Proceedings of the 3rd International Conference on Business and Policy Studies
DOI: 10.54254/2754-1169/106/20241482

Analysis of Interest Rate Determinants: Evidence from Lending Club

Yanhang Zou[1,a,*]

[1]*Department of Statistics, University of Toronto, St. George, Toronto, ON, M5G 1X6, Canada*
a. yanhang.zou@mail.utoronto.ca
**corresponding author*

Abstract: Nowadays, high interest rates hit the whole lending market, a number of people consider online lending institutions to trade. This dynamic generates a viable alternative to traditional bank services, allowing borrowers to choose their optimal lending plan. This paper focuses on online lending services, investigating the factors influencing the interest rate offered by Lending club institutions. Using OLS and machine learning models, the author analyses the influence of different dimensions of factors (debt level, FICO score, lending purpose, and normalized debt-to-income ratio) on the average interest rate. The results emphasize the results emphasize that Personal characteristics count for lending rate evaluation. People who have lower credit scores and bear a high debt ratio would be offered a higher interest rate than those who have a good individual background. Contrary to the conclusions achieved in other studies, when loans are given for longer durations, they tend to have higher interest rates. This is mainly because lenders have to consider factors like the risk of inflation and the potential earnings they might miss out on during that extended period.

Keywords: Lending Characteristics, FICO, Interest Rate, Lending Institutions, Machine Learning

1. Introduction

Currently, the emergence of online P2P finance provides a legal platform to facilitates financial transaction [1]. Those P2P lending institutions connect individual borrowers with individual investors directly, bypassing traditional financial intermediaries such as banks [2]. Specifically, the online P2P lending industry attracts the attention of global academics. Traditionally, commercial bank rates follow central bank rates, with less flexibility on interest rate for other institutions to adjust [3]. Most research regarding interest rate (benchmark rates) determinants has been explored in macroeconomic considerations and shown its significance [4]. However, whether the conclusion can be promoted to the institution's interest rate is uncertain. Following US hike, individuals have exhibited heightened concern regarding interest rates, leading to less passion for traditional bank lending [5]. In the realm of peer-to-peer (P2P) lending, the market could serve as a substitute to the conventional bank lending market [6,7]. Therefore, the market size of P2P lending platforms has surged drastically recently.

Usually, the interest rate dominates the financial markets such as debts and exchange rate when people decided to participate the financial activities. Similarly, following the monetary policy, interest rate also shows the crucial effect on shaping an economy's performance [8]. Therefore, investigating the interest rate determinants of online lending institution can provide guidelines on

DOI: 10.54254/2754-1169/106/20241482

managing economic conditions decisions. As a representative of online lending firm, LendingClub is a successful player in the global market with a substantial influence. It provides an opportunity for borrowers who might frequently be deemed ineligible for loans through traditional banking services to access credit [9]. LendingClub, as one of the pioneering successful platforms in the peer-to-peer lending sector, draws the attention from investors and venture capitalists, leading to multiple rounds of funding to support its expansion [1]. LendingClub comply with various state and federal regulations to ensure consumer protection and maintain its operations within the bounds of the law. Therefore, people trust this institution and are willing to participate in this lending-based crowdfunding activity [10].

Most existing literature contributes the interest rate determinant into the macro environment, such as Open Market Operations and Quantitative Easing methods. Researchers also regard interest rate as a consequence to measure the global economy. However, there has been a limited amount of research conducted on the comprehensive assessment of micro lending records, which considers lender behaviour would apply on offering different interest rates. To address the research gap, in this article, we will explore interest-rate-setting drivers of LendingClub, a company that emphasizes the significant standing in the industry by leveraging its expertise in assisting a wealth of small business with IPO [11]. Therefore, the large sample data from official website would make the conclusion strongly supported. Specifically, independent information recording ensures the financial professionalism of the data, reduces the impact of macroeconomic factors, and provides accuracy in studying microvariables.

Hence, this research aims to build a model to identify the comprehensive factors influencing interest rate on LendingClub. In this study, we consider six main factors: debt level; credit status; lending, lending purpose; revolving line utilization rate; social history. And we take a public disclosure dataset of interest rate listings and employ ordinary least squares (OLS) and machine learning operations to quantitatively estimate best-fitted interest rate equations.

The paper is structured as follows. Section 2 provides a literature review of the existing research about relevant interest rate drivers of demands. Section 3 illustrates the methods and data. Section 4 demonstrates the virtualization of some potential LendingClub listing interest rate indicators. Section 5 lists the results and analysis of OLS regression and Machine learning. Section 6 expresses the discussion and limitations of the results. Section 7 describes the concluding remarks.

