlighting their competitive edge in advanced screening technology and alternative data sources, allowing them to continue a strong risk assessment process even in competitive markets.

While this study provides valuable insights into the impact of bank competition on peer-to-peer lending platforms, several data and analytical limitations should be noted. The absence of information on rejected loan requests limits the understanding of how bank competition influences loan access to alternative sources of credit such as fintech lending. Furthermore, the exclusion of FICO scores as a predictor of borrower quality and the reliance on HHI index values from just two years raise concerns about the robustness of the analysis. Analytical challenges, such as multicollinearity among state dummy variables and the omission of the increase in interest rates by fintech platforms in 2016 should also be taken into account. Lastly, using loan defaults as a proxy for borrower quality may oversimplify the complex relationship between market competition and borrower riskiness. Addressing these limitations in future research will be crucial for advancing our understanding of the dynamics between bank competition and peer-to-peer lending platforms.

Further research could focus on the interaction between market competition and loan access if data for all loan requests is available, allowing for a better analysis of the target market of such lending platforms. Additionally, exploring the impact of fintech advancements in screening technologies would provide deeper insights into the competitive edge of fintech companies.

Appendix

Variable Descriptions

Table A 1: Description of Original Lending Club Variables

LendingClub Variable	Description
int_rate	Interest rate on the loan.
loan_amnt	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
annual_income	The self-reported annual income provided by the borrower during registration.
verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified.
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
pub_rec_bankruptcy	Number of public record bankruptcies
open_acc	The number of open credit lines in the borrower's credit file.
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers.
state	state in which borrower resides.
emp_length	Employment length in years.
home_ownership	The home ownership status provided by the borrower during registration or obtained from the credit report. Our values are: RENT, OWN, MORTGAGE, OTHER.
purpose	A category provided by the borrower for the loan request.
issue_year	The year which the loan was funded.
issue_date	The month the loan was funded.
badloan	Indicates whether the borrower has defaulted on the loan or not.

The descriptions above have been obtained from the LendingClub dataset (Nigmonov 2021)

Table A 2: Description of LendingClub Dummy Variables

Dummy Variable	Description

term_36	Binary variable indicating whether the loan duration is 36 months.
purpose_debt_consolidation	Binary variable indicating if the purpose of the loan is debt consolidation.
purpose_small_business	Binary variable indicating if the purpose of the loan is small business funding.
home_ownership_mortgage	Binary variable indicating if the borrower has a mortgage.
home_ownership_own	Binary variable indicating if the borrower owns a home.
10+ years	Binary variable indicating if the borrower's employment length exceeds 10 years.
< 1 year	Binary variable indicating if the borrower's employment length is less than one year.
2-5 years	Binary variable indicating if the borrower's employment length is between 2 and 5 years.
pub_rec_bankruptcies_yes	Binary variable indicating if the borrower has a public record of bankruptcies.
Not Verified	Binary variable indicating if the borrower's annual income has not been verified.

Table A 3: Description of Other Variables

Variable	**Description**
hhi	Herfindahl-Hirschman Index, a measure of market concentration to assess the level of competition within the industry.
int_rate_fed	Federal funds effective rate, typically referring to a benchmark interest rate set by the central bank.
risk_premium	The difference between the loan's interest rate and the federal interest rate (int_rate - int_rate_fred), representing the additional return expected by lenders for taking on the risk of lending to borrowers.

