UNIVERSITY OF GOTHENBURG

SCHOOL OF BUSINESS, ECONOMICS AND LAW

Graduate School

Master of Science in Finance

Credit losses in peer-to-peer lending: a quantile regression approach

Author

Alexander Campus

Supervisor

Adam Farago

Academic Year 2023/2024

Acknowledgement

I would like to express my deepest gratitude and appreciation to the University of Gothenburg and the University of Rome “Tor Vergata” for giving me the opportunity to pursue a dual degree in Rome which has been an incredible experience.

With kind regards,

Alexander Campus

Table of Contents

Figures and Tables

Abstract

Peer-to-peer (P2P) lending is a form of direct lending that connects borrowers with investors through online platforms. This study analyses factors that determine credit losses in peer-to-peer lending with a dataset consisting of 253 948 charged-off loans from the Lending Club. Most of the previous quantitative studies have focused on understanding and predicting defaults which has left a gap in the literature in understanding default amounts.

This study aims to deepen the understanding of the relatively unexplored area of credit losses in P2P lending. Through quantile regression, we have examined the relationship between borrower characteristics and credit losses to answer the question "What are the determinants of credit losses in P2P lending".

The results show that loan terms, application type, and interest rate are among the most important determinants of credit losses. Employment length is an important determinant in the lower quantiles, but the effect becomes less significant compared to other covariates in higher quantiles. Lastly, loan purpose is shown to be among the top determinants and becomes more important toward the higher quantiles.

By identifying the determinants of credit losses, we offer valuable insights for investors, P2P platforms, as well as policymakers in assessing credit risk. This study contributes to a deeper understanding of credit losses in P2P lending and we encourage future research to use this paper as a theoretical framework to go even deeper into credit losses. For example, a study that uses extremal quantile regression to analyze the tails of the distribution.

Keywords: peer-to-peer, P2P, credit losses, Lending Club, quantile regression

Chapter I: Introduction

1.1 Introduction

In recent years, the global economic landscape has been marked by challenges stemming from rising inflation rates and interest rate hikes starting in 2022. This financial strain is particularly evident in the mortgage market, as shown in a report published by the Swedish financial supervisory authority “Finansinspektionen”, on March 7, 2024. The report states that the median surplus for mortgage borrowers in Sweden has decreased from 8,900 SEK in 2022 to approximately 7,700 SEK in 2023. This number is not inflation-adjusted, meaning that the actual purchasing power has decreased even more. Moreover, the proportion of borrowers with a “low surplus” has increased from 15% to 19% during the same period. Where low surplus is defined as having less than 2,000 SEK, or roughly 170 EUR, in surplus. This indicates a deteriorating financial situation for many households (Finansinspeltionen, 2024).

During the same period the major Swedish banks, consisting of Swedbank, Handelsbanken, SEB, and Nordea, have reported record-high profits for the year 2023. Despite a 15% increase in costs, the combined EBIT for these banks surpassed 200 billion SEK, setting a new record. This was primarily attributed to the net interest, which is the difference between interest paid on deposits and gained on loans, increasing by a remarkable 42% (Molenius, 2024). The reasons for this huge increase lie beyond the scope of this study, but the increased profits have drawn criticism. Mikael Damberg, former Minister of Finance in Sweden, characterizes them as “provocative” (SVT, 2023).

However, there exists an alternative that cuts out traditional financial intermediaries such as banks, namely peer-to-peer (P2P) lending. P2P lending is a form of direct lending that connects borrowers with investors through online platforms (Kagan, 2024). It offers a diverse range of loan options including personal loans, small business loans, mortgages, and more. This form of lending has gained popularity due to the potential high returns for investors and potential lower interest rates for borrowers (CFI, 2022).

1.2 Research background

1.2.1 History and future of P2P lending

P2P lending also known as, social lending, crowd lending, or marketplace lending, is the process of obtaining financing from other individuals instead of a financial intermediary. A similar phenomenon has been around since the 18th century when local non-notarized credit makers lent money to customers, way before banks came around. These borrowers used debt to smooth consumption or to make investments, for example buying livestock (Dermineur 2019). But modern P2P lending began in 2005 with the launch of Zopa in the UK which was the first P2P platform in the world and soon after Prosper and Lending Club launched in the US 2006 and 2007 respectively (Bholat and Atz 2016).

Despite initial challenges, with limited investors and a lot of highly uncreditworthy loan requests, the P2P lending sector underwent a significant transformation during the financial crisis of 2008 - 2009. Public trust in financial institutions eroded and banks faced existential challenges which disrupted the credit supply. Suddenly individuals and businesses encountered obstacles in securing loans, with small and medium-sized enterprises (SMEs) being particularly hard hit. Investors began to shy away from traditional banking products, seeking alternative avenues for returns. The borrowers and investors found P2P lending and the platforms emerged as winners, experiencing a surge in lenders willing to embrace risk, while also witnessing an expansion in the pool of creditworthy borrowers (Bednorz, 2020).

The success of these early platforms catalyzed a wave of new entrants into the market. Rate Setter and Funding Circle launched in 2010 and subsequently, hundreds of others followed suit, marking a proliferation of P2P lending (Bednorz, 2020). The industry saw significant growth and during 2018 Lending Club had $10,9 billion in record loan originations, up 21% year-over-year (YoY), and an application growth of 35% (LendingClub, 2019). The 2020 global alternative finance market benchmarking report showed that developing countries experienced substantial growth between 2017 to 2018. YoY volumes went up 102% in Africa, and 573% in South East Asia (Ziegler et al., 2021).

China had outstanding growth within P2P lending. In 2014 the outstanding P2P loan balance was 100 billion RMB and in 2018 this number had exploded to 1,3 trillion RMB or an estimated $ 120-150 billion in the US. However due to several rogue businesses, exemplified by Ezubao which turned out to be a Ponzi scheme, China's regulators imposed increasingly strict regulations. Investors became more aware of the risks involved and in the first half of 2018, 300 P2P platforms went out of business in China (Mckinsey.com, n.d.).

Combine this with the fact that the pioneers, Zopa and Lending Club, no longer offer P2P lending and you might think the future looks dim for the industry. Nevertheless, the global P2P lending market was valued at $153 billion in 2022 with optimistic forecasts to reach $804,2 billion by 2030 or $1,7 trillion by 2032 depending on the research company (Allied Market Research, 2023; Acumen Research and Consulting, 2023).

The reason for Zopa and Lending Club quitting P2P lending is that they are now operating as banks. After Lending Club acquired Radius Bancorp in 2020, regulators had the company set aside capital against P2P loans even though the risk exposure was passed on to investors. This made P2P uneconomical and Lending Club soon after closed their P2P platform (Proud, 2021).

1.2.2 The loan process

Although the loan process might look different across platforms a typical P2P loan begins with the borrower submitting an application where they answer questions regarding the loan purpose and the amount. The Lending Club gave each borrower a loan grade which reflected the borrowers' risk, each loan grade was divided into five subgrades which determined the interest rate for the loan. The grade was set based on the credit report of the borrower, which includes things like FICO score and debt-to-income ratio, together with other factors such as loan amount and loan term (Fintech Nexus, n.d.). Platforms often evaluate a wide range of factors such as income, credit score, and employment history, when assessing the creditworthiness of a borrower. This form of lending offers an

opportunity for individuals with limited or less favorable credit histories to obtain funding, which might otherwise be challenging through a bank. Then there are the investors, which can be individuals or institutions who seek alternative investment opportunities outside the traditional stocks, bonds, or real estate markets. Investors have the flexibility to select loans to invest in based on their own criteria and spread their investments across multiple loans to diversify. This process is facilitated by investors' ability to fund partial portions of loans, enhancing accessibility and diversification within the P2P lending landscape (FasterCapital, n.d.).

Once a loan secures sufficient funding from investors through the platform, the borrower receives the loan. The platform efficiently manages the repayment process by collecting payments from the borrower and disbursing them among investors. These payments are structured as monthly installments, covering both the principal amount borrowed and interest. Due to the reduced overhead costs of P2P lending the interest rates can be lower compared to other types of unsecured lending, particularly for those with lower credit ratings. Investors receive their returns through interest, allocated proportionally based on their investment amount (FasterCapital, n.d.). If a loan defaulted, the Lending Club charged off the loan when they no longer expected further payments, usually within 30 days after entering default. After which, it was common that the debt was sold to debt collection agencies to recover funds and agree on settlements (Resolve, 2019; Meltzer, n.d.).

1.2.3 Advantages and risks of P2P lending

P2P lending has transformed the financial landscape by providing borrowers with an alternative avenue to access funds that cut out traditional intermediaries. Eliminating the middleman introduces a range of advantages and disadvantages which will be explored in this section.

P2P lending has an advantage over traditional intermediaries by having an almost automatic lending process which results in lower operating costs. This type of lending also allows to connect investors and borrowers across countries which might be difficult for a

bank (Nowak, Ross and Yencha, 2017). P2P lending typically has lower rates compared to other forms of unsecured personal bank loans. However, surveys of P2P borrowers show that this was not the main benefit, but rather the convenience in the form of lower transaction costs and a much shorter time until the loan is received. Corporate borrowers were particularly fond of not providing collateral and the flexibility of prematurely canceling loan contracts without a prepayment penalty fee (Lenz, 2016). Lastly, this type of lending allows individuals with an unfavorable credit history the opportunity to access funds (Chen et al., 2019). This might be particularly useful for debt consolidation purposes.

Though P2P lending has a set of advantages it also comes with a wide range of risks. The FDIC insurance only covers deposit products, so an American P2P loan is not protected by the FDIC, but the platform might have some legal resources (Fdic.gov, n.d.). Then there are regulatory and platform risks which were exemplified by the first half of 2018, when 300 P2P platforms went out of business in China (Mckinsey.com, n.d.). Unlike other fixed-income investments like bonds, which have an established secondary market, P2P loans are highly illiquid. Moreover, this form of lending attracts low-credit-score individuals which increases the risk of defaults (Setiawan, Suharjito and Diana, 2019). A study of the Chinese P2P Lending industry showed an average default rate of 6,45% which is much higher than in traditional banks (Tang et al., 2020). Moreover, Setiawan, Suharjito, and Diana (2019) describe that even though Lending Club had credit rating grades to help investors manage risk there were still problems with information asymmetry. They further explain that the individuals with the highest rating were not necessarily good borrowers which might suggest that traditional scoring metrics are not able to capture the dynamics in P2P lending (Setiawan, Suharjito and Diana, 2019).

1.3 Study objectives and methodology

Because most of the quantitative research in P2P lending has focused on defaults and funding success a research gap has been identified. A default does not distinguish between a borrower who defaults on a small amount or the entire loan. This leaves a gap in

understanding what drives bigger defaults. This study aims to deepen the understanding of the relatively unexplored area of credit losses in P2P lending and can serve as a foundation for future research. The objective is to analyze the relationship between borrower characteristics and credit losses. The study adopts a quantitative approach and uses *R* as the programming language. To give a comprehensive view of the credit losses in P2P lending this study uses quantile regression. Regarding the scope, this study uses a dataset, consisting of 253 948 charged-off loans, from Lending Club which means that the research is geographically limited to the US. Moreover, the dataset contains loans from 2007 to the third quarter of 2018 making it limited to this period.

1.4 Innovation points of research

This research addresses a gap in the existing literature by shifting the focus from predicting defaults to understanding the determinants of credit losses in P2P lending. By analyzing credit losses across quantiles, this study provides a comprehensive understanding of factors influencing larger defaults. The methodology of this research, quantile regression, has been used within the P2P sphere, but it is very uncommon. However, no other study has used quantile regression on credit losses.

The findings of this research are expected to contribute to the academic understanding of credit risk in P2P lending. By identifying the determinants of credit losses across quantiles, this study will offer valuable insights for investors, platforms, and policymakers in assessing and managing credit risk in P2P lending platforms. Since the literature on this subject is very scarce, this study can be used as a theoretical foundation for future research on credit losses in P2P lending.

The rest of this study is structured as follows. Chapter 2 provides a review of the previous literature and ends with our research question. Chapter 3 describes the methodology, data, and data-cleaning process. Chapter 4 presents the results and chapter 5 discusses the results and how the objective was met. Lastly, chapter 6 concludes the study, answers the research question, and gives a discussion on limitations and possible future research.

Chapter 2: Literature Review

2.1 Literature Review

Online P2P lending provides individuals and small businesses with alternative lending options to traditional banks. Because of the advantages connected to P2P lending the number of platforms and lending volume has seen significant growth (Milne and Parboteeah 2016). This has led to several empirical studies examining the phenomenon. Basha, Elgammal, and Abuzayed (2021) reviewed 198 papers on P2P lending, published between 2008-2020, and found that the research is geographically skewed to the USA and China. A popular topic to examine is regulation, with 20% of the reviewed literature discussing lending policy. Other common topics are determinants of funding success when applying for a loan and predicting loan defaults. 33% of the reviewed literature involved different regression models and a binary outcome (Basha, Elgammal and Abuzayed, 2021).

Earlier studies were more likely to use statistical methods to examine determinants of defaults and funding success, for example, Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios (2015); while more recent papers have shifted toward using machine learning to predict defaults (Aleksandrova, 2021; Xia et al., 2019).

Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios (2015) used Lending Club data from 2008 to 2014 with a sample size of 24 449 observations and studied the determinants of defaults. The study only used loans with 36 months duration since only a few 60-month loans had matured at the time. They had a subset of variables and used univariate means tests, survival analysis, and Cox regression for variable selection. The selected variables were then used in logistic regression to predict defaults. Their paper described that the grade assigned by the Lending Club decides the interest rate and found that the interest rate and loan purpose are the most important variables in predicting defaults. Other relevant variables include annual income, home ownership, credit history, and borrower

indebtedness. However, loan amount and employment length did not show statistical significance.

Michal Polena and Tobias Regner (2019) made a similar study and used Lending Club loan data from 2009 to 2012 with a 36-month duration. Their total sample size was 70 673. Backward stepwise elimination was used on a subset of 13 variables and logistic regression was used to examine determinants of defaults. Moreover, they used the credit grade to split the data into low-risk class, medium-risk class, high-risk class, very high-risk class, and all classes to see how the covariates affect borrowers with different credit grades. They found that covariates differed in both coefficient size and significance between classes. Some of the findings in the "all classes" case were that having car, credit card, debt consolidation, home improvement, or major purchase as a loan purpose decreased the probability of default. Annual income and having a mortgage or owning your house also showed to decrease defaults. While debt to income, loan amount, inquiries past 6 months, revolving line utilization, and having small business as loan purpose were associated with an increase in probability of default.

Zhou, Zhang and Luo (2018) is the only research found using Lending Club data and focusing on loss given default, similar to this study. Their study used data from 2007 to 2017. The data was subset into only containing charged-off loans and the loans where the borrower's income was not verified were screened out which led to a total sample size of 41 717. Their model consisted of 9 variables and they used linear regression with loss given default, LGD, as the dependent variable. Because of high correlation between covariates, for instance interest rate and credit grade, they ran multiple regressions where some covariates were omitted. Some of the findings were that funded amount, employment length, and home ownership were associated with a decrease in LGD. While longer term, interest rate, grade, annual income, debt to income, and delinquencies past 2 years tend to increase LGD. Moreover, they used the Kernal method and found that the LGD showed an unimodal probability distribution with a peak around 0.75 which is similar to that of subordinated bonds.

Chen et al. (2019) studied loan defaults using data from ppdai.com which is one of the biggest P2P platforms in China. They investigated, among other things, important factors to model the probability of default using logistic regression as well as exploring the probability of a loan default utilizing logistic quantile regression. The variable selection process involved comparing stepwise, LASSO, and Bayesian methods. To compare the methods the dataset was split into in-sample and out-of-sample data and then comparing which of the models was the best at predicting defaults.

While starting with 22 variables the final model only contained six, chosen by the Bayesian method. These were Periods, Loan Periods, Interest Due, Interest Rate, Loan Type Ordinary, Loan Type App, Loan Type E-commerce, and finally Regulation Change. While the ordinary logistic regression showed all six variables to be significant, they brought a more comprehensive view through logistic quantile regression. The results showed that in the lower quantiles, all six variables were insignificant while in the higher quantiles, they turned significant as well as larger in magnitude. Thus indicating that the predictors have a high impact on defaults (Chen et al., 2019).

Some of the findings were that higher loan periods and interest rates increase the probability of default. Interest due, which is the amount that was planned to be repaired in the last installment, showed in high quintiles to be more likely to default. Regulation change, which was a dummy variable for all loans with more stringent requirements, surprisingly showed an increase in defaults (Chen et al., 2019).

Aleksandrova (2021) used data from the Lending Club and compared different machine learning methods, some of which were neural networks, eXtreme Gradient Boosting (XGBoost), Random Forest, and Stacked Ensemble of all models. Overall the results showed that XGBoost and Stack Ensemble had the best performance, but interestingly, most of the models agreed when looking at variable importance. All the models except random forest had credit term 60 months and FICO score as the top two most important

variables. Other important variables were dti and dti_loan, which shows the indebtedness level of the borrower, as well as loan amount, number of mortgage accounts, and total limit.

Xia et al. (2019) made a similar study and also used data from the Lending Club. However, they incorporated soft information in the form of text descriptions. Their final model used an advanced gradient-boosting decision tree technique called CatBoost and the text information extraction technique was based on keyword clustering. The results showed that the variables extracted from the texts were useful and significantly improved the ability to predict defaults compared to only using hard information (Xia et al., 2019).