2. Literature Review

This research intends to robustly analyze the interest rate determinants of LendingClub listings. Firstly, we note that debt levels have been shown to affect the lending interest rate. Based on the research by the Emekter team, they discovered that lenders might consider borrowers with elevated debt levels as riskier, resulting in higher interest rates being applied to offset the higher probability of loan default [12]. However, citation points that borrowers probably characterized by elevated debt levels might opt to pledge collateral or security as a means to secure the loan from the lending institution [13]. This act of providing collateral serves to mitigate the lender's risk exposure, leading the lender to potentially offer a reduced interest rate in return for the added security. The rate is also related to the FICO (credit score). Normally, borrowers with high FICO values indicate a more favorable credit profile and responsible financial behavior [14]. However, limited credit history might deviate from risk assessment through FICO value. In instances where a borrower possesses a high FICO score yet has a constrained credit history, lenders might perceive them as posing a higher risk in comparison to borrowers with more extensive credit backgrounds. Consequently, due to the limited availability of comprehensive credit data for assessment, lenders may impose slightly elevated interest rates [15]. The notion of lending purpose represents the loan flow direction. Normally,

DOI: 10.54254/2754-1169/106/20241482

lending for purchasing property or equipment, often involves collateral that can be seized by the bank in case of default. And loans for such purposes would be relatively lower than those to invest.

The other research gap would be related to the analysis of endogenous factors on interest rate. In the prevailing scholarly discourse, loan interest rates are typically assessed through static indicators like personal social background and creditworthiness. Nonetheless, in this study, given the inherent information asymmetry prevalent in online loans, the incorporation of dynamic indicators, such as revolving rate, is advocated to mitigate the risk of information asymmetry. Despite of limited existing research on this topic, through OLS and machine learning methods, we aim to provide confirmatory evidence about the effectiveness for several aspects on LendingClub interest rate.

3. Data and Methodology

3.1. Data and Context

The data used in this analysis contains information about the lending records of LendingClub and obtained from Kaggle ‘Lending Club Loan Data from 2007 to 2015’, which was originally gained from the public LendingClub website. First, we scan the whole dataset and find non-null data inside and the number of data remains consistent. Although data collection spanned eight years, it still follows a Cross-Sectional Data approach because time is muted. Next, given our investigation into the determinants of interest rates, establishing causality is crucial. This study aims to identify a set of factors influencing interest rates. In alignment with fundamental financial causality principles, we excluded the variable 'installment' since changes in interest rates are the sole drivers of installment variations. And then, we focus on a pair of similar indexes: ‘log_income ’and ‘di’. Given its capacity to precisely gauge the borrower's debt load their income, the debt-to-income ratio offers a more lucid indication of the borrower's aptitude to handle supplementary debt responsibilities [16]. Hence, we drop ‘log_income ’and keep ‘di ’as a unique debt level index. Similarly, we also maintain the 'revolving line utilization rate ’and remove the ‘revolving balance’. After that, the column ‘not_fully_paid ’describes whether the loan is paid which is irrelevant in our study, so, we also give up this measurement index. Since most of the listings record the days of lending loans as lower than 5730, which is approximately 15 years, we decided to shift the days to the year to shrink the unit. We conduct ‘year_with_cr_line ’to see the term level, which the divided into 365 of each listing. This resulted in a final dataset with 9578 observations of each lending record that includes 10 different attributes.

Since we plan to establish a model and use the same data to check the model availability, we randomly divided the whole dataset as half and half, namely 50% as train data to build models and 50% as test data to fit the model.

3.2. Methodology

3.2.1. Data Description

- **credit_policy:** whether the customer meets the credit underwriting criteria of LendingClub website.
- **purpose:** The purpose of the loan.
- **dti_normalization:** The normalized debt-to-income ratio of the borrower (amount of debt divided by annual income).
- **fico:** The FICO credit score of the borrower.
- **log_days:** The log number of days the borrower has had a credit line.

- **revol_util:** The borrower's revolving line utilization rate (the amount of the credit line used relative to total credit available).
- **inq_last_6mths:** The borrower's number of inquiries by creditors in the last 6 months.
- **delinq_2yrs:** The number of times the borrower had been 30+ days past due on a payment in the past 2 years.
- **pub_rec:** The borrower's number of derogatory public records (bankruptcy filings, tax liens, or judgments).