Regression Tables

Figure A 1: OLS Regression Results (State and Year Coefficients Omitted)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           risk_premium   R-squared:                       0.240
Model:                            OLS   Adj. R-squared:                  0.240
Method:                 Least Squares   F-statistic:                 1.205e+04
Date:                Sun, 26 May 2024   Prob (F-statistic):               0.00
Time:                        17:38:50   Log-Likelihood:             4.6432e+06
No. Observations:             2703430   AIC:                        -9.286e+06
Df Residuals:                 2703358   BIC:                        -9.285e+06
Df Model:                          71
Covariance Type:            nonrobust
==============================================================================
                                  coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
------------------------
const                           0.1319      0.001    217.092      0.000
0.131       0.133
term_36                        -0.0398    6.4e-05   -621.173      0.000
-0.040      -0.040
application_type_individual     0.0023      0.000     20.549      0.000
0.002       0.002
purpose_debt_consolidation      0.0067   5.46e-05    121.900      0.000
0.007       0.007
purpose_small_business          0.0297      0.000    108.371      0.000
0.029       0.030
10+ years                      -0.0005   7.34e-05     -6.156      0.000
-0.001      -0.000
< 1 year                        0.0012      0.000     11.206      0.000
0.001       0.001
2-5 years                       0.0017   7.14e-05     23.504      0.000
0.002       0.002
logannual_inc                  -0.0046   3.35e-05   -137.028      0.000
-0.005      -0.005
logdti                          0.0063    3.1e-05    204.186      0.000
0.006       0.006
pub_rec_bankruptcies_yes        0.0066   8.23e-05     80.209      0.000
0.006       0.007
Not Verified                   -0.0168   5.64e-05   -297.969      0.000
-0.017      -0.017
logopen_acc                    -0.0023   3.07e-05    -76.389      0.000
-0.002      -0.002
logloan_amnt                   -0.0040   3.34e-05   -120.635      0.000
-0.004      -0.004
loghhi                         -0.0008      0.000     -3.420      0.001
-0.001      -0.000
=============================================================================
Omnibus:                   298877.990   Durbin-Watson:                   1.979
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           424369.856
Skew:                           0.863   Prob(JB):                         0.00
Kurtosis:                       3.889   Cond. No.                     2.12e+03
==============================================================================
```

Figure A 2: Logit Regression (State and Year Coefficients Omitted)

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                badloan   No. Observations:              2703430
Model:                          Logit   Df Residuals:                  2703357
Method:                           MLE   Df Model:                           72
Date:                Sun, 26 May 2024   Pseudo R-squ.:                  0.1154
Time:                        17:40:03   Log-Likelihood:            -6.6743e+05
converged:                       True   LL-Null:                   -7.5448e+05
Covariance Type:            nonrobust   LLR p-value:                     0.000
==============================================================================
                                 coef    std err          z      P>|z|
[0.025      0.975]
------------------------------------------------------------------------------
const                         -3.2323      0.056    -57.931      0.000
-3.342      -3.123
logannual_inc                 -0.1764      0.003    -56.384      0.000
-0.183      -0.170
logloan_amnt                   0.0867      0.003     26.754      0.000
0.080       0.093
term_36                       -0.4601      0.005    -84.810      0.000
-0.471      -0.450
purpose_debt_consolidation     0.0857      0.005     17.522      0.000
0.076       0.095
purpose_small_business         0.6213      0.020     30.461      0.000
0.581       0.661
home_ownership_mortgage       -0.3355      0.005    -62.219      0.000
-0.346      -0.325
home_ownership_own            -0.1378      0.008    -17.440      0.000
-0.153      -0.122
10+ years                     -0.1404      0.006    -22.044      0.000
-0.153      -0.128
< 1 year                      -0.0446      0.009     -4.866      0.000
-0.063      -0.027
2-5 years                     -0.0737      0.006    -11.995      0.000
-0.086      -0.062
logdti                         0.0971      0.003     32.578      0.000
0.091       0.103
pub_rec_bankruptcies_yes       0.1917      0.007     27.852      0.000
0.178       0.205
Not Verified                  -0.3704      0.005    -67.409      0.000
-0.381      -0.360
logopen_acc                    0.0478      0.003     17.230      0.000
0.042       0.053
loghhi                         0.0776      0.018      4.209      0.000
0.041       0.114
==============================================================================
```

Bibliography

Balyuk, Tetyana, and Sergei Davydenko. "Reintermediation in FinTech: Evidence from Online Lending." *Journal of Financial and Quantitative Analysis*, June 20, 2023, 1–41. https://doi.org/10.1017/S0022109023000789.

Beaumont, Paul, Huan Tang, and Eric Vansteenberghe. "Collateral Effects: The Role of FinTech in Small Business Lending." *SSRN Electronic Journal*, 2022. https://doi.org/10.2139/ssrn.4260842.

Debt.org. "Debt.Org - America's Debt Help Organization." Accessed May 26, 2024. https://www.debt.org/.

Degryse, Hans, and Steven Ongena. "Distance, Lending Relationships, and Competition." *The Journal of Finance*, 2005.

FDIC. "Pro Forma (HHI) Report Provided." FDIC, 2011. https://www7.fdic.gov/sod/sodMarketBank.asp?barItem=2

FDIC. "Pro Forma (HHI) Report Provided." FDIC, 2015. https://www7.fdic.gov/sod/sodMarketBank.asp?barItem=2

FRED. "Federal Funds Effective Rate," 2024. https://fred.stlouisfed.org/series/FEDFUNDS.

Frost, Jon. "The Economic Forces Driving FinTech Adoption across Countries." *SSRN Electronic Journal*, 2020. https://doi.org/10.2139/ssrn.3515326.

FSB. "FinTech and Market Structure in Financial Services: Market Developments and Potential Financial Stability Implications," 2019. https://data.mendeley.com/datasets/wb3ndt69gf/3

Gopal, Manasa, and Philipp Schnabl. "The Rise of Finance Companies and FinTech Lenders in Small Business Lending." *The Review of Financial Studies* 35, no. 11 (November 1, 2022): 4859–4901. https://doi.org/10.1093/rfs/hhac034.

Petersen, Mitchell A., and Raghuram G. Rajan. "The Benefits of Lending Relationships: Evidence from Small Business Data." *The Journal of Finance* 49, no. 1 (1994): 3–37. https://doi.org/10.1111/j.1540-6261.1994.tb04418.x.

Nigmonov, Asror. "Dataset from the US Peer-to-Peer Lending Platform with Macroeconomic Variables." Mendeley Data, November 29, 2021. https://doi.org/DOI:10.17632/wb3ndt69gf.3.

Tang, Huan. "Peer-to-Peer Lenders Versus Banks: Substitutes or Complements?" *The Review of Financial Studies* 32, no. 5 (May 1, 2019): 1900–1938. https://doi.org/10.1093/rfs/hhy137.

Yannelis, Constantine, and Anthony Lee Zhang. "Competition and Selection in Credit Markets," Journal of Financial Economics 150, Issue 2, (November 2023): 103710 . https://doi.org/10.1016/j.jfineco.2023.103710.