2.2 Research Question

As shown in this literature review a lot of the quantitative research regarding P2P lending, especially Lending Club, has been focused on understanding or predicting defaults. While predicting defaults is great, it does not distinguish between a borrower who defaults on a small amount and a borrower who defaults on the entire loan. Considering that P2P loans have relatively high levels of defaults this leaves a gap in the literature in understanding what drives bigger defaults, which are the ones you really want to avoid as an investor. Zhou, Zhang, and Luo (2018) is the only study found focusing on credit losses while using Lending Club data. Their study does not really explain their variable choices and uses multivariate linear regression which only examines the conditional mean. To deepen the understanding and to bring a more comprehensive view of credit losses this study uses quantile regression and aims to answer the following research question:

What are the determinants of credit losses in P2P lending?

Chapter 3: Method and Data

3.1 Stepwise AIC

This section discusses the stepwise variable selection process used in this study to help us choose important covariates among the available ones. With a large number of covariates, interpretation can be difficult. Therefore we are willing to sacrifice some of the covariates to get a smaller subset with the strongest effects (Hastie, Tibshirani and Freidman, 2009). If we have p predictors there are 2^p candidate models, so for 30 predictors there are over a billion subsets to consider. There are several stepwise variable selection methods to choose from those subsets, for example partial F, partial correlation, and semi-partial correlation. However, we decided to use the stepwise AIC (Akaike information criterion). The previously mentioned methods do not take proper account of the multiple testing issues. Moreover, stepwise AIC is more general and can be applied to nonlinear models as well as non-normally distributed data (Yamashita, Yamashita and Kamimura, 2007).

$$AIC = -2log(\hat{\theta}) + 2p \qquad (1)$$

The AIC is calculated through (1) where $\hat{\theta}$ is the log-likelihood of the model and p is number of covariates. AIC measures the goodness of the model together with the complexity and penalizes models that overfit the data. The goal is to minimize the AIC value, where a lower value indicates a better trade-off between model fit and complexity.

Forward selection using stepwise AIC starts with an empty model only containing the intercept. It then iteratively adds one more covariate that improves the AIC the most. This is repeated, adding one predictor at a time, until no further improvement in the AIC is achieved or until a stopping criterion is met (Hastie, Tibshirani and Freidman, 2009). A backward stepwise using AIC works very similarly but instead starts with the full model and iteratively drops covariates until no further improvement. This is implemented in R using the $Step$ function where you can choose to do forward, backward, or even both.

In addition to AIC, we also examine the variance inflation factor (VIF) values to make sure that the final model does not suffer from multicollinearity issues. Multicollinearity can inflate the standard error of coefficients and even lead to coefficients having opposite sign of what is expected. The coefficients become unstable and highly sensitive to changes in the data (Stock and Watson, 2015). Therefore, we make sure that the selected covariates are not highly correlated which could distort the results. The VIF values are examined using the $Jtools$ package.

3.2 Quantile Regression

After selecting a promising subset of covariates using the variable selection process, we examine the influence of those covariates under different quantiles on credit loss using quantile regression (QR). In short, quantile regression estimates the relationship between conditional quantiles of the response variable given a set of explanatory variables. Because it can be applied to the whole distribution of the conditional quantiles of the response variable it gives a more complex view of relationships. QR does not rely on parametric assumptions about errors or the response variables distribution and is a distribution-free model, while also being less sensitive to outliers compared to least squares regression (Davino, Furno and Vistocco, 2014). Below is an introduction to QR.

QR was first introduced by Koenker and Basset (1978) with the motivation to come up with a model that has comparable efficiency to Gaussian models while being more robust to outliers and addressing the problem of non-Gaussian error distributions. The original paper brings, among other things, four equivariance properties. When you have estimated a coefficient $\hat{\beta}(\tau; y, X)$ between $0 < \tau < 1$ using the observations (y, X), then for any $p\ x\ p$ nonsingular matrix where $A, \gamma \in \mathbb{R}^p$, and $a > 0$ then Eq. (2-5) holds. What this means is that quantile regression estimates have linear transformation properties so if you for example multiply the dependent variable y with 100 then the estimates $\hat{\beta}$ scale accordingly.

$$\hat{\beta}(\tau; ay, X) = a\hat{\beta}(\tau; ay, X) \qquad (2)$$

$$\hat{\beta}(\tau; -ay, X) = a\hat{\beta}(1 - \tau; ay, X) \qquad (3)$$

$$\hat{\beta}(\tau; y + X\gamma, X) = \hat{\beta}(\tau; y, X) + \gamma \qquad (4)$$

$$\hat{\beta}(\tau; y, XA) = A^{-1}\hat{\beta}(\tau; y, X) + \gamma \qquad (5)$$

But a more powerful property of QR is that the model is also equivariance to monotone transformations (Koenker, 2001). This means that for any monotone function $h(\cdot)$, the relationship in Eq. (6) holds, which follows from observing Eq. (7) where T is a random variable.

$$\mathrm{Q}_{h(T)}(\tau|x) = h\big(\mathrm{Q}_T(\tau|x)\big) \qquad (6)$$

$$Prob(T < t|x) = Prob(h(T) < h(t)|x) \qquad (7)$$

This means that taking the log, which is a monotone transformation, of the dependent variable before estimation will yield the same result as applying the log transformation after estimating the conditional quantile function at τ of the original variable. Or simply $\mathrm{Q}_{log(Y)}(\tau|x) = log\big(\mathrm{Q}_Y(\tau|x)\big)$, which does not hold for conditional mean relationships. This property allows one to not worry about conflicting objectives of transforming the dependent variable (Koenker, 2001).

Moreover, this property is particularly important when you censor the observed response variable Eq. (8) where y_i^* is a latent and unobservable response and is assumed to be generated from the linear model with u_i being iid from a distribution F.

$$y_i^* = x_i\beta + u_i \quad i = 1, \dots, n \qquad (8)$$

$$y_i = max\{0, y_i^*\} \qquad (9)$$

Since y_i^* is censored you cannot observe it directly, but instead you can observe Eq. (9). Powel (1986) showed that in this model the property of quantile equivariance under monotone transformation implies that the conditional quantile functions of y_i depend only on the censoring point and are independent of the distribution F. This gives a distribution free model. The τ-th conditional quantile for y_i in this model can be seen in Eq. (10). The parameters can then be estimated by solving Eq. (11) and it is assumed that the vector x_i contains an intercept that absorbs the effect of $F_u^{-1}(u)$. ρ_τ is a loss function that is described later in this chapter.

$$\mathrm{Q}_i(\tau|x) = max\{0, x_i^T\beta + F_u^{-1}(\tau)\} \qquad (10)$$

$$\underset{b}{argmin}\sum_{i=1}^{n}\rho_\tau\big(y_i - max\{0, x_i^T b\}\big) \qquad (11)$$

Furthermore, QR is more robust to outliers than OLS since it inherits robustness properties of ordinary sample quantiles. Koenker and Hallock (2001) illustrate this by revisiting Engel's (1857) analysis of how household income affected household food expenditure. They estimated quantile regression for τ = 0,05; 0,1; 0,25; 0,5; 0,75; 0,9; 0,95 and compared these with the least squares estimate. The conditional mean fit was quite different from the conditional median because of two unusual points with high income but low food expenditure. These two unusual data points caused the least squares fit to be a poor estimate for the poorest households, with the regression line passing above all the low-income households. However, quantile regression did not suffer from these unusual observations and provided a better fit which shows the robustness of the model. Additionally, they could tell that the conditional distribution of food expenditure was left-skewed by the spacing between the quantile regression lines, offering further insights into the data distribution.

Least squares regression focuses on the conditional mean of a response variable Y conditional on a set of variables X. Assuming the Gauss-Markov conditions are met the regression model can be written as $E(Y|X) = X^T\beta$ where you get the coefficients through minimizing the squared error Eq. (12) (Koenker, 2001).

$$\hat{\beta} = \underset{\beta}{argmin}\sum_{i=1}^{n}\big(y_i - x_i^T\beta\big)^2 \qquad (12)$$

Least squares regression is restricted to a specific location of Y's conditional distribution. QR expands on least squares regression changing the focus from the conditional mean to estimates of conditional quantiles. QR was first introduced by Koenker and Basset (1978) and gives a more global view since you can examine the whole distribution of the

conditional quantiles of the response variable. A generic QR model takes the following form.

$$\mathrm{Q}_\tau(\hat{y}|X) = X\hat{\beta}(\tau) \qquad (13)$$

τ specifies the quantile and must be $0 < \tau < 1$. $\mathrm{Q}_\tau(|)$ is the conditional quantile function and even though it is theoretically possible to examine an infinite number of quantiles only a finite number is distinct from one another, which is called the quantile process. Unlike least squares regression, quantile regression does not make any parametric assumptions regarding the error distribution and is a distribution-free model.

The coefficients $\hat{\beta}_p(\tau)$ are interpreted as the rate of change in the τ-th conditional quantile of the response variable given a unit change in the p-th regressor, all else equal (Davino, Romano and Vistocco, 2022). OLS covariates answer the question "how does the conditional mean of Y depend on the covariates X" while quantile regression answers the question "how does the conditional distribution of Y given X = x depend on x" (Koenker, 2001). To contextualize, let us say that there is a variable for age in years and the dependent variable is income. If the coefficient is 1 500 at τ = 0,7 then for each additional year of age, the income at the 70th percentile increases by 1 500 dollars, assuming all other covariates are held constant. This can thus tell us how age affects higher earners, specifically those at τ = 0,7 rather than the average effect. The coefficients are estimated through a modified version of median regression (14) where you put asymmetric weights on the errors through $\rho_\tau()$which is a loss function seen in Eq. (15).

$$\hat{\beta}(\tau) = \underset{\beta(\tau)}{argmin} \sum_{i=1}^{n} \rho_\tau\left(y_i - x_i^T\beta(\tau)\right) \qquad (14)$$

$$\rho_\tau(y) = [\tau - I(y < 0)]y = [(1-\tau)I(y < 0) + \tau I(y > 0)]|y| \qquad (15)$$

Where the loss function (15) puts a $(1-\tau)$ weight on negative errors $(y < 0)$ while a τ weight is put on the positive $(y > 0)$. It thus results in a weighted sum of absolute deviations and we can rewrite the coefficient estimate.

$$\hat{\beta}(\tau) = \underset{\beta(\tau)}{argmin} \sum_{y_i < x_i^T \beta(\tau)} (1-\tau)\left|y_i - x_i^T \beta(\tau)\right| + \sum_{y_i \geq x_i^T \beta(\tau)} \tau\left|y_i - x_i^T \beta(\tau)\right| \qquad (16)$$

In QR the coefficient vector $\hat{\beta}(\tau)$ can vary depending on different quantile levels τ. A special case of quantile regression is when you put $\tau = 0.5$, it is then equivalent to minimizing the sum of absolute residuals and yields median regression line. To assess the model you can use common goodness of fit metrics such as R^2, however it is more common to use pseudo-R^2 when it comes to QR (Davino, Romano and Vistocco, 2022). When predictors in QR are highly correlated, the variance tends to increase. The variance of QR estimators is influenced by the degree of correlation among predictors, which affects the precision of the estimates. QR is implemented in R using the $quantreg$ package.

3.3.1 Data and Data Cleaning

This thesis uses a public dataset from Lending Club which contains loans from 2007 to 2018 third quarter. Since Lending Club no longer offers P2P lending the dataset is no longer available on their website. However, Nathan George has uploaded the full dataset on Kaggle (Kaggle.com, 2019). It includes all the information collected by the Lending Club during its loan process but because the dataset was public on their website the loans are unfortunately de-identified and do not include personal information like age or gender. The entire dataset consists of 2 260 701 observations and 151 covariates. Since we are interested in credit losses the first step was to subset the data using the "loan_status" into only having "charged-off" loans which left us with 268 599 observations.

The next step was to remove redundant covariates. Some were useless for this analysis because they were free-text answers such as "decs", "title", and "emp_title". Or because they would not have a meaning for the regression, for example, membership ID "id" or "url". While others only had one unique value such as "policy_code" or "pymnt_plan". We continued the data-cleaning process by handling missing values where all covariates with more than 15% missing values were removed. Moreover, 10 318 observations were removed because they had missing values in the same 19 covariates.

Some categorical covariates were converted or manipulated. The “term” covariate was turned into a dummy variable where 1 = 60 months and 0 = 36 months. As for the “purpose” the category “credit_card” was put into “debt_consolidation” as taking a loan to pay back credit card debt can be seen as a form of debt consolidation. Moreover, the categories “renewable_enegry” and “ wedding” were dropped because they had very few observations and “house”, “moving” and “home_improvment” was clumped as “house_related”.

Since there was no covariate of “credit_loss” we had to create our own. The dependent variable credit_loss is defined in Eq. (17) which the focus is on the principal amount that is not recovered excluding recoveries, interest payments, and other fees. The loss is then divided by the funded_amnt to get the proportion not paid back. It is then multiplied by 100 to get it in whole percentages. This results in a bounded range between 0 to 100 but to still see the effect of loan amount funded_amnt was used as a covariate in the model.

$$credit\_loss_i = \frac{funded\_amnt_i - total\_rec\_prncp_i}{funded\_amnt_i} \times 100 \qquad (17)$$

Lastly, all the covariates that were not known at loan origination were removed such as “last_fico_range_high”, “hardship_status”, or “last_pymnt_d” together with some problematic covariates for example, “initial_list_status” which is described in the data dictionary as “Initial listing status of the loan. Possible values are W, F” which we cannot interpret the meaning of. After removing the last observations with missing values as well as a few outliers the dataset contained 253 948 observations and 50 candidate covariates.

3.3.2 Variable Selection

The $Step$ function was used together with quantile regression and all the 50 covariates were consistently kept in the model. Using the $jtools$ package the VIF values were

examined, and there were indeed problems with multicollinearity. Covariates were dropped one at a time based on their VIF value and based on which made more sense to keep in the model. For example, "grade" and "int_rate" had high VIF values because your grade is what determines your interest rate. Out of the two, "int_rate" was kept in the model because it was easier and more transparent to understand than an internally calculated loan grade. The threshold used for VIF values was 5. All in all, 14 covariates were dropped because of high VIF values.

After no more covariates had VIF values exceeding 5 attention was set to the correlation matrix. The threshold for what was considered to high correlation was 0,7 and two additional covariates were dropped and a correlation matrix heatmap can be found in Appendix A. The $Step$ function was used again and all covariates were kept in the model. This leaves 34 variables in the model and the dictionary for these can be found in table 1. One-hot encoding was used for the categorical variables in the regression.

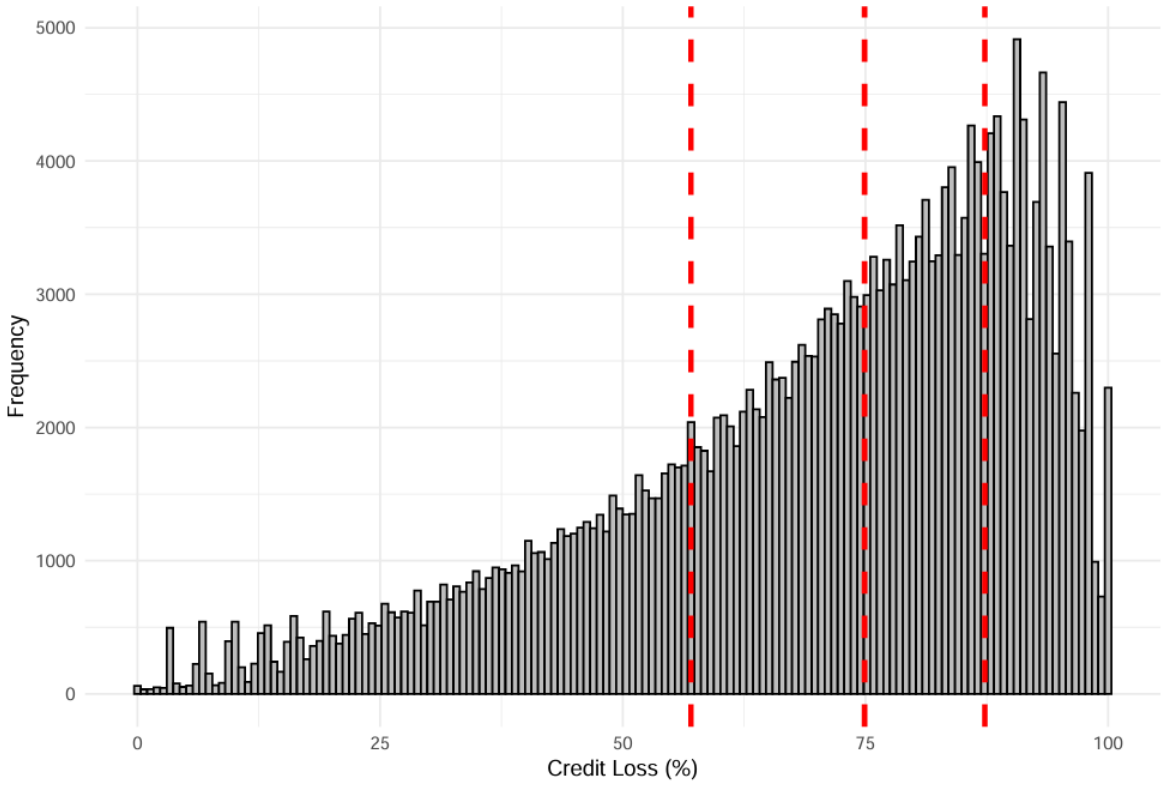
Table 1: Variable Descriptions

Variable	**Meaning**
acc_open_past_24mths	Number of trades opened in past 24 months.
annual_inc	The self-reported annual income provided by the borrower during registration.
application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers
avg_cur_bal	Average current balance of all accounts
chargeoff_within_12_mths	Number of charge-offs within 12 months
collections_12_mths_ex_med	Number of collections in 12 months excluding medical collections
delinq_2yrs	Number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
delinq_amnt	The past-due amount owed for the accounts on which the borrower is now delinquent.
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.
emp_length	Employment length in years.
fico_range_high	The upper boundary range the borrower's FICO at loan origination belongs to.
funded_amnt	The total amount funded to that loan at that point in time.
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
inq_last_6mths	The number of inquiries in past 6 months (excluding auto and mortgage inquiries)
int_rate	Interest Rate on the loan
mo_sin_old_rev_tl_op	Months since oldest revolving account opened
mo_sin_rcnt_rev_tl_op	Months since most recent revolving account opened
mo_sin_rcnt_tl	Months since most recent account opened
mort_acc	Number of mortgage accounts.
num_accts_ever_120_pd	Number of accounts ever 120 or more days past due
num_rev_accts	Number of revolving accounts
num_sats	Number of satisfactory accounts
num_tl_90g_dpd_24m	Number of accounts 90 or more days past due in last 24 months
pct_tl_nvr_dlq	Percent of trades never delinquent
pub_rec	Number of derogatory public records
purpose	A category provided by the borrower for the loan request.
revol_bal	Total credit revolving balance
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
tax_liens	Number of tax liens
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
tot_coll_amt	Total collection amounts ever owed
total_bal_ex_mort	Total credit balance excluding mortgage
total_bc_limit	Total bankcard high credit/credit limit
verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified

Chapter 4: Findings / Results

4.1 Preliminary Analysis

Figure 1: Histogram and Summary Statistics of Credit Loss



Statistic	Mean	Std. Dev	Min	Q1	Median	Q3	Max
Value	69.94	21.71	0.00	57.02	74.91	87.29	100.00

To get an understanding of the data and identify initial patterns a preliminary analysis was done. Figure 1 shows a histogram of the credit losses in the dataset along with summary statistics. The red lines represent the 1st, 2nd, and 3rd quartiles which are 57,02%; 74,91% and 87,29% respectively. The average credit loss in the data is 69,92% with a standard deviation of 21,71%. From this, we can tell that the distribution is left-skewed and that defaults are large in proportion to the funded amount as well as relatively concentrated.