According to the different lending purpose, ’purpose ’contains six types of purpose: credit card; debt consolidation; educational; major purchase; small business; and others. Therefore, setting dummy variables would be easy to compare different lending purpose on the rate setting effect. Due to the six levels of purpose, we choose five variables named ‘credit_card’; ‘debt_ consolidation’; ‘educational’; ‘major_purchase’; ‘small _business’. If the recordings are signed as credit card purpose, the value of “credit_card” would be 1, otherwise 0. Same as other variables. If the purpose is “others”, both five index would be 0.

3.2.2. OLS Regression

To examine the interest rate setting behaviour of LendingClub, we apply ordinary least squares (OLS) approach in rate-setting regression models. The baseline model is a linear regression model (LM) with selected features as explanatory variables. This method assumes that a characteristic vector can represent each aspect factor under consideration. Besides, we also take the log on days with credit line to find the how term floating by each unit change in rate change. On the other hand, large range of the days holding would produce outliers to those endpoints data, and logged days would tolerate range wise dataset.

$$int\ rate = \beta_0 + \beta_i X_i + \varepsilon_i$$

where $i = 1, \ldots, n$; and n is the number of observations.

where ***int_rate*** represents the LendingClub interest rate of ***i*** ; $\boldsymbol{X_i}$ refers to a set of characteristics that are associated with a particular recordings, perhaps be measured in dummy or levels; $\boldsymbol{\beta_0}$ is the intercept; $\boldsymbol{\beta_i}$ is the parameter of each indicator and ε_i is a random error term with the usual properties. The partial derivative of the interest rate setting function with respect to how much each listing characteristic that contributes to the interest rate change. Particularly, those attributes, ***Xi***, can be roughly divided into six categories: (1) debt level (e.g., debt-to-income ratio); (2) credit status (e.g.,FICO, credit policy satisfaction); (3) social history (e.g., public records, number of inquiries); (4) time (e.g., days with credit line); (5) dynamic endogeneity (revolving line utilization rate); and (6) purpose (educational or purchasing). And we expected that debt level, lending term, number of inquiries and small business would have a positive effect and others such as FICO, credit policy would have a negative effect on the rate-setting mechanism.

The final model is presented as follows:

$$\begin{aligned} int\ rate = {} & \beta_0 + \beta_1 * FICO + \beta_2 * dti\ normalized + \beta_3 * logdays + \beta_4 * credit\ policy + \beta_5 \\ & * revolving\ rate + \beta_6 * credit\ card + \beta_7 * debt\ consolidation \quad + \beta_8 \\ & * eductaion\ + \beta_9 * major\ purchase + \beta_{10} * inquiries + \beta_{11} * small\ business \\ & + \beta_{12} * public\ records + \beta_{13} * decq\ 2yrs + \varepsilon \end{aligned}$$

After constructing a comprehensive model using OLS regression, we observe that certain predictor variables do not exhibit statistical significance at the 95% confidence level, as indicated by their p-values exceeding 0.05. For instance, the variables 'public_records' and 'decq_2yrs' possess p-values of 0.725 and 0.627, respectively, signifying their lack of significance in the regression model

DOI: 10.54254/2754-1169/106/20241482

(Specification 1). Consequently, we employ both forward and backward variable selection methods to refine the model. Ideally, both approaches should yield equivalent outcomes **(Specification 2)**. The results of these selection methods show that all variables are significant within the 95% confidence interval. However, the presence of multicollinearity undermines the reliability of the regression results. To address this issue, we generate a correlation matrix to explore potential correlations among the variables. This analysis reveals two pairs of variables: 'FICO' & 'revolving rate' and 'inquiries' & 'credit policy'. Subsequently, we incorporate these two sets of variables into the regression model obtained from the variable selection **(Specification 3)**. Furthermore, we employ Lasso and Ridge regression to identify the best fit model. Although Lasso and Ridge regression differ from OLS regression, we evaluate their performance using mean squared error (MSE), mean absolute error (MAE), and R-square (R^2) as empirical measures of model fit.

3.2.3. Machine Learning

In this paper, we focus on three main machine learning methods: decision tree, bootstrapping, and Random Forest. Initially, we employ the decision tree to predict a continuous target variable through a set of input features. We prepare the same indicators that are chosen in the OLS regression to the regression tree. And we would get the objective function. Based on objective function, we can calculate the sum of the squared differences between the predicted and actual values of the target variables. The goal is to determine the split that minimizes SSR at each node. The algorithm evaluates all possible splits on the interest rate of all recording determinants at each node of the tree and selects the one that results in the greatest decrease in SSR. The difference between the SSR before the split and the total of the SSR after the split is used to determine the reduction in SSR. Then, we form a decision tree and get the MSE value. Next, we apply the bootstrapping method to find the best fit model. And we also collected the MSE value of the model in bagging fitness. And then we fit the bootstrapping model to compare the prediction model and true model fitness. After that, we use the Random Forest methods to predict the model and we produce a variable importance plot. Random forest algorithm one of the ensembles learning techniques that integrate predictions from decision trees [17]. It is a bagging strategy that allows each tree to run autonomously before aggregating the results at the end without giving any tree preference. Finally, linked with advantages and disadvantages in each method, we compare the MSE, MAE and R^2 value of among three methods. Since the same data structure, we also compare those three index value to the OLS regression. And then, we choose the OLS model **(Specification 3)** that fully considering the applicability of various models.