Tables 2 and 3 present summary statistics for the categorial and numerical variables. In Table 3 it is shown that the interest rates ranged from 5,31% to 30,99%, with a standard

deviation of 4,93% which indicates a big variety in interest rates. From Looking at the "application type" in Table 2 it was found that only 2,25% of the observations were joint applications but because the dataset is large this still equates to 5 725 observations. However, the loan purposes "vacation" and "car" only had 1 624 and 1 880 observations respectively meaning the results for these categories might be less precise.

Although the "grade" variable was not included in the final model Table 4 subsets the dataset by grade and presents various statistics. The table shows that there are much fewer observations in grades A, F, and G. This is explained by the fact that few A-graded loans default, and few F- and G-graded loans are funded (Zhou, Zhang and Luo, 2018). We can also see that the median interest rate grows as the grade decreases which makes sense since we have described earlier that the grade, more specifically subgrade, is what determines the interest rate on the loan. Moreover, the debt-to-income ratio (DTI) is partly what decides a borrower's grade and thus shows to increase as the grade decreases.

Table 4 further shows that the median credit loss increases as loan grade decreases and it is also worth noting that the higher grades have much higher median absolute deviation, MAD, indicating greater variability. Moreover, the median funded amount shows to increase as the grade decreases, whereas the annual income does not follow the same trend. The lowest grade had a similar median annual income to the B-graded borrowers. This might lead one to wonder if these borrowers are more inclined to report a higher annual income than they are earning. However, the column "% not verified" showed that the G-graded borrowers actually had the lowest percentage of "not verified" incomes.

Interestingly, the table also showed that the percentage of loans in each category that had payment a term of 60 months, as opposed to 36 months, grew as the grade decreased. For the joint applications, there are no particular patterns, and most grades had a similar proportion of joint applications to the total dataset of 2,25%.

Table 2: Categorical Variables

term			
	Freq	% Total	% Total Cum.
0 = 36 months	154176	60.71	60.71
1 = 60 months	99772	39.29	100.00
Total	253948	100.00	100.00
emp_length			
	Freq	% Total	% Total Cum.
< 1 year	40932	16.12	16.12
1 year	17214	6.78	22.90
2 years	22809	8.98	31.88
3 years	20267	7.98	39.86
4 years	14845	5.85	45.71
5 years	15385	6.06	51.76
6 years	11327	4.46	56.22
7 years	10914	4.30	60.52
8 years	11529	4.54	65.06
9 years	9682	3.81	68.87
10+ years	79044	31.13	100.00
Total	253948	100.00	100.00
home_ownership			
	Freq	% Total	% Total Cum.
MORTGAGE	108088	42.56	42.56
OWN	28439	11.20	53.76
RENT	117421	46.24	100.00
Total	253948	100.00	100.00
verification_status			
	Freq	% Total	% Total Cum.
Not Verified	55232	21.75	21.75
Source Verified	104822	41.28	63.03
Verified	93894	36.97	100.00
Total	253948	100.00	100.00
purpose			
	Freq	% Total	% Total Cum.
debt_consolidation	205202	80.80	80.80
house_related	18107	7.13	87.93
major_purchase	5059	1.99	89.93
medical	3157	1.24	91.17
small_business	3746	1.48	92.65
vacation	1624	0.64	93.28
car	1880	0.74	94.03
other	15173	5.97	100.00
Total	253948	100.00	100.00
application_type			
	Freq	% Total	% Total Cum.
Individual	248223	97.75	97.75
Joint App	5725	2.25	100.00
Total	253948	100.00	100.00

Table 3: Summary Statistics Numerical Variables

Variable	Mean	Std. Dev	Min	Median	Max
acc_open_past_24mths	5.32	3.39	0.00	5.00	46.00
annual_inc ($ 1000)	69.13	38.64	0.75	60.00	350.00
avg_cur_bal	10785.04	12821.23	0.00	5706.50	262093.00
chargeoff_within_12_mths	0.01	0.12	0.00	0.00	8.00
collections_12_mths_ex_med	0.02	0.16	0.00	0.00	9.00
credit_loss	69.94	21.71	0.00	74.91	100.00
delinq_2yrs	0.36	0.94	0.00	0.00	20.00
delinq_amnt	20.04	1003.64	0.00	0.00	249925.00
dti	20.19	8.75	0.00	19.97	59.83
fico_range_high	691.41	25.62	664.00	684.00	850.00
funded_amnt ($ 1000)	15.61	8.78	1.00	14.40	40.00
inq_last_6mths	0.77	1.00	0.00	0.00	6.00
int_rate	15.76	4.93	5.31	15.05	30.99
mo_sin_old_rev_tl_op	171.32	94.44	2.00	154.00	842.00
mo_sin_rcnt_rev_tl_op	11.32	14.41	0.00	7.00	372.00
mo_sin_rcnt_tl	6.89	7.89	0.00	5.00	289.00
mort_acc	1.36	1.81	0.00	1.00	29.00
num_accts_ever_120_pd	0.53	1.32	0.00	0.00	34.00
num_rev_accts	14.45	8.02	1.00	13.00	75.00
num_sats	11.88	5.49	1.00	11.00	40.00
num_tl_90g_dpd_24m	0.10	0.52	0.00	0.00	20.00
pct_tl_nvr_dlq	94.01	8.78	10.50	97.40	100.00
pub_rec	0.25	0.62	0.00	0.00	19.00
revol_bal	14804.53	13970.54	0.00	11076.00	149626.00
revol_util	54.65	23.70	0.00	55.40	129.50
tax_liens	0.06	0.39	0.00	0.00	19.00
tot_coll_amnt	143.15	613.19	0.00	0.00	7994.00
total_bal_ex_mort	48611.10	42265.55	0.00	37775.00	893784.00
total_bc_limit	18386.90	17957.25	0.00	13100.00	337200.00

Table 4: Comparison of Key Statistics by Grade

	Observations	% of total	Median credit_loss (%)	Median funded_amnt ($ 1000)	Median annual_inc ($ 1000)	% not verified	Median dti (%)	Median int_rate (%)	% 60 months	% joint app
A	12863	5.07	61.08 (27.52)	12.00 (7.41)	70.00 (34.1)	40.81	16.91 (8.57)	7.49 (0.71)	4.70	2.40
B	49005	19.30	65.47 (24.88)	12.00 (7.41)	62.00 (29.65)	31.60	18.16 (8.78)	10.99 (1.25)	15.98	1.89
C	81867	32.24	73.23 (20.45)	13.00 (8.60)	60.00 (29.65)	22.90	19.64 (9.21)	13.99 (1.48)	34.07	2.34
D	57996	22.84	77.39 (18.46)	15.00 (8.71)	58.50 (27.43)	18.24	21.08 (9.82)	17.57 (1.45)	44.59	2.43
E	34239	13.48	82.31 (15.05)	16.95 (9.67)	60.00 (28.17)	11.29	21.91 (9.81)	20.99 (2.59)	67.05	2.11
F	13669	5.38	87.04 (11.59)	18.58 (9.67)	60.10 (26.84)	7.83	22.22 (9.74)	24.50 (1.90)	80.57	2.22
G	4309	1.70	90.39 (8.68)	19.52 (10.27)	63.00 (27.94)	5.43	22.17 (9.95)	28.14 (3.42)	84.03	3.32

Note: Median values are shown with Median Absolute Deviation (MAD) in parentheses.

4.2 Quantile Regression Results

The pseudo-R^2 ranged from 0,034 (τ = 0,95) to 0,1276 (τ = 0,15) meaning that we had the best model fit at τ = 0,15. Due to the large amounts of regressions conducted, we have chosen to not present the full regression outputs with standard errors and t values in this thesis. However, interested readers are welcome to request the complete output by contacting the author directly. Table 5 shows the results for all quantiles examined where the bold numbers indicate significance at a significance level of 0,05. Below is each covariate result explained.

"funed_amnt" was not significant at the middling quantiles τ = 0,50; 0,55; 0,60. The coefficients were positive in the lower quantiles but interestingly turned negative at τ = 0,60 onwards, indicating that for loans with high levels of credit loss, an increase in the funded amount was associated with lower credit loss. However, the effects were not very big. An increase of $ 1 000 in funded amount showed the biggest positive coefficient being 0,16 (τ = 0,05) and the biggest negative being -0,031 (τ = 0,80) percentage points.

The variables "term" and "int_rate" were significant across all quantiles and the coefficients indicate that longer loan terms and higher interest rates are associated with higher credit losses. The coefficients for "term" started off very high in the lower quantiles 19,85 (τ = 0,05) and gradually decreased to 1,48 (τ = 0,95). "int_rate" saw an increase up to 1,45 (τ = 0,15) to then diminish toward the higher quantiles 0,31 (τ = 0,95). Overall these two covariates showed a big effect on credit loss.

For all employment lengths, the significant coefficients were negative and in general, the size of the coefficients grew with longer employment lengths. Thus, indicating that borrowers with longer employment are associated with lower credit loss compared to the baseline of "< 1 year". "1 year" was not significant in any quantile while "2 years", "3 years", "4 years" and "5 years" had a few quantiles that were not significant, usually towards the tails. As we move across quantiles, we can see that most employment lengths

peaked around τ = 0,10 to τ = 0,30 to then diminish toward the higher quantiles. In the lower quantiles “7 years” had the biggest effect of -3,91 (τ = 0,10; 0,15) while in the extreme case of τ = 0,95 “8 years” had the biggest effect of -0,81.

When it comes to homeownership “rent” was significant across all quantiles while “own” had one not significant case of τ = 0,15. Both had positive coefficients which indicates that renting and owning a home is associated with more risk in terms of credit loss compared to the baseline “mortgage”. The result for “own” was a bit surprising since you would intuitively associate owning your property with lower-risk individuals. Both categories’ coefficients showed non-monotonic patterns as they grew and shrunk across quantiles. “own” went from 0,92 (τ = 0,05) to 0,61 (τ = 0,95). “rent” went from 0,93 (τ = 0,05) to 0,29 (τ = 0,95). “own“ showed a moderate effect on credit loss that was relatively stable across quantiles. “annual_inc” was significant across quantiles and was consistently associated with an increase in credit loss. The effects were not very big though as an increase of $ 1 000 was associated with 0,024 (τ = 0,20) to 0,010 (τ = 0,95) percentage points.

For the income verification status both “source verified” and “verified” were significant in all quantiles and consistently had negative coefficients indicating lower credit loss compared to the baseline “not verified”. “verified” generally showed a bigger effect, except for the top quantile, and had coefficients ranging from -3,54 (τ = 0,05) to -0,22 (τ = 0,95). “source verified” ranged from -1,20 (τ = 0,05) to -0,23 (τ = 0,95).

The first three categories of loan purpose, “house_related”, “major_purchase”, and “medical”, were significant across all quantiles. “small_business”, “vacation”, and “other” had a few lower quantiles that were not significant in the lower quantiles while “car” was only significant in nine quantiles, which was τ = 0,45 to 0,75 and τ = 0,90; 0,95. Overall all the significant coefficients in every category were positive indicating greater credit loss compared to the baseline “debt_consolidation”. There were only two cases that showed negative coefficients and they were not significant. “medical” consistently had the biggest

coefficients across all quantiles ranging from 4,31 (τ = 0,20) to 2,12 (τ = 0,95). Excluding the not significant cases “small_business” and “ car” had the lowest coefficients from τ = 0,30 onwards while in the lower quantiles “house_related” and “other” had the lowest. Even in the extreme case of τ = 0,95 all categories except “small_business” are associated with an increase in credit loss of more than 1 percentage point which highlights the importance of loan purpose.

The debt-to-income ratio “dti” was significant in all cases except for τ = 0,90. The coefficients are positive across most quantiles but diminish in size from 0,15 (τ = 0,05) to 0,007 (τ = 0,85). Interestingly the effect turns negative -0,011 at τ = 0,95. In general, “dti” is associated with higher credit loss but for the extreme case (τ = 0,95) the results suggest that borrowers who experience high levels of credit loss a higher “dti” is associated with a decrease in credit loss.

“delinq_2yrs” was significant in eleven quantiles. The coefficients started at 0,47 (τ = 0,05), then diminished across quantiles, and from τ = 0,70 onwards no coefficients were significant. “fico_range_high” was not significant up to τ = 0,10 but was significant onwards. The coefficients were positive across all quantiles and ranged from 0,0002 (τ = 0,05) to 0,017 (τ = 0,65). Coefficients were relatively stable but diminished a little toward the highest quantile0,013 (τ = 0,95). The results indicate that higher FICO scores are associated with higher credit losses which is quite counterintuitive since higher FICO scores are associated with lower credit risk.

The results of “inq_last_6mths” were not significant between τ = 0,45 to τ = 0,65 and the coefficients were negative at lower quantiles but turned positive toward higher quantiles. Given the small coefficients and that most observations have had zero inquiries in the past six months the effect is negligible. “pub_rec” were significant from τ = 0,30 onwards and the coefficients were negative except for the extreme case of τ = 0,05. The significant coefficients varied in magnitude from -0,57 (τ = 0,50) to -0,16 (τ = 0,95) suggesting that an increase in derogatory public records is associated with a decrease in credit loss.

Total revolving balance “revol_bal” showed consistently a positive relationship with credit loss and was significant in all quantiles except $\tau = 0,90; 0,95$. In general, the effect was bigger in the lower quantiles and shrunk toward the higher quantiles. The coefficients ranged from 3,17e-05 ($\tau = 0,05$) to 1,00e-05 ($\tau = 0,85$). If you consider the mean 14 804,53 and the standard deviation 13 970,54 of the revolving balance you can see a big variety in the data but the effect is not. The revolving utilization rate “revol_util” consistently had a negative relationship with credit loss and was significant in all quantiles. The effects were stronger in the lower quantiles and the coefficients ranged from -0.101 ($\tau = 0,15$) to -0.021 ($\tau = 0,95$).

“collections_12_mths_ex_med” was not significant in $\tau = 0,05; 0,75; 0,95$ and had consistently a positive effect on credit loss. The effect was bigger in lower quantiles but over all showed a negligible effect. For the application type “joint app” was significant in all quantiles and was consistently positive indicating that a joint application is associated with higher credit loss compared to the baseline “individual”. The coefficients were extremely high in the lower quantiles and shrunk toward higher quantiles ranging from 22,80 ($\tau = 0,05$) to 1,88 ($\tau = 0,95$). In the median case, a joint application was associated with 8,60 percentage points higher credit loss compared to individual applications.

“tot_coll_amt” was only significant at $\tau = 0,55$ and 0,85. The coefficients were mostly negative, but turned positive toward the tails. The most negative was -0,00015 ($\tau = 0,55$) and showed a minor effect on credit loss. “acc_open_past_24mths” showed positive coefficients and was significant in all quantiles. The coefficients showed the biggest effect at $\tau = 0,05$ of 0,72 and consistently shrank across quantiles to 0,038 ($\tau = 0,95$).

Average current balance on all accounts “avg_cur_bal” was significant across quantiles and the coefficients peaked at 5,28e-05 ($\tau = 0,10$) to then shrink to 1,00e-05 ($\tau = 0,95$). “chargeoff_within_12_mths” was significant from $\tau = 0,40$ onwards except in $\tau = 0,75$.