DOI: 10.54254/2754-1169/106/20241482

4. Data Visualization

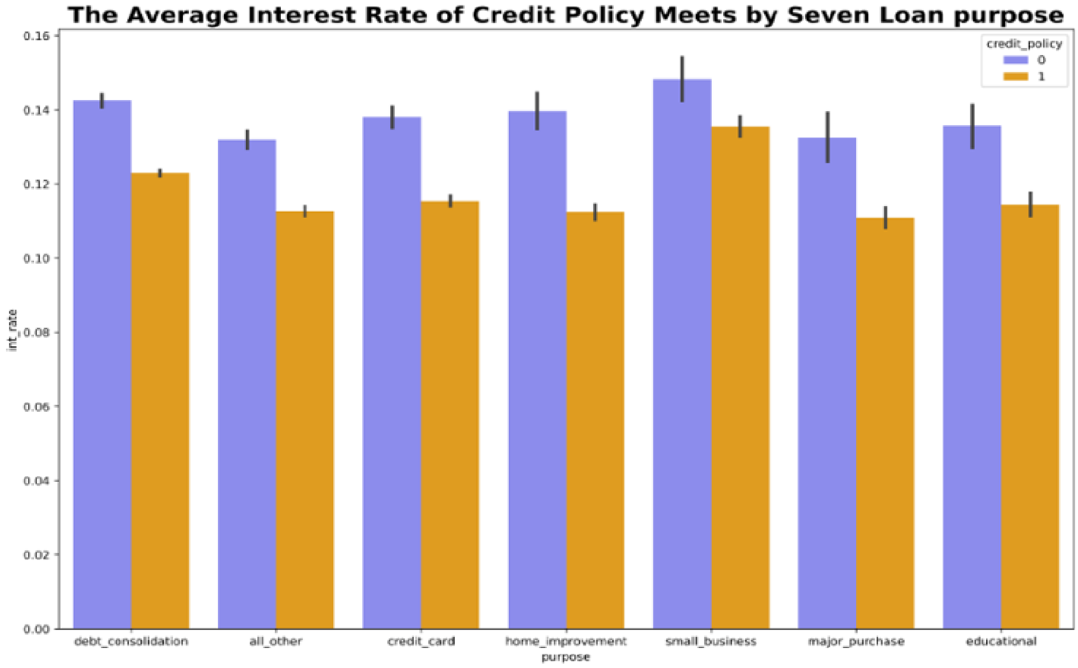

Figure 1: The average interest rate of credit policy implementations across seven loan purposes

Generally, our findings suggest that adhering to credit policy guidelines generally results in lower interest rates across all loan purposes. Moreover, the interest rates for records meeting the credit policy criteria tend to be relatively lower than those for records that do not comply with the policy. Small business loans consistently exhibit the highest interest rates, regardless of credit policy adherence. Specifically, small business loans failing to meet the credit policy criteria have an average interest rate of approximately 0.15, whereas those meeting the policy have an average rate of 0.13. Loans for purchase purposes demonstrate the lowest interest rates for both compliance categories, with rates of 0.13 and 0.11, respectively.

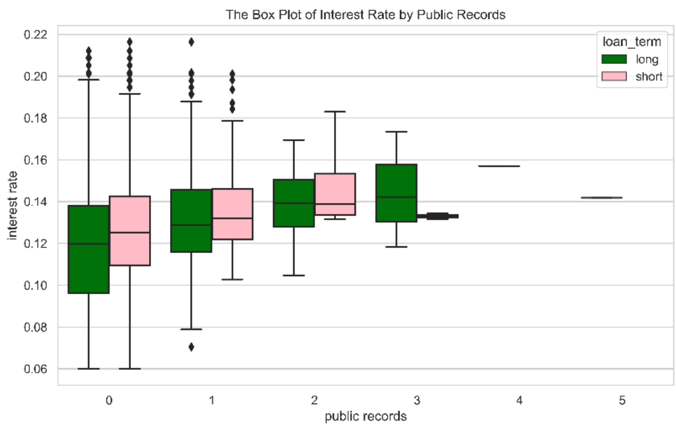

Figure 2: the distribution of interest rates for two types of loans, namely long-term and short-term loans, across various public record categories such as bankruptcy filings, tax liens, or judgments. Since we discover half of the recordings has over 10 years, we classify the loan term over 10 years as ‘long-term’, otherwise ‘short-term’.