The coefficients were positive and followed a non-monotonic pattern, growing and shrinking across quantiles. Significant coefficients showed a moderate effect on credit loss ranging from 1,16 (τ = 0,55) to 0,61 (τ = 0,95).

"delinq_amnt" was only significant at τ = 0,20; 0,25; 0,60 and showed a negligible positive effect credit loss. "mo_sin_old_rev_tl_op" was significant in all quantiles except τ = 0,05. The coefficients were all negative indicating that an increase in months of a person's oldest revolving account is associated with a decrease in credit loss. Coefficients peaked at -0,0030 (τ = 0,25) and then shrank to -0,0007 (τ = 0,95).
"mo_sin_rcnt_rev_tl_op" was significant in all quantiles and the coefficients were relatively stable across quantiles but shrank toward the higher ones. "mo_sin_rcnt_tl" was not significant at τ = 0,10 and τ = 0,85; 0,90. Coefficients were mostly negative but turned positive toward the highest two quantiles. They ranged from -0,039 (τ = 0,45) to 0,010 (τ = 0,95).

Number of mortgage accounts "mort_acc" was significant across all quantiles and consistently showed to decrease credit loss. The coefficients were bigger in the lower quantiles, shrunk towards higher ones and ranged between -0,45 (τ = 0,15) to -0,038 (τ = 0,95). "num_accts_ever_120_pd" was not significant at τ = 0,65 to 0,75 and at τ = 0,95. The coefficients were all positive and the significant ones ranged from 0,32 (τ = 0,05) to 0,032 (τ = 0,90). The number of revolving accounts "num_rev_accts" was significant in all quantiles except τ = 0,95 and consistently showed negative coefficients. Coefficients peaked at -0,14 (τ = 0,10) and then shrunk to -0,006 (τ = 0,95).

"num_sats" was significant in all quantiles and the coefficients showed an association with decreased credit loss. The coefficients were relatively stable in the lower half of the quantiles examined and shrunk toward the higher quantiles. They ranged from -0,25 (τ = 0,10) to -0,11 (τ = 0,95). "num_tl_90g_dpd_24m" was not significant in any quantile and the coefficients changed sign depending on the quantile. Overall, the covariate showed a negligible effect on credit loss.

Percent of trades never delinquent “pct_tl_nvr_dlq” was significant in all quantiles and consistently had positive coefficients which is quite counterintuitive. The coefficients were bigger in the middling quantiles and smaller towards the tails. They ranged from 0,055 (τ = 0,45) to 0,020 (τ = 0,95). “tax_liens” was only significant at τ = 0,10; 0,15 and the coefficients changed signs depending on the quantile but were mostly negative. The biggest was -0,67 (τ = 0,10).

“total_bal_ex_mort” was significant in all quantiles except τ = 0,05. The coefficients were all positive and ranged from 6,92e-06 (τ = 0,05) to 2,20e-06 (τ = 0,95). “total_bc_limit” was significant across all quantiles and the coefficients were consistently positive. They peaked at 7,32e-05 (τ = 0,25) and then shrunk to 2,44e-05 (τ = 0,95).

Table 5: Quantile Regression Coefficients

	τ				
Variable	0.05	0.10	0.15	0.20	0.25
(Intercept)	3,69	**8,67**	**10,55**	**13,80**	**18,64**
funded_amnt ($ 1000)	**0,16**	**0,13**	**0,11**	**0,085**	**0,062**
term	**19,85**	**18,59**	**17,01**	**15,77**	**14,51**
int_rate	**1,15**	**1,39**	**1,45**	**1,43**	**1,39**
emp_length1 year	0,01	-0,05	-0,07	0,01	0,01
emp_length2 years	-0,43	-0,40	-0,51	**-0,66**	**-0,59**
emp_length3 years	-0,49	**-0,85**	**-1,07**	**-1,22**	**-1,03**
emp_length4 years	**-1,01**	-0,77	**-0,92**	**-1,25**	**-1,29**
emp_length5 years	**-1,50**	**-1,39**	**-1,62**	**-1,58**	**-1,41**
emp_length6 years	**-2,14**	**-2,77**	**-2,72**	**-2,58**	**-2,29**
emp_length7 years	**-3,55**	**-3,91**	**-3,91**	**-3,83**	**-3,66**
emp_length8 years	**-2,51**	**-3,08**	**-3,33**	**-3,59**	**-3,26**
emp_length9 years	**-0,32**	**-1,36**	**-1,46**	**-2,00**	**-2,03**
emp_length10+ years	**-1,66**	**-1,97**	**-1,99**	**-2,23**	**-2,10**
home_ownershipOWN	**0,92**	**0,85**	0,53	**0,72**	**0,80**
home_ownershipRENT	**0,93**	**0,73**	**0,45**	**0,49**	**0,54**
annual_inc ($ 1000)	**0,010**	**0,015**	**0,020**	**0,024**	**0,023**
verification_statusSource Verified	**-1,20**	**-1,14**	**-1,25**	**-1,16**	**-1,13**
verification_statusVerified	**-3,54**	**-3,09**	**-2,69**	**-2,29**	**-1,99**
purposehouse_related	**0,90**	**0,81**	**1,33**	**1,50**	**1,67**
purposemajor_purchase	**2,18**	**2,43**	**1,87**	**1,78**	**2,11**
purposemedical	**2,82**	**3,11**	**3,86**	**4,31**	**4,03**
purposesmall_business	0,55	-0,14	-0,04	0,88	1,35
purposevacation	0,91	1,07	**2,36**	**2,45**	**2,89**
purposecar	0,33	0,25	0,39	0,74	0,90
purposeother	**1,02**	0,37	**0,86**	**1,04**	**1,13**
dti	**0,15**	**0,11**	**0,10**	**0,091**	**0,075**
delinq_2yrs	**0,47**	**0,29**	0,19	**0,25**	**0,31**
fico_range_high	0,000	0,005	**0,012**	**0,015**	**0,016**
inq_last_6mths	**-0,61**	**-0,68**	**-0,65**	**-0,47**	**-0,40**
pub_rec	0,13	-0,00011	0,067	-0,055	-0,22
revol_bal	**3,17E-05**	**3,10E-05**	**2,84E-05**	**2,71E-05**	**2,51E-05**
revol_util	**-0,099**	**-0,109**	**-0,101**	**-0,094**	**-0,087**
collections_12_mths_ex_med	0,81	**1,84**	**2,00**	**1,41**	**1,37**
application_typeJoint App	**22,80**	**17,77**	**15,39**	**14,29**	**13,17**
tot_coll_amt	6,61E-06	3,20E-05	-2,89E-05	-6,80E-05	-8,18E-05
acc_open_past_24mths	**0,72**	**0,66**	**0,57**	**0,52**	**0,50**
avg_cur_bal	**4,14E-05**	**5,28E-05**	**4,73E-05**	**4,68E-05**	**4,51E-05**
chargeoff_within_12_mths	0,43	0,59	0,33	0,46	0,38
delinq_amnt	1,21E-05	0,00013	9,76E-05	**0,00011**	**7,86E-05**
mo_sin_old_rev_tl_op	-0,0021	**-0,0024**	**-0,0026**	**-0,0028**	**-0,0030**
mo_sin_rcnt_rev_tl_op	**0,025**	**0,032**	**0,033**	**0,037**	**0,041**
mo_sin_rcnt_tl	**-0,023**	-0,026	**-0,030**	**-0,038**	**-0,038**
mort_acc	**-0,41**	**-0,44**	**-0,45**	**-0,43**	**-0,37**
num_accts_ever_120_pd	**0,32**	**0,19**	**0,20**	**0,18**	**0,17**
num_rev_accts	**-0,13**	**-0,14**	**-0,14**	**-0,13**	**-0,12**
num_sats	**-0,23**	**-0,25**	**-0,24**	**-0,25**	**-0,24**
num_tl_90g_dpd_24m	-0,49	-0,12	0,009	-0,018	-0,19
pct_tl_nvr_dlq	**0,030**	**0,036**	**0,037**	**0,042**	**0,038**
tax_liens	-0,56	-0,55	**-0,67**	**-0,56**	-0,41
total_bal_ex_mort	6,92E-06	**1,15E-05**	**9,96E-06**	**8,85E-06**	**7,26E-06**
total_bc_limit	**5,81E-05**	**6,29E-05**	**6,82E-05**	**7,11E-05**	**7,32E-05**

Table 5: (Continued)

	τ				
Variable	0.30	0.35	0.40	0.45	0.50
(Intercept)	**22,44**	**27,43**	**31,85**	**35,02**	**39,66**
funded_amnt ($ 1000)	**0,041**	**0,032**	**0,024**	**0,019**	0,011
term	**13,34**	**12,24**	**11,33**	**10,48**	**9,69**
int_rate	**1,34**	**1,26**	**1,19**	**1,11**	**1,04**
emp_length1 year	-0,07	-0,04	-0,07	-0,07	-0,03
emp_length2 years	**-0,55**	**-0,54**	**-0,57**	**-0,50**	**-0,55**
emp_length3 years	**-1,04**	**-0,68**	**-0,81**	**-0,68**	**-0,62**
emp_length4 years	**-1,30**	**-1,23**	**-1,02**	**-0,95**	**-0,91**
emp_length5 years	**-1,14**	**-1,08**	**-1,06**	**-0,88**	**-0,75**
emp_length6 years	**-2,33**	**-2,01**	**-1,68**	**-1,57**	**-1,55**
emp_length7 years	**-3,18**	**-2,84**	**-2,79**	**-2,67**	**-2,52**
emp_length8 years	**-3,29**	**-2,91**	**-2,78**	**-2,55**	**-2,40**
emp_length9 years	**-1,97**	**-1,87**	**-1,81**	**-1,71**	**-1,67**
emp_length10+ years	**-2,10**	**-1,94**	**-1,85**	**-1,76**	**-1,66**
home_ownershipOWN	**0,79**	**0,88**	**0,85**	**1,05**	**0,96**
home_ownershipRENT	**0,58**	**0,59**	**0,57**	**0,55**	**0,55**
annual_inc ($ 1000)	**0,023**	**0,022**	**0,022**	**0,022**	**0,022**
verification_statusSource Verified	**-1,00**	**-0,97**	**-0,87**	**-0,77**	**-0,73**
verification_statusVerified	**-1,70**	**-1,47**	**-1,26**	**-1,15**	**-1,01**
purposehouse_related	**1,81**	**1,86**	**1,82**	**1,81**	**1,83**
purposemajor_purchase	**1,86**	**2,21**	**2,51**	**2,79**	**2,97**
purposemedical	**4,06**	**4,08**	**4,24**	**4,22**	**3,86**
purposesmall_business	**1,37**	**1,10**	**1,29**	**1,37**	**1,12**
purposevacation	**2,89**	**3,07**	**2,43**	**2,61**	**2,57**
purposecar	1,13	1,02	1,34	**1,10**	**1,12**
purposeother	**1,56**	**1,85**	**1,99**	**2,16**	**2,19**
dti	**0,064**	**0,059**	**0,051**	**0,051**	**0,046**
delinq_2yrs	**0,27**	**0,23**	**0,23**	**0,26**	**0,18**
fico_range_high	**0,017**	**0,016**	**0,016**	**0,016**	**0,016**
inq_last_6mths	**-0,31**	**-0,23**	**-0,12**	-0,076	-0,029
pub_rec	**-0,38**	**-0,37**	**-0,39**	**-0,43**	**-0,57**
revol_bal	**3,11E-05**	**3,02E-05**	**2,86E-05**	**2,48E-05**	**2,00E-05**
revol_util	**-0,082**	**-0,077**	**-0,071**	**-0,065**	**-0,060**
collections_12_mths_ex_med	**1,14**	**0,95**	**0,95**	**0,85**	**0,77**
application_typeJoint App	**12,16**	**11,26**	**10,32**	**9,37**	**8,60**
tot_coll_amt	-7,29E-05	-0,00012	-0,00014	-0,00013	-0,00013
acc_open_past_24mths	**0,44**	**0,39**	**0,36**	**0,31**	**0,29**
avg_cur_bal	**4,54E-05**	**4,02E-05**	**3,52E-05**	**3,64E-05**	**3,33E-05**
chargeoff_within_12_mths	0,58	0,75	**0,88**	**0,81**	**1,14**
delinq_amnt	6,03E-05	1,00E-04	8,97E-05	7,15E-05	4,87E-05
mo_sin_old_rev_tl_op	**-0,0029**	**-0,0021**	**-0,0017**	**-0,0014**	**-0,0013**
mo_sin_rcnt_rev_tl_op	**0,035**	**0,037**	**0,039**	**0,038**	**0,035**
mo_sin_rcnt_tl	**-0,032**	**-0,035**	**-0,039**	**-0,039**	**-0,035**
mort_acc	**-0,33**	**-0,32**	**-0,29**	**-0,25**	**-0,23**
num_accts_ever_120_pd	**0,16**	**0,16**	**0,14**	**0,17**	**0,13**
num_rev_accts	**-0,11**	**-0,10**	**-0,093**	**-0,075**	**-0,067**
num_sats	**-0,24**	**-0,24**	**-0,24**	**-0,25**	**-0,25**
num_tl_90g_dpd_24m	-0,11	-0,098	-0,094	-0,13	-0,032
pct_tl_nvr_dlq	**0,046**	**0,049**	**0,050**	**0,055**	**0,052**
tax_liens	-0,26	-0,29	-0,19	-0,15	0,11
total_bal_ex_mort	**6,52E-06**	**7,84E-06**	**7,16E-06**	**5,62E-06**	**4,81E-06**
total_bc_limit	**6,95E-05**	**6,85E-05**	**6,52E-05**	**6,39E-05**	**6,43E-05**

Table 5: (Continued)

	τ				
Variable	0.55	0.60	0.65	0.70	0.75
(Intercept)	**43,34**	**46,97**	**51,28**	**56,04**	**60,75**
funded_amnt ($ 1000)	0,001	-0,010	**-0,018**	**-0,023**	**-0,027**
term	**8,86**	**8,07**	**7,31**	**6,45**	**5,51**
int_rate	**0,97**	**0,90**	**0,83**	**0,77**	**0,69**
emp_length1 year	0,06	-0,02	-0,10	-0,08	-0,13
emp_length2 years	**-0,55**	**-0,45**	**-0,61**	**-0,53**	**-0,50**
emp_length3 years	**-0,60**	**-0,59**	**-0,70**	**-0,60**	**-0,49**
emp_length4 years	**-0,88**	**-0,78**	**-0,76**	**-0,64**	**-0,63**
emp_length5 years	**-0,74**	**-0,52**	**-0,65**	**-0,60**	**-0,67**
emp_length6 years	**-1,39**	**-0,99**	**-0,99**	**-0,84**	**-0,82**
emp_length7 years	**-2,18**	**-1,86**	**-1,76**	**-1,60**	**-1,43**
emp_length8 years	**-2,25**	**-2,10**	**-2,02**	**-1,69**	**-1,52**
emp_length9 years	**-1,52**	**-1,60**	**-1,65**	**-1,45**	**-1,23**
emp_length10+ years	**-1,59**	**-1,48**	**-1,43**	**-1,30**	**-1,14**
home_ownershipOWN	**0,98**	**0,93**	**0,86**	**0,78**	**0,79**
home_ownershipRENT	**0,53**	**0,45**	**0,48**	**0,42**	**0,43**
annual_inc ($ 1000)	**0,021**	**0,020**	**0,021**	**0,019**	**0,019**
verification_statusSource Verified	**-0,70**	**-0,45**	**-0,37**	**-0,35**	**-0,37**
verification_statusVerified	**-0,92**	**-0,66**	**-0,48**	**-0,47**	**-0,46**
purposehouse_related	**1,82**	**1,79**	**1,77**	**1,78**	**1,76**
purposemajor_purchase	**2,94**	**3,07**	**2,98**	**2,94**	**2,81**
purposemedical	**3,87**	**3,95**	**3,99**	**3,69**	**3,52**
purposesmall_business	**1,10**	**1,23**	**1,08**	**1,09**	**0,93**
purposevacation	**2,25**	**2,11**	**2,02**	**2,37**	**2,25**
purposecar	**1,14**	**1,31**	**1,11**	**1,01**	**0,77**
purposeother	**2,21**	**2,24**	**2,19**	**2,24**	**2,18**
dti	**0,042**	**0,038**	**0,034**	**0,025**	**0,018**
delinq_2yrs	**0,17**	0,12	**0,11**	0,065	0,063
fico_range_high	**0,017**	**0,018**	**0,017**	**0,016**	**0,015**
inq_last_6mths	-0,001	0,045	0,065	**0,074**	**0,084**
pub_rec	**-0,50**	**-0,48**	**-0,49**	**-0,45**	**-0,43**
revol_bal	**2,12E-05**	**1,97E-05**	**1,92E-05**	**1,61E-05**	**1,41E-05**
revol_util	**-0,056**	**-0,051**	**-0,048**	**-0,045**	**-0,041**
collections_12_mths_ex_med	**0,75**	**0,63**	**0,49**	**0,43**	0,40
application_typeJoint App	**7,73**	**6,93**	**6,28**	**5,68**	**5,08**
tot_coll_amt	**-0,00015**	-0,00011	-7,29E-05	-6,05E-05	-7,13E-05
acc_open_past_24mths	**0,25**	**0,21**	**0,18**	**0,15**	**0,13**
avg_cur_bal	**3,01E-05**	**2,86E-05**	**2,77E-05**	**2,41E-05**	**1,83E-05**
chargeoff_within_12_mths	**1,16**	**1,08**	**1,06**	**0,81**	0,67
delinq_amnt	3,01E-05	**3,00E-05**	2,52E-05	1,28E-05	3,35E-05
mo_sin_old_rev_tl_op	**-0,0012**	**-0,0010**	**-0,0008**	**-0,0008**	**-0,0012**
mo_sin_rcnt_rev_tl_op	**0,035**	**0,032**	**0,030**	**0,030**	**0,028**
mo_sin_rcnt_tl	**-0,036**	**-0,032**	**-0,026**	**-0,025**	**-0,019**
mort_acc	**-0,21**	**-0,18**	**-0,16**	**-0,12**	**-0,11**
num_accts_ever_120_pd	**0,11**	**0,067**	0,025	0,032	0,054
num_rev_accts	**-0,056**	**-0,048**	**-0,040**	**-0,037**	**-0,028**
num_sats	**-0,24**	**-0,23**	**-0,22**	**-0,21**	**-0,20**
num_tl_90g_dpd_24m	-0,074	0,003	-0,006	0,067	0,003
pct_tl_nvr_dlq	**0,048**	**0,042**	**0,039**	**0,033**	**0,032**
tax_liens	-0,059	-0,032	-0,026	-0,022	-0,035
total_bal_ex_mort	**4,22E-06**	**4,31E-06**	**3,46E-06**	**3,51E-06**	**4,04E-06**
total_bc_limit	**5,96E-05**	**5,67E-05**	**5,35E-05**	**5,20E-05**	**4,83E-05**

Table 5: (Continued)

	τ			
Variable	0.80	0.85	0.90	0.95
(Intercept)	**64,98**	**68,45**	**73,38**	**79,89**
funded_amnt ($ 1000)	**-0,031**	**-0,030**	**-0,028**	**-0,03**
term	**4,61**	**3,64**	**2,68**	**1,48**
int_rate	**0,61**	**0,53**	**0,43**	**0,31**
emp_length1 year	-0,16	-0,04	0,02	-0,12
emp_length2 years	**-0,53**	**-0,42**	**-0,33**	**-0,33**
emp_length3 years	**-0,39**	**-0,45**	**-0,30**	**-0,32**
emp_length4 years	**-0,46**	**-0,33**	-0,14	-0,11
emp_length5 years	**-0,54**	**-0,35**	**-0,25**	-0,13
emp_length6 years	**-0,77**	**-0,62**	**-0,50**	**-0,48**
emp_length7 years	**-1,37**	**-1,07**	**-0,66**	**-0,39**
emp_length8 years	**-1,20**	**-1,03**	**-0,96**	**-0,81**
emp_length9 years	**-1,17**	**-0,97**	**-0,92**	**-0,65**
emp_length10+ years	**-1,03**	**-0,86**	**-0,67**	**-0,54**
home_ownershipOWN	**0,89**	**0,86**	**0,78**	**0,61**
home_ownershipRENT	**0,43**	**0,45**	**0,38**	**0,29**
annual_inc ($ 1000)	**0,018**	**0,015**	**0,013**	**0,010**
verification_statusSource Verified	**-0,34**	**-0,32**	**-0,24**	**-0,23**
verification_statusVerified	**-0,36**	**-0,32**	**-0,25**	**-0,22**
purposehouse_related	**1,68**	**1,54**	**1,38**	**1,03**
purposemajor_purchase	**2,57**	**2,41**	**1,86**	**1,44**
purposemedical	**3,31**	**3,02**	**2,66**	**2,12**
purposesmall_business	**0,98**	**0,97**	**0,66**	**0,48**
purposevacation	**2,19**	**1,87**	**1,74**	**1,28**
purposecar	0,66	0,44	**0,97**	**1,02**
purposeother	**2,10**	**1,77**	**1,61**	**1,42**
dti	**0,015**	**0,007**	0,000	**-0,011**
delinq_2yrs	0,050	0,067	0,049	0,039
fico_range_high	**0,014**	**0,015**	**0,014**	**0,013**
inq_last_6mths	**0,10**	**0,13**	**0,13**	**0,14**
pub_rec	**-0,35**	**-0,31**	**-0,29**	**-0,16**
revol_bal	**1,35E-05**	**1,00E-05**	2,53E-06	3,77E-06
revol_util	**-0,039**	**-0,034**	**-0,029**	**-0,021**
collections_12_mths_ex_med	**0,55**	**0,54**	**0,40**	0,21
application_typeJoint App	**4,28**	**3,43**	**2,72**	**1,88**
tot_coll_amt	-0,00011	**-8,73E-05**	-4,29E-05	1,41E-05
acc_open_past_24mths	**0,096**	**0,089**	**0,069**	**0,038**
avg_cur_bal	**1,83E-05**	**1,73E-05**	**1,57E-05**	**1,00E-05**
chargeoff_within_12_mths	**0,96**	**0,77**	**0,61**	**0,61**
delinq_amnt	2,84E-05	3,25E-05	1,36E-05	-2,51E-08
mo_sin_old_rev_tl_op	**-0,0011**	**-0,0008**	**-0,0008**	**-0,0007**
mo_sin_rcnt_rev_tl_op	**0,026**	**0,021**	**0,018**	**0,012**
mo_sin_rcnt_tl	**-0,014**	-0,0060	0,0014	**0,010**
mort_acc	**-0,087**	**-0,073**	**-0,062**	**-0,038**
num_accts_ever_120_pd	**0,058**	**0,038**	**0,032**	0,003
num_rev_accts	**-0,025**	**-0,020**	**-0,014**	-0,006
num_sats	**-0,18**	**-0,17**	**-0,14**	**-0,11**
num_tl_90g_dpd_24m	-0,050	0,022	0,031	0,076
pct_tl_nvr_dlq	**0,030**	**0,028**	**0,026**	**0,020**
tax_liens	-0,15	-0,15	-0,023	0,031
total_bal_ex_mort	**3,24E-06**	**3,66E-06**	**3,27E-06**	**2,20E-06**
total_bc_limit	**4,26E-05**	**3,83E-05**	**3,44E-05**	**2,44E-05**

4.3 Effect Comparison

To more easily compare covariates we examined the effect size for the continuous variables. Which is the standard deviation of the covariate multiplied by the coefficient. The resulting number then measures the impact on credit loss associated with one standard deviation increase in the covariate. The categorial variables are still interpreted the same as before. The full table for the standardized coefficients can be found in Appendix A, and to keep the results cleaner table 6 summarizes the top five positive and negative effects and their corresponding covariate name. If the covariates are compared in absolute value we can see that “term”, “joint app” and “int_rate” are the top three covariates in every quantile up to $\tau = 0{,}75$ where “medical” as loan purpose takes the spot of “int_rate”. This changed ones again at the extreme case of $\tau = 0{,}95$ where “medical” takes the number one spot, followed by “joint app” and “int_rate”.

The verification status “verified” shows one of the biggest effects in the bottom two quantiles but becomes less significant in comparison to other covariates onwards. Employment length, specifically 7 years, was shown as one of the top five covariates in the lower quantiles examined, except in $\tau = 0{,}35$. While in the higher quantiles the effect becomes less important compared to other covariates. The loan purpose “major_purchase” is among the top 5 from $\tau = 0{,}45$ onwards and purpose “medical” is among the top five from $\tau = 0{,}10$ onwards.

When moving along quantiles employment becomes less important while loan purpose becomes more important. To further illustrate this shift in effects, Figures 2 to 6 show plots of the loan purposes “medical”, “major_purchase” and “vacation” as well as the employment lengths “7 years” and “8 years”. The effect of the covariates is on the y-axis and the quantile is on the x-axis. Solid dots indicate significance at a 5% significance level while the hollow dots indicate that the coefficient was not significant. The blue bands around the estimates are the standard errors for the coefficients. In Figures 2 and 5 we can see that in the lower quantiles the effects of “medical” and “7 years” are similar in absolute value around 4 percentage points. But after $\tau = 0{,}25$ the effect of “7 years” starts to

diminish and almost halves in value at $\tau = 0{,}60$ while "medical" is much more stable and is still around 4 percentage points. In the extreme case of $\tau = 0{,}95$ "medical" is above 2 percentage points while "7 years" is below -0,5 percentage points.

Table 6: Top Effects

	$\tau = 0{,}05$				$\tau = 0{,}10$		
application_typeJoint App	22,80	emp_length7 years	-3,55	term	18,59	emp_length7 years	-3,55
term	19,85	verification_statusVerified	-3,54	application_typeJoint App	17,77	verification_statusVerified	-3,54
int_rate	5,69	emp_length8 years	-2,51	int_rate	6,83	emp_length8 years	-2,51
purposemedical	2,82	revol_util	-2,36	purposemedical	3,11	revol_util	-2,36
acc_open_past_24mths	2,45	emp_length6 years	-2,14	purposemajor_purchase	2,43	emp_length6 years	-2,14
	$\tau = 0{,}15$				$\tau = 0{,}20$		
term	17,01	emp_length7 years	-3,91	term	15,77	emp_length7 years	-3,91
application_typeJoint App	15,39	emp_length8 years	-3,33	application_typeJoint App	14,29	emp_length8 years	-3,33
int_rate	7,13	emp_length6 years	-2,72	int_rate	7,06	emp_length6 years	-2,72
purposemedical	3,86	verification_statusVerified	-2,69	purposemedical	4,31	verification_statusVerified	-2,69
purposevacation	2,36	revol_util	-2,40	purposevacation	2,45	revol_util	-2,40
	$\tau = 0{,}25$				$\tau = 0{,}30$		
term	14,51	emp_length7 years	-3,66	term	13,34	emp_length8 years	-3,29
application_typeJoint App	13,17	emp_length8 years	-3,26	application_typeJoint App	12,16	emp_length7 years	-3,18
int_rate	6,85	emp_length6 years	-2,29	int_rate	6,60	emp_length6 years	-2,33
purposemedical	4,03	emp_length10+ years	-2,10	purposemedical	4,06	emp_length10+ years	-2,10
purposevacation	2,89	revol_util	-2,06	purposevacation	2,89	emp_length9 years	-1,97
	$\tau = 0{,}35$				$\tau = 0{,}40$		
term	12,24	emp_length8 years	-2,91	term	11,33	emp_length7 years	-2,79
application_typeJoint App	11,26	emp_length7 years	-2,84	application_typeJoint App	10,32	emp_length8 years	-2,78
int_rate	6,23	emp_length6 years	-2,01	int_rate	5,86	emp_length10+ years	-1,85
purposemedical	4,08	emp_length10+ years	-1,94	purposemedical	4,24	emp_length9 years	-1,81
purposevacation	3,07	emp_length9 years	-1,87	purposemajor_purchase	2,51	emp_length6 years	-1,68
	$\tau = 0{,}45$				$\tau = 0{,}50$		
term	10,48	emp_length7 years	-2,67	term	9,69	emp_length7 years	-2,52
application_typeJoint App	9,37	emp_length8 years	-2,55	application_typeJoint App	8,60	emp_length8 years	-2,40
int_rate	5,48	emp_length10+ years	-1,76	int_rate	5,11	emp_length9 years	-1,67
purposemedical	4,22	emp_length9 years	-1,71	purposemedical	3,86	emp_length10+ years	-1,66
purposemajor_purchase	2,79	emp_length6 years	-1,57	purposemajor_purchase	2,97	emp_length6 years	-1,55
	$\tau = 0{,}55$				$\tau = 0{,}60$		
term	8,86	emp_length7 years	-2,52	term	8,07	emp_length8 years	-2,10
application_typeJoint App	7,73	emp_length8 years	-2,40	application_typeJoint App	6,93	emp_length7 years	-1,86
int_rate	4,77	emp_length9 years	-1,67	int_rate	4,44	emp_length9 years	-1,60
purposemedical	3,87	emp_length10+ years	-1,66	purposemedical	3,95	emp_length10+ years	-1,48
purposemajor_purchase	2,94	emp_length6 years	-1,55	purposemajor_purchase	3,07	num_sats	-1,26

Table 6: (Continued)

	$\tau = 0{,}65$				$\tau = 0{,}70$		
term	7,31	emp_length8 years	-2,02	term	6,45	emp_length8 years	-1,69
application_typeJoint App	6,28	emp_length7 years	-1,76	application_typeJoint App	5,68	emp_length7 years	-1,60
int_rate	4,11	emp_length9 years	-1,65	int_rate	3,78	emp_length9 years	-1,45
purposemedical	3,99	emp_length10+ years	-1,43	purposemedical	3,69	emp_length10+ years	-1,30
purposemajor_purchase	2,98	num_sats	-1,22	purposemajor_purchase	2,94	num_sats	-1,14
	$\tau = 0{,}75$				$\tau = 0{,}80$		
term	5,51	emp_length8 years	-1,52	term	4,61	emp_length7 years	-1,37
application_typeJoint App	5,08	emp_length7 years	-1,43	application_typeJoint App	4,28	emp_length8 years	-1,20
purposemedical	3,52	emp_length9 years	-1,23	purposemedical	3,31	emp_length9 years	-1,17
int_rate	3,41	emp_length10+ years	-1,14	int_rate	3,02	emp_length10+ years	-1,03
purposemajor_purchase	2,81	num_sats	-1,11	purposemajor_purchase	2,57	num_sats	-0,98
	$\tau = 0{,}85$				$\tau = 0{,}90$		
term	3,64	emp_length7 years	-1,07	application_typeJoint App	2,72	emp_length8 years	-0,96
application_typeJoint App	3,43	emp_length8 years	-1,03	term	2,68	emp_length9 years	-0,92
purposemedical	3,02	emp_length9 years	-0,97	purposemedical	2,66	num_sats	-0,78
int_rate	2,61	num_sats	-0,92	int_rate	2,11	revol_util	-0,69
purposemajor_purchase	2,41	emp_length10+ years	-0,86	purposemajor_purchase	1,86	emp_length10+ years	-0,67
	$\tau = 0{,}95$						
purposemedical	2,12	emp_length8 years	-0,81				
application_typeJoint App	1,88	emp_length9 years	-0,65				
int_rate	1,53	num_sats	-0,58				
term	1,48	emp_length10+ years	-0,54				
purposemajor_purchase	1,44	revol_util	-0,50				