DOI: 10.54254/2754-1169/106/20241482

We observe that the interest rates for long-term loans tend to be relatively lower compared to short-term loans (represented by the green and pink boxes, respectively). Within the long-term loan category (green box), an increasing trend is noticeable in interest rates as the number of public records listings rises. Specifically, for loans without any public record listings, approximately half of the interest rates for short-term and long-term loans cluster around 0.12 and 0.13, respectively. There are potential outliers in the data, represented by individual data points beyond the whiskers, with interest rates exceeding 0.2.

5. Results and Analysis

5.1. OLS Regression

First, we conduct an ordinary multiple linear regression model **(Table 1, Specification (1))**. Based on the strong linear relation on scatter plot related to the FICO and price, we can ensure the feasibility on linearity assumption. Besides, most relative factors are considered in the linear regression, which ensures LSA 1 to be valid.

Table 1: Specification Table

Variable	(1) full	(2) Back/forward	(3) with intersection
const	0.4232*** (0.007)	0.4238*** (0.006)	0.3605*** (0.009)
fico	-0.0005*** (9.6e-06)	-0.0005*** (8.94e-06)	-0.0004*** (1.21e-05)
dti normalized	0.0027** (0.001)	0.0027*** (0.001)	0.0038*** (0.001)
log days	0.0018*** (0.000)	0.0018*** (0.000)	0.0017*** (0.000)
credit policy	-0.016* (0.001)	-0.016*** (0.001)	—
inquiries in last 6 months	0.001*** (0.000)	0.001*** (0.000)	0.0011*** (0.000)
revolving utility	0.0001*** (1.13e-05)	0.0001*** (1.1e-05)	0.0020*** (0.000)
credit card	-0.0028*** (0.001)	-0.0028*** (0.001)	-0.0023*** (0.001)
debt consolidation	0.0018*** (0.001)	0.0018*** (0.001)	0.0022*** (0.001)
educational	-0.0001 (0.001)	—	—
major purchase	-6.405e-05 (0.001)	—	—
small business	0.0232*** (0.001)	0.0232*** (0.001)	0.0236*** (0.001)
public records	-0.0004*** (0.001)	—	—
delinq 2 years	0.0002 (0.000)	—	—
fico x revolving utility	—	—	-2.631e-06*** (2.49e-07)

DOI: 10.54254/2754-1169/106/20241482

Table 2: Model Fitness Summary Table

Method-Index	MSE	MAE	R-Square
(1)Full	0.0003	0.0135	0.5784
(2)Back/Forward	0.0003	0.0135	0.5783
(3)Interaction	0.00031	0.0139	0.587
(4)Ridge	0.00031	0.0139	0.5635
(5)Lasso	0.00035	0.0147	0.5076
(6)Decision Tree	0.00050	0.0173	0.2674
(7)Bootstrapping	0.00030	0.0129	0.6060
(8)Random Forest	0.00020	0.0128	0.6242

Generally, we find that three models hold significant parameters (with "***" behind). Compare three specifications, we can find that the specification (3) holds the largest R^2 and adjusted R^2, means higher (58.8%) volume of the Y (interest rate) in the observations would be explained by this model. Besides, it holds the smallest residual error, which shows the observed values (true interest rate) deviate from the predicted values (the predicted interest rate) by approximately 0.017. The results in Model fitness summary table (**Table 2**) indicate that specification (3) is the most appropriate among OLS regression. The adjust R^2 of the model is 0.587, which indicates 58.7% of the variability of interest rate that can be explained by the selected independent variables. The mean squared error is 0.00029. And the value of Durbin Watson (DW) statistic is 1.952, close to 2, shows that there is no autocorrelation detected in the model.

$$
\begin{aligned}
int\ rate = {} & 0.3605 - 0.0004 * FICO + 0.0038 * dti\ normalized + 0.0017 * logdays - 0.0023 \\
& * credit\ card + 0.0022 * debt\ consolidation + 0.0011 * inquies + 0.0020 \\
& * revolving\ rate + 0.0236 * small\ business - 2.631 * 10^{-6} \\
& * FICO\ X\ revolving\ rate
\end{aligned}
$$