Figure 2: Purpose = Medical

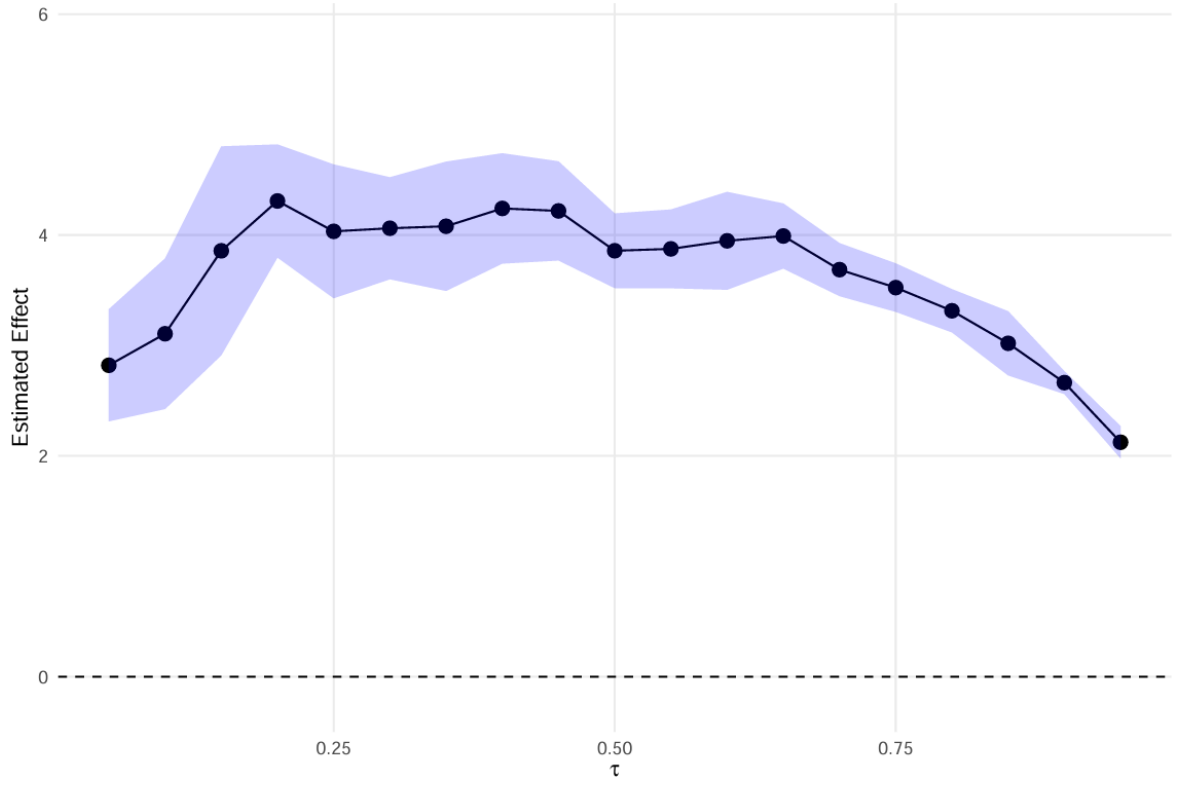


Figure 3: Purpose = Major Purchase

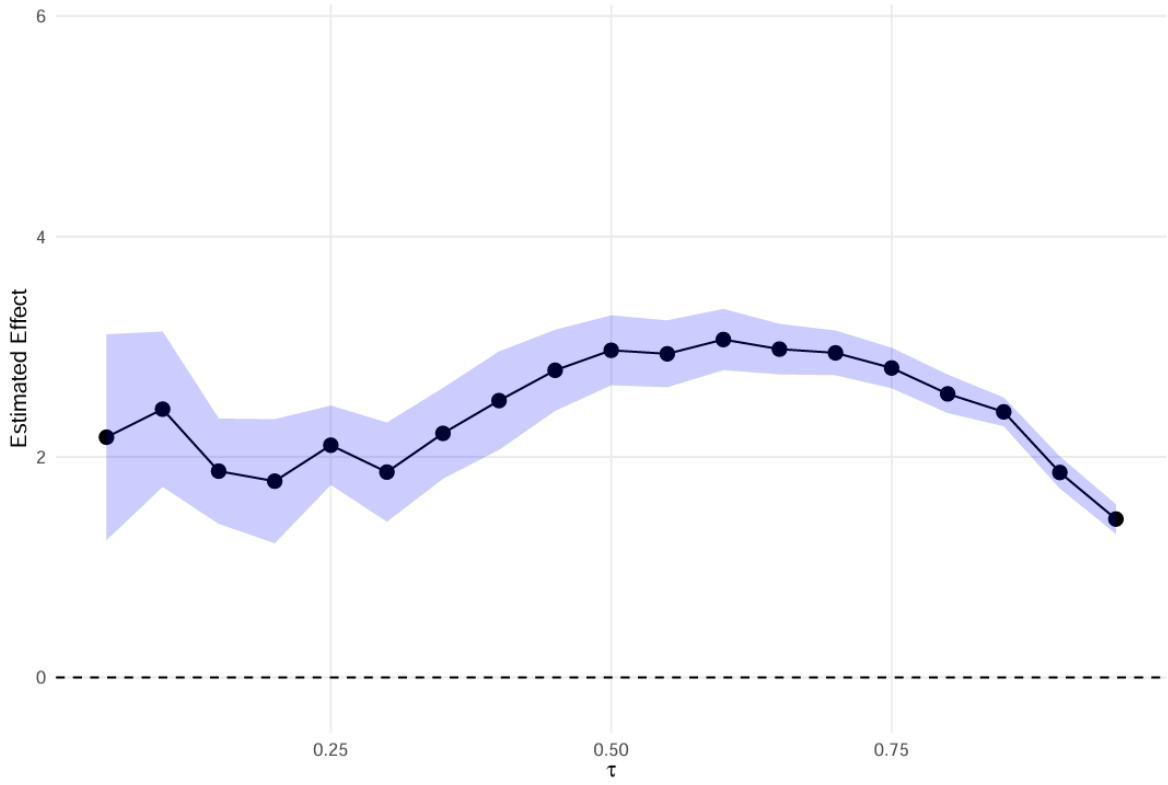


Figure 4: Purpose = Vacation

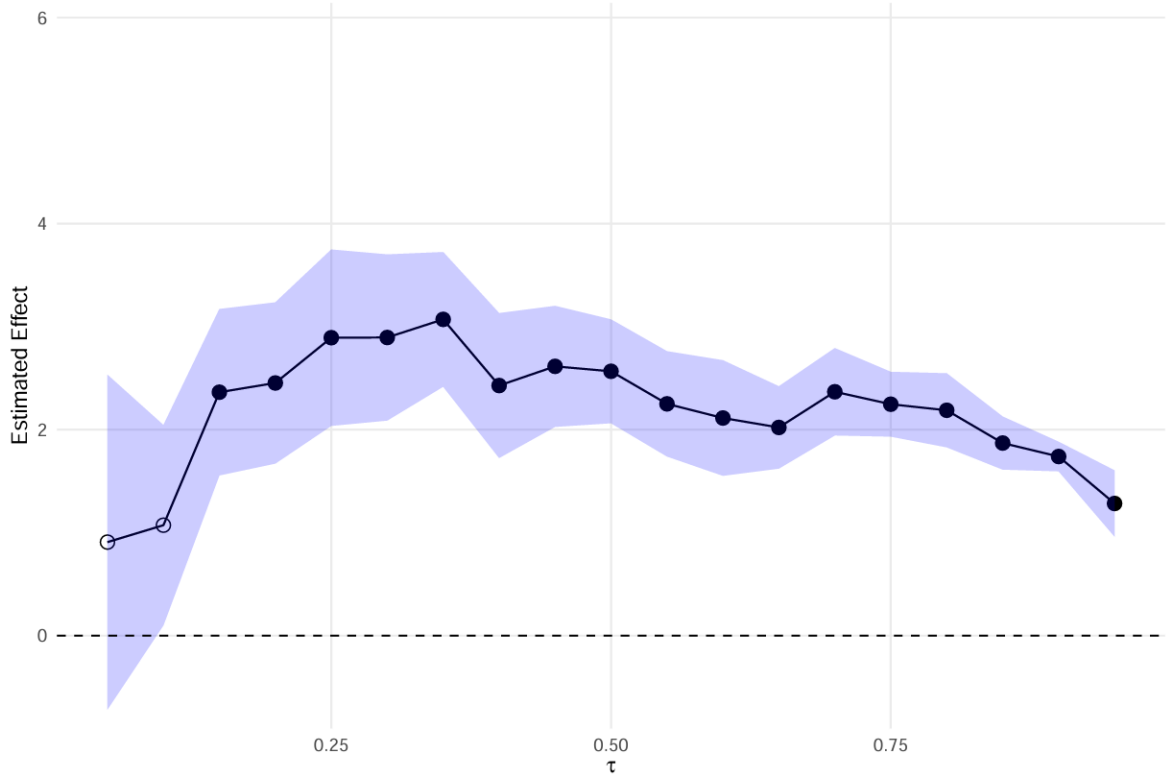


Figure 5: Employment Length = 7 years

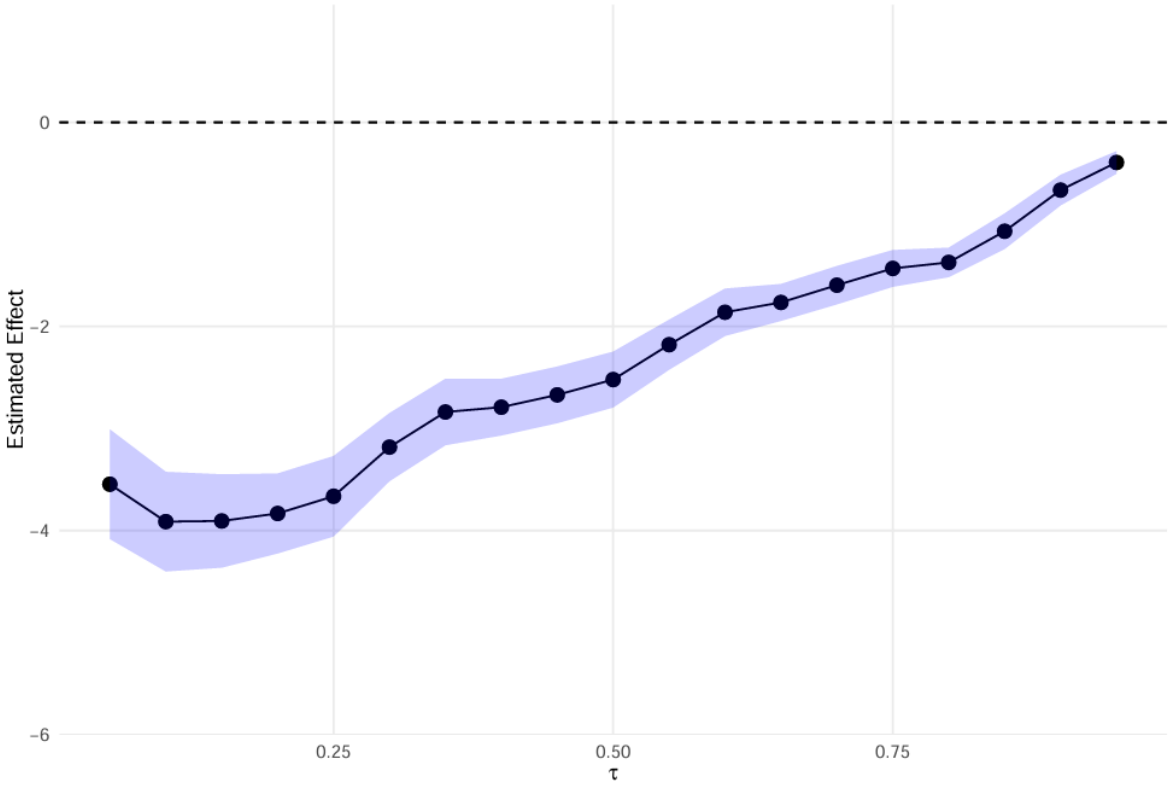
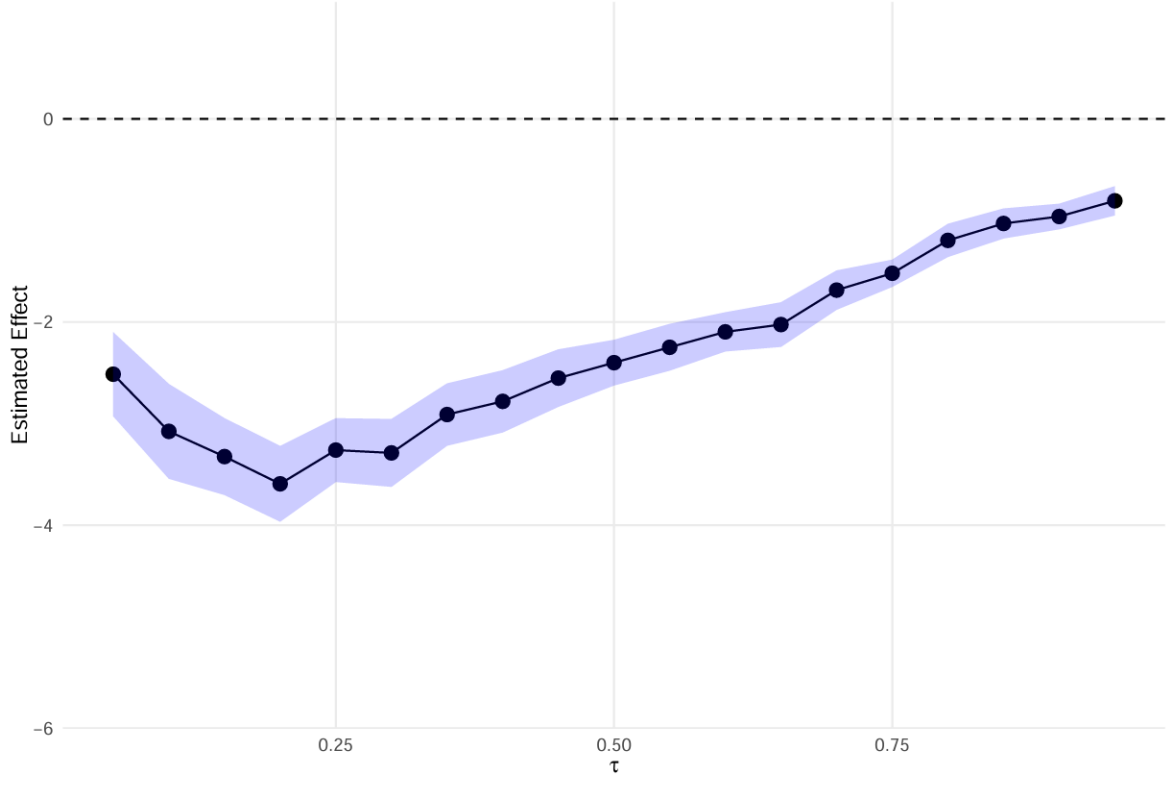


Figure 6: Employment Length = 8 years



Chapter 5: Discussion

5.1 Discussion

To summarize the answer to our research question “What are the determinants of credit losses in P2P lending”. The results show that “term”, “joint app” and “int_rate” are among the most important determinants of credit loss across all quantiles. Having “verified” income shows a big effect in the lowest two quantiles but becomes less important along quantiles. Longer employment length shows a clear association with decreased credit loss across quantiles, but the effect is most prominent in lower quantiles and becomes less important in relation to other covariates in higher quantiles. Versus the loan purpose “medical” which is shown to be a key determinant in all quantiles and gains increased importance to other covariates toward the higher quantiles. Lastly, as we move along quantiles, the loan purposes “major purchase” and “vacation” are also shown to be important determinants of high quantile credit losses. These results are more reliable than previous studies because of the larger sample size of 254 948 observations as well as the robust methodology.

Tabel 6 summarizes the top five positive and negative effects on credit losses and their corresponding covariate name. From this, we can see that the loan term is one of the biggest determinants of increased credit loss. It ranks as one of the top two in every quantile except in the very high quantiles of 0,90 and 0,95 where it ranks in the 3rd and 4th place respectively. The preliminary analysis showed that the proportion of 60-month loans moved together with loan grade, but we control for this relationship by including “int-rate” in the model which helps to isolate the effect of the loan term. Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios (2015) and Polena and Regner (2019) only studied loans with 36 months duration with the motivation that there were too few loans with 60 months duration. This is not the case in our version of the dataset as 39,29% of the observations are 60-month duration loans. The results are in line with Zhou, Zhang, and Lou (2018) who found that longer loan terms increase loss given default, LGD. Moreover, Aleksandrova’s

(2021) machine-learning methods further confirm this result as they showed 60 months as one of the top two most important variables to predict defaults.

Having a joint application type with a co-borrower as opposed to an individual one is shown to be one of the most important determinants of credit losses. The effect was one of the biggest two in every quantile examined. Considering that the preliminary analysis showed all grades to have a similar proportion of joint applications further shows the importance of these results. One explanation could be that a co-borrower is often used to help an individual obtain a loan they otherwise would not be able to get. Combining this with the fact that P2P lending naturally attracts high-risk borrowers might result in ultra-high-risk borrowers which could explain the results.

None of the previous studies reviewed in this thesis included application type as a variable in their analysis and we can therefore not compare the results. An explanation for this could be that Polena and Regner (2019) describe that 60-month duration loans were first introduced in 2010, so it might be that the joint application type became a feature after the period they examined which also would explain the smaller sample size in our data.

Serrano-Cinca, Gutierrez-Nieto, and López-Palacios (2015) describe that the loan grade was what determined the interest rate at the Lending Club and that those two covariates therefore have a very high correlation. Depending on which of the two were included they have consistently shown to have a large effect on defaults and LGD in the previous literature. The results of this study show that higher interest rates are one of the most important determinants of higher credit losses which is in line with the previous literature (Serrano-Cinca, Gutierrez-Nieto and López-Palacios, 2015; Polena and Regner, 2019; Zhou, Zhang and Lou, 2018; Chen et al., 2019).

The results show employment length to be an important determinant of credit losses, especially in the lower quantiles. This is in line with Zhou, Zhang, and Lou (2018) who

found longer employment lengths to be associated with lower LGD. Whereas Serrano-Cinca, Gutierrez-Nieto, and López-Palacios (2015) did not find employment length to have a statistically significant impact on default, though the coefficient implied a slight decrease in defaults.

Polena and Regner (2019) found that the loan purposes debt consolidation and credit card lowered the probability of default, this cannot be directly compared to this study but we can see that all other categories exhibit higher credit losses in the significant cases compared to our combined baseline category of debt consolidation and credit card. Moreover, they found that the purpose small business had the highest association with defaults whereas the result in this study shows the purpose small business to have among the smallest effect on credit losses compared to the other examined categories.

Another interesting finding to compare to the literature is the funded amount. Polena and Regner (2019) found that the funded amount increases defaults and Zhou, Zhang and Lou (2018) found that the funded amount decreases LGD. This study finds that the funded amount is associated with increased credit loss in the lower half of the quantiles, but in the upper half, the effect switches and turns negative. The result is therefore partly in line with and partly contradicts the results of Zhou, Zhang and Lou (2018) depending on what quantiles you consider.

This research has met the objective of analyzing the relationship between borrower characteristics and credit losses in P2P lending. Some of the results are in line with previous literature, some contradict, and some are completely new, like “joint application”. The findings contribute to the academic understanding of credit risk in P2P lending and offer valuable insights for investors, P2P platforms, and policymakers. Moreover, the research has fulfilled the aim of adding knowledge to the relatively unexplored subject of credit losses in P2P lending.

Chapter 6: Conclusions and Prospects

6.1 Conclusions

P2P lending has transformed the financial landscape by providing borrowers with an alternative avenue to access funds that cut out traditional intermediaries such as banks. It has grown in popularity due to faster loan processes, lower interest rates compared to other unsecured personal loans, and the possibility of lending with an unfavorable credit history (Chen et al., 2019; Lenz, 2016). However, this form of lending comes with an array of risks such as liquidity risk, platform risk, information asymmetry, and credit risk. These risks have led to a lot of quantitative research in P2P lending to focus on understanding and predicting defaults (Aleksandrova, 2021; Chen et al., 2019; Polena and Regner, 2019; Serrano-Cinca, Gutierrez-Nieto and López-Palacios, 2015, Xia et al., 2019). However, a default does not distinguish between a small amount or an entire loan which leaves a gap in the literature in understanding credit losses.

This study fills this gap in the literature on the relatively unexplored area of credit losses in peer-to-peer lending. Through quantile regression, we have examined the relationship between borrower characteristics and credit losses to answer the question “What are the determinants of credit losses in P2P lending”.