The specification (3) shown above identifies that each point increase in the debt-to-income ratio score between the recordings is associated with 0.04% shrink on the interest rate. The intercept β_0 is 0.3605, keep all else indicator to be 0, indicating if the recording with no debt and credit score, no term requirement on zero inquiries and public records, the interest rate of lending in other purpose would be 36.05% on average. The p-value is 0, and thus we reject the null hypothesis that $\beta_1 = 0$ at 5% level of significance. It emphasizes that FICO point is significantly negatively correlated with the interest rate. We consider this result to be economically significant as one point higher in the FICO is reasonable if prime rates for high credit borrowers and well demonstrated creditworthiness is attainable. Specifically, the lower FICO owner might have limited negotiating power, which indicates they would have limited bargaining opportunities from lenders to advocate for better loan terms. As a result, the institution may need to provide subprime or high-interest rate loans for users to lower the default risk.

The OLS results **(Table 1)** indicate that all indicators falling under lending-purpose and debt-level attributes are significant determinants of the rate-setting strategy. The coefficient for credit card repayment is -0.0023, suggesting that the recording rate would be 0.23% lower for credit card repayments compared to non-credit-card repayments. Similarly, the coefficient for small business opening and expansion is 0.0236, indicating a 2.36% higher rate for business loans compared to other types of loans. These rate differences may be attributed to varying risk profiles, where business loans are considered riskier due to larger loan amounts and variable cash flows. The normalized debt-to-income ratio shows a positive relationship with the rate, with a 0.38% average increase in rate for each additional unit increase in the ratio. The p-value of 0.01 indicates the debt level's significance

DOI: 10.54254/2754-1169/106/20241482

as a positive determinant of the recording rate. Elevated debt levels may strain the borrower's capacity to manage additional debt, leading to lender concerns about loan payment promptness.

Surprisingly, we discover that each one hundred percent change in the number of days contributes to a 0.17% increase in the listing interest rate, which contrasts our initial expectation that longer loan durations would lead to proportionally lower rates. This unexpected result can be attributed to potential inflation risk associated with longer lending periods. Lenders may consider inflation risk when determining interest rates for longer-term loans. On a macroeconomic level, inflation erodes the purchasing power of money over time, prompting lenders to charge higher rates to compensate for potential decreases in the loan's real value. Micro-economically, longer lending periods entail higher potential opportunity costs for lenders. By allocating funds to long-term loans, lenders forgo the chance to invest in potentially more profitable ventures. Higher interest rates can help mitigate this opportunity cost.

5.2. Machine Learning

5.2.1. Regression Tree

Then, we put all variables that are chosen in the OLS regression to the regression tree. The objective function shows below:

$$\begin{aligned}\underset{tree \subset T}{min} \sum (& \beta_0 + \beta_1 * FICO_i + \beta_2 * dti_i + \beta_3 * year\ with\ cr\ line_i + \beta_4 * not\ fully\ paid_i \\ & + \beta_5 * credit\ policy_i + \beta_6 * inquiries_i + \beta_7 * revolving\ rate_i + \beta_8 \\ & * credit\ card_i + \beta_9 * debt\ consolidation_i + \beta_{10} * eductaion_i + \beta_{11} \\ & * major\ purchase + \beta_{12} * small\ business_i + \beta_{13} * public\ records_i + \beta_{14} \\ & * decq\ 2yrs_i - \left(log(int\ rate_i)\right)^2 + \alpha |terminal\ \ nodes\ in\ tree|\end{aligned}$$

where the terminal nodes in tree in our regression tree equals to 12 (see the tree below). The maximum depth in the tree is 3. And "alpha" means a regularization parameter to control the complexity of the tree and prevent overfitting. For example, if we employ the Lasso regularization, the regularization parameter sets the intensity of the penalty on the sum of absolute values of the coefficients in the price predicted model. A higher value of 'alpha' results in more regularization, while a lower value allows the price predicted model to fit the data more closely. As a result, the interpretation of alpha is related to the price predicted model's sparsity, with greater alpha resulting in a sparser model. However, the optimal regularization parameter (alpha) we choose should balance the trade-off between model complexity and overfitting.