Our result results find that the loan term, the application type, and the interest rate are among the most important determinants of credit losses. Employment length is an important determinant in the lower quantiles, but the effect becomes less significant compared to other covariates as we move along the quantiles. Lastly, loan purpose is shown to be among the top determinants and becomes more important toward the higher quantiles.

These results give new insights and contribute to the academic understanding of credit risk in P2P lending. By identifying the determinants of credit losses, we offer valuable insights

for investors, P2P platforms, as well as policymakers in assessing credit risk. For investors, these insights can aid risk management, reduce information asymmetry, and help to identify loans that could result in a higher loss in the case of default. Similarly, P2P platforms can use these findings to enhance their credit grading process and improve transparency. These insights can also be used by policymakers to guide future regulation changes. Moreover, as the literature on this subject is very scarce, we have set a theoretical framework for future studies to build upon.

6.2 Limitations and Future Research

The variable selection process, stepwise AIC, was set to select a subset of important covariates for us to examine. However, since all covariates were selected, it did not add much value to the research. Additionally, stepwise AIC is usually used in prediction applications to not get an overparameterized model. Future research should therefore consider using other variable selection processes. Although the dataset is large it is limited to a specific P2P platform, the Lending Club, and one country. Therefore, more studies should be conducted on different platforms to verify the results and assess their generalizability. Another limitation of this study is that it does not consider any macroeconomic factors. As shown in the introduction of this study these factors play an important role for borrowers and future research could do a longitudinal study exploring macroeconomic factors. Lastly, the results shown are purely observational and we can only report associations. We have not run an experiment and we cannot claim any causal effect.

This research finds “term” and “joint app” as important determinants of increased credit loss. However, it would be valuable to explore how these covariates affect defaults, as covariates can affect defaults and credit loss differently. Therefore, we suggest that future research include these variables when revisiting the topic of defaults. Another suggestion is to use this study as a theoretical framework for future research that goes beyond the quantiles examined in this case. A study that uses extremal quantile regression to examine the tails of the distribution would be a natural continuation of this topic.

Appendix A: Tables

Table A1: Standardized Coefficients

	τ				
Variable	0.05	0.10	0.15	0.20	0.25
funded_amnt	**1,40**	**1,14**	**0,97**	**0,74**	**0,54**
term	**19,85**	**18,59**	**17,01**	**15,77**	**14,51**
int_rate	**5,69**	**6,83**	**7,13**	**7,06**	**6,85**
emp_length1 year	0,01	-0,05	-0,07	0,01	0,01
emp_length2 years	-0,43	-0,40	-0,51	**-0,66**	**-0,59**
emp_length3 years	-0,49	**-0,85**	**-1,07**	**-1,22**	**-1,03**
emp_length4 years	**-1,01**	-0,77	**-0,92**	**-1,25**	**-1,29**
emp_length5 years	**-1,50**	**-1,39**	**-1,62**	**-1,58**	**-1,41**
emp_length6 years	**-2,14**	**-2,77**	**-2,72**	**-2,58**	**-2,29**
emp_length7 years	**-3,55**	**-3,91**	**-3,91**	**-3,83**	**-3,66**
emp_length8 years	**-2,51**	**-3,08**	**-3,33**	**-3,59**	**-3,26**
emp_length9 years	**-0,32**	**-1,36**	**-1,46**	**-2,00**	**-2,03**
emp_length10+ years	**-1,66**	**-1,97**	**-1,99**	**-2,23**	**-2,10**
home_ownershipOWN	**0,92**	**0,85**	0,53	**0,72**	**0,80**
home_ownershipRENT	**0,93**	**0,73**	**0,45**	**0,49**	**0,54**
annual_inc	**0,38**	**0,60**	**0,76**	**0,92**	**0,89**
verification_statusSource Verified	**-1,20**	**-1,14**	**-1,25**	**-1,16**	**-1,13**
verification_statusVerified	**-3,54**	**-3,09**	**-2,69**	**-2,29**	**-1,99**
purposehouse_related	**0,90**	**0,81**	**1,33**	**1,50**	**1,67**
purposemajor_purchase	**2,18**	**2,43**	**1,87**	**1,78**	**2,11**
purposemedical	**2,82**	**3,11**	**3,86**	**4,31**	**4,03**
purposesmall_business	0,55	-0,14	-0,04	0,88	**1,35**
purposevacation	0,91	1,07	**2,36**	**2,45**	**2,89**
purposecar	0,33	0,25	0,39	0,74	0,90
purposeother	**1,02**	0,37	**0,86**	**1,04**	**1,13**
dti	**1,34**	**0,99**	**0,84**	**0,79**	**0,66**
delinq_2yrs	**0,44**	**0,27**	0,18	**0,23**	**0,29**
fico_range_high	0,01	0,12	**0,30**	**0,37**	**0,41**
inq_last_6mths	**-0,61**	**-0,68**	**-0,65**	**-0,47**	**-0,40**
pub_rec	0,08	0,00	0,04	-0,03	-0,13
revol_bal	**0,44**	**0,43**	**0,40**	**0,38**	**0,35**
revol_util	**-2,36**	**-2,59**	**-2,40**	**-2,24**	**-2,06**
collections_12_mths_ex_med	0,13	**0,29**	**0,32**	**0,23**	**0,22**
application_typeJoint App	**22,80**	**17,77**	**15,39**	**14,29**	**13,17**
tot_coll_amt	0,00	0,02	-0,02	-0,04	-0,05
acc_open_past_24mths	**2,45**	**2,25**	**1,92**	**1,77**	**1,70**
avg_cur_bal	**0,53**	**0,68**	**0,61**	**0,60**	**0,58**
chargeoff_within_12_mths	0,05	0,07	0,04	0,06	0,05
delinq_amnt	0,01	0,13	0,10	**0,11**	**0,08**
mo_sin_old_rev_tl_op	-0,20	**-0,23**	**-0,25**	**-0,26**	**-0,28**
mo_sin_rcnt_rev_tl_op	**0,35**	**0,47**	**0,47**	**0,54**	**0,59**
mo_sin_rcnt_tl	**-0,18**	-0,20	**-0,23**	**-0,30**	**-0,30**
mort_acc	**-0,74**	**-0,80**	**-0,82**	**-0,78**	**-0,67**
num_accts_ever_120_pd	**0,42**	**0,25**	**0,26**	**0,24**	**0,22**
num_rev_accts	**-1,03**	**-1,16**	**-1,12**	**-1,01**	**-0,98**
num_sats	**-1,28**	**-1,36**	**-1,30**	**-1,39**	**-1,33**
num_tl_90g_dpd_24m	-0,25	-0,06	0,00	-0,01	-0,10
pct_tl_nvr_dlq	**0,27**	**0,32**	**0,33**	**0,37**	**0,34**
tax_liens	-0,22	-0,21	**-0,26**	**-0,22**	-0,16
total_bal_ex_mort	0,29	**0,49**	**0,42**	**0,37**	**0,31**
total_bc_limit	**1,04**	**1,13**	**1,22**	**1,28**	**1,31**

Table A1: (Continued)

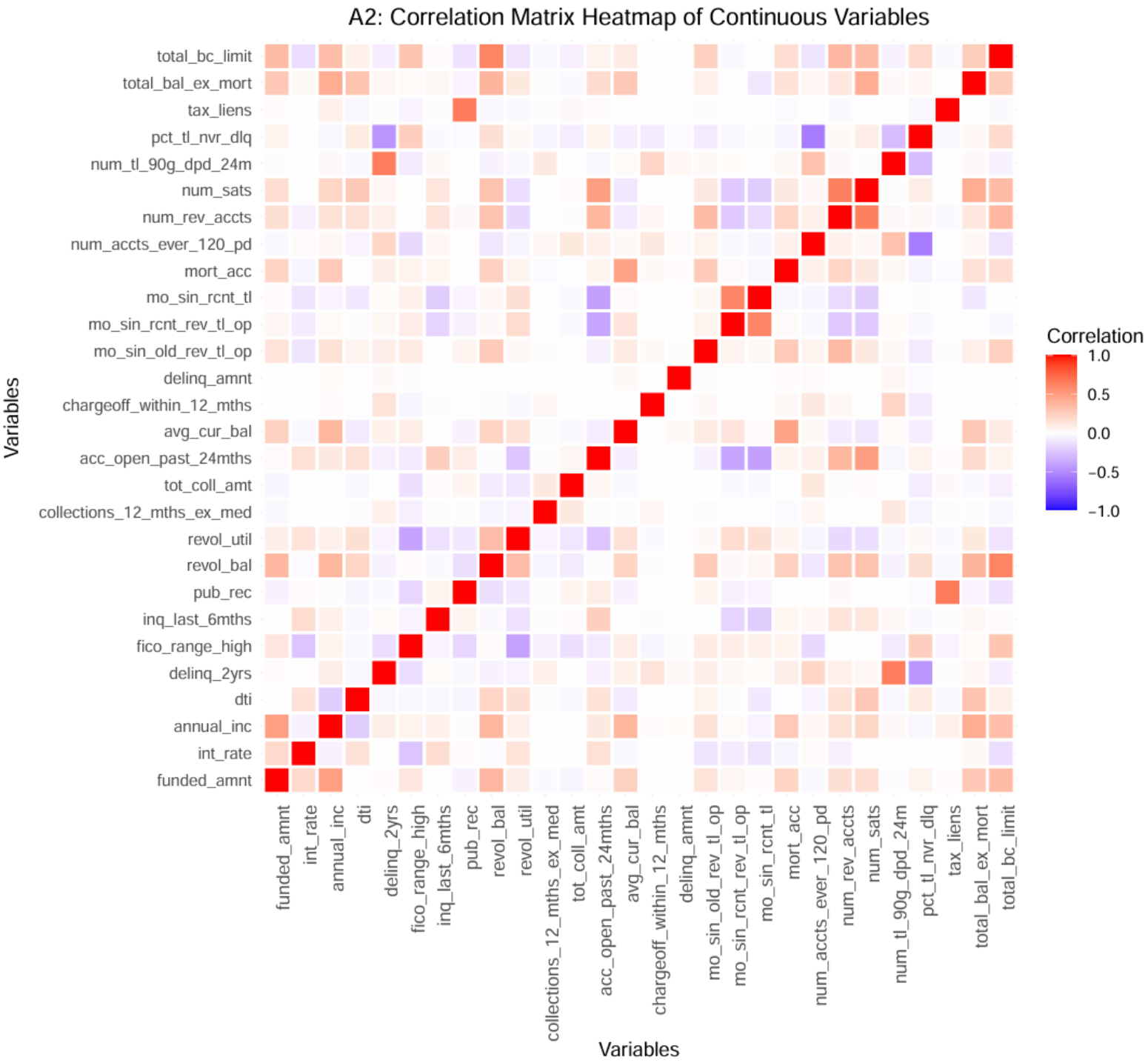
	τ				
Variable	0.30	0.35	0.40	0.45	0.50
funded_amnt	**0,36**	**0,28**	**0,21**	**0,17**	0,09
term	**13,34**	**12,24**	**11,33**	**10,48**	**9,69**
int_rate	**6,60**	**6,23**	**5,86**	**5,48**	**5,11**
emp_length1 year	-0,07	-0,04	-0,07	-0,07	-0,03
emp_length2 years	**-0,55**	**-0,54**	**-0,57**	**-0,50**	**-0,55**
emp_length3 years	**-1,04**	**-0,68**	**-0,81**	**-0,68**	**-0,62**
emp_length4 years	**-1,30**	**-1,23**	**-1,02**	**-0,95**	**-0,91**
emp_length5 years	**-1,14**	**-1,08**	**-1,06**	**-0,88**	**-0,75**
emp_length6 years	**-2,33**	**-2,01**	**-1,68**	**-1,57**	**-1,55**
emp_length7 years	**-3,18**	**-2,84**	**-2,79**	**-2,67**	**-2,52**
emp_length8 years	**-3,29**	**-2,91**	**-2,78**	**-2,55**	**-2,40**
emp_length9 years	**-1,97**	**-1,87**	**-1,81**	**-1,71**	**-1,67**
emp_length10+ years	**-2,10**	**-1,94**	**-1,85**	**-1,76**	**-1,66**
home_ownershipOWN	**0,79**	**0,88**	**0,85**	**1,05**	**0,96**
home_ownershipRENT	**0,58**	**0,59**	**0,57**	**0,55**	**0,55**
annual_inc	**0,88**	**0,85**	**0,86**	**0,86**	**0,84**
verification_statusSource Verified	**-1,00**	**-0,97**	**-0,87**	**-0,77**	**-0,73**
verification_statusVerified	**-1,70**	**-1,47**	**-1,26**	**-1,15**	**-1,01**
purposehouse_related	**1,81**	**1,86**	**1,82**	**1,81**	**1,83**
purposemajor_purchase	**1,86**	**2,21**	**2,51**	**2,79**	**2,97**
purposemedical	**4,06**	**4,08**	**4,24**	**4,22**	**3,86**
purposesmall_business	**1,37**	**1,10**	**1,29**	**1,37**	**1,12**
purposevacation	**2,89**	**3,07**	**2,43**	**2,61**	**2,57**
purposecar	1,13	1,02	1,34	**1,10**	**1,12**
purposeother	**1,56**	**1,85**	**1,99**	**2,16**	**2,19**
dti	**0,56**	**0,51**	**0,44**	**0,45**	**0,40**
delinq_2yrs	**0,25**	**0,21**	**0,21**	**0,25**	**0,17**
fico_range_high	**0,43**	**0,41**	**0,41**	**0,42**	**0,41**
inq_last_6mths	**-0,31**	**-0,23**	**-0,12**	-0,08	-0,03
pub_rec	**-0,23**	**-0,23**	**-0,24**	**-0,27**	**-0,35**
revol_bal	**0,43**	**0,42**	**0,40**	**0,35**	**0,28**
revol_util	**-1,95**	**-1,82**	**-1,67**	**-1,53**	**-1,42**
collections_12_mths_ex_med	**0,18**	**0,15**	**0,15**	**0,14**	**0,12**
application_typeJoint App	**12,16**	**11,26**	**10,32**	**9,37**	**8,60**
tot_coll_amt	-0,04	-0,07	-0,09	-0,08	-0,08
acc_open_past_24mths	**1,50**	**1,34**	**1,21**	**1,06**	**0,98**
avg_cur_bal	**0,58**	**0,52**	**0,45**	**0,47**	**0,43**
chargeoff_within_12_mths	0,07	0,09	**0,11**	**0,10**	**0,14**
delinq_amnt	0,06	0,10	0,09	0,07	0,05
mo_sin_old_rev_tl_op	**-0,27**	**-0,20**	**-0,16**	**-0,14**	**-0,12**
mo_sin_rcnt_rev_tl_op	**0,50**	**0,53**	**0,55**	**0,54**	**0,50**
mo_sin_rcnt_tl	**-0,25**	**-0,28**	**-0,31**	**-0,30**	**-0,27**
mort_acc	**-0,59**	**-0,58**	**-0,53**	**-0,46**	**-0,43**
num_accts_ever_120_pd	**0,21**	**0,21**	**0,19**	**0,23**	**0,17**
num_rev_accts	**-0,88**	**-0,82**	**-0,75**	**-0,60**	**-0,54**
num_sats	**-1,33**	**-1,34**	**-1,34**	**-1,35**	**-1,38**
num_tl_90g_dpd_24m	-0,06	-0,05	-0,05	-0,07	-0,02
pct_tl_nvr_dlq	**0,40**	**0,43**	**0,44**	**0,48**	**0,46**
tax_liens	-0,10	-0,11	-0,07	-0,06	0,04
total_bal_ex_mort	**0,28**	**0,33**	**0,30**	**0,24**	**0,20**
total_bc_limit	**1,25**	**1,23**	**1,17**	**1,15**	**1,15**

Table A1: (Continued)