DOI: 10.54254/2754-1169/106/20241482

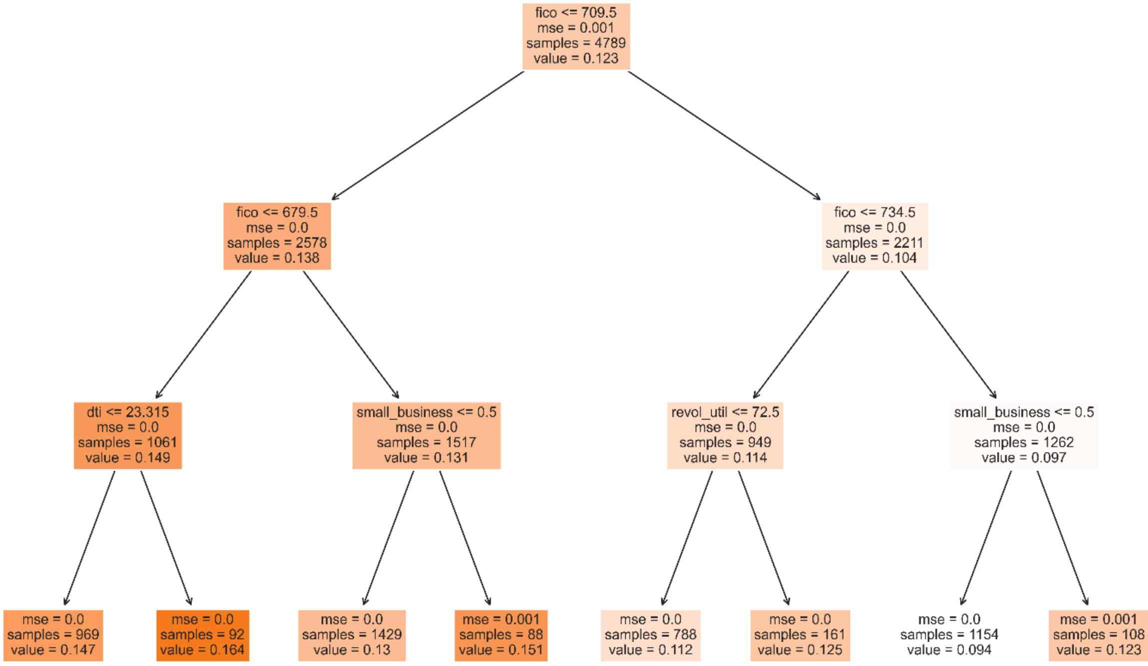

Figure 3: the decision tree in the LendingClub interest rate prediction model.

We can observe that the root (decision) node on the top. This node gathered full data that have 4789 training observations. Start from the root, it divided the whole training data into two parts: the fico score lower than 679.5 and lower than 734.5. If the fico score lower than 679.5, it would comes to the left branch; otherwise, right branch. Therefore, this division lead the tree to the second degree. In the left node, it has would have 1061 observations inside while the right branch has 1517 observations. Since our tree holds max depth for three, means we would have three degree of nodes. Each node divided the observations according to the conditions that yes to the left and no to the right. Eventually, we would have 2^3 = 8 leaf (in the terminant node).For example, there is 969 observations that holds fico score lower than 679.5 with the "dti" ratio less than 23.315, and in the regression, the interest rate value would be predicted as 0.147. And the value 0.147 for those data holds the min MSE value at 0.

5.2.2. Bootstrapping

Next, we apply the bootstrapping method to find the best fit model. The drawback for the single tree would relate to the high variance. And then we fit the bootstrapping model to compare the prediction model and true model fitness. Compare with the MSE value of bagging, OLS regression, and regression tree, we find that the MSE of bootstrapping method only 0.00028 which truly much lower than the regression tree and OLS regression.

5.2.3. Random Forest

Then, we use the Random Forest methods to predict the model. We know that FICO value is the most important variable in all features, with is over 40 Mean Decrease Gini. This means FICO score level largely determines on the interest rate setting. The X-axis expresses the mean decrease accuracy and represents how much accuracy the model losses when move each variable.

5.2.4. MSE & MAE & R square Comparison and Final Model Choose

Similarly, we also employ the three main model accuracy index to measure the model fitness.

From **Table 2,** the MSE and MAE values of Random Forest are 0.00027 and 0.0128, which are the lowest among other methods. Meanwhile, the R square of random forest would be relatively the highest, at 0.624, which means 62.4% of the data can be fit through the RF model. Based on three measurement value comparison, we can conclude that the machine learning prediction (Random Forest) would have better model fitness. However, the best model chosen would not only focus on the accuracy. Due to the slight difference with among the model compete and model promptness, we consider a the model that are more easy to understand. Overall, OLS regression would be better to explain each variable and would have more economical significance. Therefore, we still choose the linear regression model as our results.