	τ				
Variable	0.55	0.60	0.65	0.70	0.75
funded_amnt	0,01	-0,09	**-0,16**	**-0,20**	**-0,24**
term	**8,86**	**8,07**	**7,31**	**6,45**	**5,51**
int_rate	**4,77**	**4,44**	**4,11**	**3,78**	**3,41**
emp_length1 year	0,06	-0,02	-0,10	-0,08	-0,13
emp_length2 years	**-0,55**	**-0,45**	**-0,61**	**-0,53**	**-0,50**
emp_length3 years	**-0,60**	**-0,59**	**-0,70**	**-0,60**	**-0,49**
emp_length4 years	**-0,88**	**-0,78**	**-0,76**	**-0,64**	**-0,63**
emp_length5 years	**-0,74**	**-0,52**	**-0,65**	**-0,60**	**-0,67**
emp_length6 years	**-1,39**	**-0,99**	**-0,99**	**-0,84**	**-0,82**
emp_length7 years	**-2,18**	**-1,86**	**-1,76**	**-1,60**	**-1,43**
emp_length8 years	**-2,25**	**-2,10**	**-2,02**	**-1,69**	**-1,52**
emp_length9 years	**-1,52**	**-1,60**	**-1,65**	**-1,45**	**-1,23**
emp_length10+ years	**-1,59**	**-1,48**	**-1,43**	**-1,30**	**-1,14**
home_ownershipOWN	**0,98**	**0,93**	**0,86**	**0,78**	**0,79**
home_ownershipRENT	**0,53**	**0,45**	**0,48**	**0,42**	**0,43**
annual_inc	**0,82**	**0,78**	**0,80**	**0,74**	**0,73**
verification_statusSource Verified	**-0,70**	**-0,45**	**-0,37**	**-0,35**	**-0,37**
verification_statusVerified	**-0,92**	**-0,66**	**-0,48**	**-0,47**	**-0,46**
purposehouse_related	**1,82**	**1,79**	**1,77**	**1,78**	**1,76**
purposemajor_purchase	**2,94**	**3,07**	**2,98**	**2,94**	**2,81**
purposemedical	**3,87**	**3,95**	**3,99**	**3,69**	**3,52**
purposesmall_business	**1,10**	**1,23**	**1,08**	**1,09**	**0,93**
purposevacation	**2,25**	**2,11**	**2,02**	**2,37**	**2,25**
purposecar	**1,14**	**1,31**	**1,11**	**1,01**	**0,77**
purposeother	**2,21**	**2,24**	**2,19**	**2,24**	**2,18**
dti	**0,37**	**0,33**	**0,29**	**0,22**	**0,16**
delinq_2yrs	**0,16**	0,11	**0,10**	0,06	0,06
fico_range_high	**0,43**	**0,45**	**0,43**	**0,41**	**0,38**
inq_last_6mths	0,00	0,04	0,07	**0,07**	**0,08**
pub_rec	**-0,31**	**-0,30**	**-0,31**	**-0,28**	**-0,27**
revol_bal	**0,30**	**0,28**	**0,27**	**0,23**	**0,20**
revol_util	**-1,32**	**-1,21**	**-1,15**	**-1,06**	**-0,98**
collections_12_mths_ex_med	**0,12**	**0,10**	**0,08**	**0,07**	0,06
application_typeJoint App	**7,73**	**6,93**	**6,28**	**5,68**	**5,08**
tot_coll_amt	**-0,09**	-0,07	-0,04	-0,04	-0,04
acc_open_past_24mths	**0,85**	**0,71**	**0,62**	**0,51**	**0,43**
avg_cur_bal	**0,39**	**0,37**	**0,35**	**0,31**	**0,24**
chargeoff_within_12_mths	**0,14**	**0,13**	**0,13**	**0,10**	0,08
delinq_amnt	0,03	**0,03**	0,03	0,01	0,03
mo_sin_old_rev_tl_op	**-0,12**	**-0,10**	**-0,08**	**-0,08**	**-0,11**
mo_sin_rcnt_rev_tl_op	**0,50**	**0,47**	**0,43**	**0,43**	**0,40**
mo_sin_rcnt_tl	**-0,28**	**-0,25**	**-0,21**	**-0,20**	**-0,15**
mort_acc	**-0,38**	**-0,32**	**-0,29**	**-0,22**	**-0,20**
num_accts_ever_120_pd	**0,15**	**0,09**	0,03	0,04	0,07
num_rev_accts	**-0,45**	**-0,38**	**-0,32**	**-0,29**	**-0,22**
num_sats	**-1,34**	**-1,26**	**-1,22**	**-1,14**	**-1,11**
num_tl_90g_dpd_24m	-0,04	0,00	0,00	0,04	0,00
pct_tl_nvr_dlq	**0,42**	**0,37**	**0,34**	**0,29**	**0,28**
tax_liens	-0,02	-0,01	-0,01	-0,01	-0,01
total_bal_ex_mort	**0,18**	**0,18**	**0,15**	**0,15**	**0,17**
total_bc_limit	**1,07**	**1,02**	**0,96**	**0,93**	**0,87**

Table A1: (Continued)

	τ			
Variable	0.80	0.85	0.90	0.95
funded_amnt	**-0,27**	**-0,27**	**-0,24**	**-0,23**
term	**4,61**	**3,64**	**2,68**	**1,48**
int_rate	**3,02**	**2,61**	**2,11**	**1,53**
emp_length1 year	-0,16	-0,04	0,02	-0,12
emp_length2 years	**-0,53**	**-0,42**	**-0,33**	**-0,33**
emp_length3 years	**-0,39**	**-0,45**	**-0,30**	**-0,32**
emp_length4 years	**-0,46**	**-0,33**	-0,14	-0,11
emp_length5 years	**-0,54**	**-0,35**	**-0,25**	-0,13
emp_length6 years	**-0,77**	**-0,62**	**-0,50**	**-0,48**
emp_length7 years	**-1,37**	**-1,07**	**-0,66**	**-0,39**
emp_length8 years	**-1,20**	**-1,03**	**-0,96**	**-0,81**
emp_length9 years	**-1,17**	**-0,97**	**-0,92**	**-0,65**
emp_length10+ years	**-1,03**	**-0,86**	**-0,67**	**-0,54**
home_ownershipOWN	**0,89**	**0,86**	**0,78**	**0,61**
home_ownershipRENT	**0,43**	**0,45**	**0,38**	**0,29**
annual_inc	**0,69**	**0,60**	**0,51**	**0,41**
verification_statusSource Verified	**-0,34**	**-0,32**	**-0,24**	**-0,23**
verification_statusVerified	**-0,36**	**-0,32**	**-0,25**	**-0,22**
purposehouse_related	**1,68**	**1,54**	**1,38**	**1,03**
purposemajor_purchase	**2,57**	**2,41**	**1,86**	**1,44**
purposemedical	**3,31**	**3,02**	**2,66**	**2,12**
purposesmall_business	**0,98**	**0,97**	**0,66**	**0,48**
purposevacation	**2,19**	**1,87**	**1,74**	**1,28**
purposecar	0,66	0,44	**0,97**	**1,02**
purposeother	**2,10**	**1,77**	**1,61**	**1,42**
dti	**0,13**	**0,06**	0,00	**-0,09**
delinq_2yrs	0,05	0,06	0,05	0,04
fico_range_high	**0,37**	**0,39**	**0,37**	**0,34**
inq_last_6mths	**0,10**	**0,13**	**0,13**	**0,14**
pub_rec	**-0,22**	**-0,19**	**-0,18**	**-0,10**
revol_bal	**0,19**	**0,14**	0,04	0,05
revol_util	**-0,92**	**-0,80**	**-0,69**	**-0,50**
collections_12_mths_ex_med	**0,09**	**0,09**	**0,06**	0,03
application_typeJoint App	**4,28**	**3,43**	**2,72**	**1,88**
tot_coll_amt	-0,07	**-0,05**	-0,03	0,01
acc_open_past_24mths	**0,32**	**0,30**	**0,23**	**0,13**
avg_cur_bal	**0,23**	**0,22**	**0,20**	**0,13**
chargeoff_within_12_mths	**0,11**	**0,09**	**0,07**	**0,07**
delinq_amnt	0,03	0,03	0,01	0,00
mo_sin_old_rev_tl_op	**-0,10**	**-0,08**	**-0,08**	**-0,07**
mo_sin_rcnt_rev_tl_op	**0,37**	**0,31**	**0,26**	**0,18**
mo_sin_rcnt_tl	**-0,11**	-0,05	0,01	**0,08**
mort_acc	**-0,16**	**-0,13**	**-0,11**	**-0,07**
num_accts_ever_120_pd	**0,08**	**0,05**	**0,04**	0,00
num_rev_accts	**-0,20**	**-0,16**	**-0,11**	-0,05
num_sats	**-0,98**	**-0,92**	**-0,78**	**-0,58**
num_tl_90g_dpd_24m	-0,03	0,01	0,02	0,04
pct_tl_nvr_dlq	**0,26**	**0,24**	**0,22**	**0,17**
tax_liens	-0,06	-0,06	-0,01	0,01
total_bal_ex_mort	**0,14**	**0,15**	**0,14**	**0,09**
total_bc_limit	**0,77**	**0,69**	**0,62**	**0,44**


A2: Correlation Matrix Heatmap of Continuous Variables
total_bc_limit
total_bal_ex_mort
tax_liens
pct_tl_nvr_dlq
num_tl_90g_dpd_24m
num_sats
num_rev_accts
num_accts_ever_120_pd
mort_acc
mo_sin_rcnt_tl
mo_sin_rcnt_rev_tl_op
mo_sin_old_rev_tl_op
delinq_amnt
chargeoff_within_12_mths
avg_cur_bal
acc_open_past_24mths
tot_coll_amt
collections_12_mths_ex_med
revol_util
revol_bal
pub_rec
inq_last_6mths
fico_range_high
delinq_2yrs
dti
annual_inc
int_rate
funded_amnt
Variables
funded_amnt
int_rate
annual_inc
dti
delinq_2yrs
fico_range_high
inq_last_6mths
pub_rec
revol_bal
revol_util
collections_12_mths_ex_med
tot_coll_amt
acc_open_past_24mths
avg_cur_bal
chargeoff_within_12_mths
delinq_amnt
mo_sin_old_rev_tl_op
mo_sin_rcnt_rev_tl_op
mo_sin_rcnt_tl
mort_acc
num_accts_ever_120_pd
num_rev_accts
num_sats
num_tl_90g_dpd_24m
pct_tl_nvr_dlq
tax_liens
total_bal_ex_mort
total_bc_limit
Variables
Correlation
1.0
0.5
0.0
−0.5
−1.0

References

Acumen Research and Consulting. (2023). *P2P Lending Market Size to Hit USD 804.2 Billion by 2030*. [online] Available at: https://www.acumenresearchandconsulting.com/p2p-lending-market.

Allied Market Research. (2023). *Peer to Peer (P2P) Lending Market Size, Share, Competitive Landscape and Trend Analysis report by Business Model, by Loan Type, by End User: Global*

Aleksandrova, Y. (2021). Comparing Performance of Machine Learning Algorithms for Default Risk Prediction in Peer to Peer Lending. *TEM Journal*, pp.133–143. doi:https://doi.org/10.18421/tem101-16.

Basha, S., Elgammal, M.M. and Abuzayed, B.M. (2021). Online peer-to-peer lending: A review of the literature. *Electronic Commerce Research and Applications*, 48(2), p.101069. doi: https://doi.org/10.1016/j.elerap.2021.101069.

Bednorz, J. (2020). *The History of P2P Lending*. [online] P2PMarketData. Available at: https://p2pmarketdata.com/articles/p2p-lending-history/.

Bholat, D.M. and Atz, U. (2016). Peer-to-Peer Lending and Financial Innovation in the United Kingdom. *SSRN Electronic Journal*. doi:https://doi.org/10.2139/ssrn.2772585.

CFI. (2022). *Peer-to-Peer Lending*. [online] Corporate Finance Institute. Available at: https://corporatefinanceinstitute.com/resources/wealth-management/peer-to-peer-lending/.

Chen, C.W.S., Dong, M.C., Liu, N. and Sriboonchitta, S. (2019). Inferences of default risk and borrower characteristics on P2P lending. *The North American Journal of Economics and Finance*, 50, p.101013. doi:https://doi.org/10.1016/j.najef.2019.101013.

Meltzer, C. (n.d.). *How to Beat a Lending Club Charge Off | SoloSuit Blog*. [online] Available at: https://www.solosuit.com/posts/beat-lending-club-charge-off [Accessed 15 May. 2024].

Davino, C., Furno, M. and Vistocco, D. (2014). *Quantile regression*. Chichester, West Sussex: Wiley.

Davino, C., Romano, R. and Vistocco, D. (2022). Handling multicollinearity in quantile regression through the use of principal component regression. *METRON*. doi:https://doi.org/10.1007/s40300-022-00230-3.

Dermineur, E.M. (2019). Peer-to-peer lending in pre-industrial France. *Financial History Review*, pp.1–30. doi:https://doi.org/10.1017/s0968565019000143.

FasterCapital. (n.d.). *Peer to Peer Lending: The Future of Borrowing and Investing*. [online] Available at: https://fastercapital.com/content/Peer-to-Peer-Lending--The-Future-of-Borrowing-and-Investing.html [Accessed 8 May 2024].

Finansinspektionen. (2024). *Räntan pressar högt skuldsatta hushåll*. [online] Available at: https://www.fi.se/sv/publicerat/rapporter/bolanerapporter/den-svenska-bolanemarknaden-2024/ [Accessed 7 May 2024].

Fintech Nexus. (n.d.). *Lending Club Review for New Investors*. [online] Available at: https://www.fintechnexus.com/lending-club-review-at-for-new-investors/.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The elements of statistical learning : data mining, inference, and prediction*. 2nd ed. New York: Springer.

Kagan, J. (2024). *What is Peer-to-Peer (P2P) Lending?*. [online] Investopedia. Available at: https://www.investopedia.com/terms/p/peer-to-peer-lending.asp.

Kaggle.com. (2019). *All Lending Club loan data*. [online] Available at: https://www.kaggle.com/datasets/wordsforthewise/lending-club/data.

Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1), p.33. doi:https://doi.org/10.2307/1913643.

Koenker, R. (2001). Linear Hypothesis: Regression (Quantile). *International Encyclopedia of the Social Behavioral Sciences,* pp. 8893-8899 doi: https://doi.org/10.1016/B0-08-043076-7/00456-3.

Koenker, R and F. Hallock, K. (2001). Quantile Regression. *The Journal of Economic Perspectives*, 15(4), pp. 143-156. doi: https://doi.org/10.1257/jep.15.4.143

LendingClub. (2019). *LendingClub Reports Fourth Quarter and Full Year 2018 Results.* [online] Available at: https://ir.lendingclub.com/news/news-details/2019/LendingClub-Reports-Fourth-Quarter-and-Full-Year-2018-Results/default.aspx

Lenz, R. (2016). Peer-to-Peer Lending: Opportunities and Risks. *European Journal of Risk Regulation*, 7(4), pp.688–700. doi:https://doi.org/10.1017/s1867299x00010126.

Mckinsey.com. (n.d.). *What today's shake-out in China's peer-to-peer lending market means for fintech | McKinsey & Company*. [online] Available at: https://www.mckinsey.com/cn/our-insights/perspectives-on-china-blog/what-todays-shakeout-in-chinas-peer-to-peer-lending-market-means-for-fintech.

Milne, A. and Parboteeah, P. (2016). The Business Models and Economics of Peer-to-Peer Lending. *SSRN Electronic Journal*, (17). doi:https://doi.org/10.2139/ssrn.2763682.

Molenius, H. (2024). *Svenska banker gör rekordvinster – men det finns kanariefåglar i gruvan*. [online] Affärsvärlden. Available at: https://www.affarsvarlden.se/kronika/svenska-banker-gor-rekordvinster-men-det-finns-kanariefaglar-i-gruvan?fbclid=IwZXh0bgNhZW0CMTAAAR3pi2gKqJoGkebx9254AFxCFhhMnlBdDDQ4OgP4S9-

pYyvQQR138ftSFUE_aem_AcQIXIv9g2f2TpKqWpGhbcNqFJmHWiYynnijTY_au2mTAPMvZjP87TEOUJo0jb0-Uwq7ht09xvZt02H4yODYqmo- [Accessed 4 May 2024].

Nowak, A., Ross, A. and Yencha, C. (2017). SMALL BUSINESS BORROWING AND PEER-TO-PEER LENDING: EVIDENCE FROM LENDING CLUB. *Contemporary Economic Policy*, 36(2), pp.318–336. doi:https://doi.org/10.1111/coep.12252.

Opportunity Analysis and Industry Forecast, 2023-2032. [online] Available at: https://www.alliedmarketresearch.com/peer-to-peer-lending-market.

Polena, M. and Regner, T. (2018). Determinants of Borrowers' Default in P2P Lending under Consideration of the Loan Risk Class. *Games*, 9(4), p.82. doi:https://doi.org/10.3390/g9040082.

Powell, J.L. (1986). Censored regression quantiles. *Journal of Econometrics*, 32(1), pp.143–155. doi:https://doi.org/10.1016/0304-4076(86)90016-3.

Proud, L. (2021). *Peer-to-peer lending's demise is cautionary tale.* [online] Reuters. Available at: https://www.reuters.com/markets/asia/peer-to-peer-lendings-demise-is-cautionary-tale-2021-12-13/

Resolve. (2019). *What happens when you default on your Lending Club debt?* [online] Available at: https://www.helloresolve.com/debt-management/what-happens-when-you-default-on-your-lending-club-debt/.

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

Setiawan, N., Suharjito and Diana. (2019). A Comparison of Prediction Methods for Credit Default on Peer to Peer Lending using Machine Learning. *Procedia Computer Science*, 157, pp.38–45. doi:https://doi.org/10.1016/j.procs.2019.08.139.

Stock, J.H. and Watson, M.W. (2015). *Introduction to econometrics.* Boston: Pearson

Svt.se. (2023). S: Bankernas vinster provocerar. *SVT Nyheter*. [online] 18 Jul. Available at: https://www.svt.se/nyheter/ekonomi/s-om-bankerna [Accessed 7 May 2024].

Tang, M., Mei, M., Li, C., Lv, X., Li, X. and Wang, L. (2020). How does an individual's default behavior on an online peer-to-peer lending platform influence an observer's default intention? *Financial Innovation*, 6(1). doi:https://doi.org/10.1186/s40854-020-00197-y.

Xia, Y., He, L., Li, Y., Liu, N. and Ding, Y. (2019). Predicting loan default in peer-to-peer lending using narrative data. *Journal of Forecasting*, 39(2), pp.260–280. doi:https://doi.org/10.1002/for.2625.

Yamashita, T., Yamashita, K. and Kamimura, R. (2007). A Stepwise AIC Method for Variable Selection in Linear Regression. *Communications in Statistics - Theory and Methods*, 36(13), pp.2395–2403. doi:https://doi.org/10.1080/03610920701215639.

Ziegler, T., Shneor, R., Wenzlaff, K., wang, britney, Kim, J., Paes, F.F. de C., Suresh, K., Zhang, B.Z., Mammadova, L. and Adams, N. (2020). The Global Alternative Finance Market Benchmarking Report. *SSRN Electronic Journal*. doi:https://doi.org/10.2139/ssrn.3771509.