6. Discussions and Limitation

Discoveries in this paper illustrate the significance of debt level, credit status, social history, lending term, lending purpose and dynamic revolving rate as interest rate setting determinants in LendingClub. Additionally, these findings are consistent with earlier studies with appropriate explanations. The highlight of the analysis is that we take the dynamic individual financial index into account as issues to affect rate setting. Generally, higher revolving rate would guide the interest rate goes up. A higher revolving line utilization rate signals that borrowers are utilizing a larger portion of their available credit, leading to a perception of heightened credit risk by lenders. Borrowers with elevated credit utilization are viewed as riskier due to their extensive usage of credit limits, potentially indicating financial strain or an increased probability of default. In response to this perceived higher risk, lenders may impose higher interest rates on borrowers with elevated revolving line utilization rates. Such elevated interest rates serve to offset the potential rise in default risk, enhancing the profitability of the loan for the lender and potentially mitigating the risk associated with heightened credit utilization. Therefore, the results (**Specification 3**) would provide some guidelines to the LendingClub users, managers and also policymakers. LendingClub users are encouraged to be vigilant in assessing their lending purpose and considering potential rate variations. Moreover, they can prioritize upholding a robust credit profile by ensuring punctual payments and responsible credit management, thereby fostering the possibility of securing more favorable loan terms. As for website managers, the cultivation of increased transparency is recommended, wherein comprehensive information is furnished regarding the impact of diverse factors on interest rates. By promoting transparency in rate-setting criteria, borrowers gain an understanding of the underlying reasoning behind the rates presented to them, thus empowering them to make well-informed decisions. The more aspects LendingClub officials inspect regarding to the users background details, the lower the risk of information asymmetry they would bear. Additionally, the future expansion of the LendingClub and its potential shocks on the traditional commercial bank industry may need to be concerned by financial market policymakers. Some cutting-edge research has mentioned the effect of online banking on lending performance [18]. To maintain a sustainable consistence for all stakeholders, policymakers will need to address these concerns.

Indeed, Drawbacks is inevitable in our results. First, the size of data set is not large. It would extend three different restriction of the study: imprecise data measurement, inaccuracy of old data, and statistical invalidation. For the data measurement imprecisely, due to the constrained data dimensions concerning the rate listings, there exists a potential for deviations in the OLS regression and machine learning predictions. The incorporation of additional descriptive and pertinent data related to the listings is anticipated to diminish the error term in both OLS and machine learning approaches. For the data accuracy, since the research draws upon data sourced from the LendingClub

DOI: 10.54254/2754-1169/106/20241482

online platform, encompassing loan records spanning the years 2007 to 2015. It cannot describe current lending environment. Employing more current data would enhance the pertinence and precision of the findings. For statistical assumption validation, the statistical power to identify meaningful associations or distinctions between variables is diminished. Limited data size would lead a heightened risk of Type II errors (False Negatives) occurring, as significant effects may go undetected due to the inadequate sample size.

Besides, Although the study identifies correlations between variables and interest rates, it may not establish a causal relationship. There could be additional omitted or unobservable factors that are influencing the interest rates, and thus, drawing definitive causal conclusions may be challenging.

7. Conclusion

The study delves into the factors influencing interest rates within the P2P lending platform, LendingClub, including debt level, FICO score, lending purpose, and normalized debt-to-income ratio. Notably, longer lending periods exhibit a positive correlation with higher interest rates, owing to inflation risk and opportunity costs faced by lenders. To estimate the interest rate equations, both traditional OLS regression and machine learning techniques are employed. While Random Forest offers superior predictive accuracy, OLS regression retains its significance in providing economic interpretation and comprehensibility. The findings contribute valuable insights into interest rate determinants in the P2P lending industry, emphasizing the significance of considering both micro and macro variables in rate-setting strategies. However, the research acknowledges certain limitations concerning data sources and model interpretability and suggests avenues for future research to further explore interest rate determination in the P2P lending market.

References

[1] *Moenninghoff, S. C., & Wieandt, A. (2013). The future of peer-to-peer finance. Schmalenbachs Zeitschrift Für Betriebswirtschaftliche Forschung, 65(5), 466–487.*

[2] *Bavoso, V. (2021). Financial intermediation in the age of Fintech: P2p lending and the reinvention of banking. Oxford Journal of Legal Studies, 42(1), 48–75. https://doi.org/10.1093/ojls/gqab022*

[3] *Friedman, B. M., & Kuttner, K. N. (2010). Implementation of monetary policy. Handbook of Monetary Economics, 1345–1438. https://doi.org/10.1016/b978-0-444-53454-5.00012-8*

[4] *Saeed, M. S. (2014). Bank-related, industry-related and macroeconomic factors affecting bank profitability: A case of the United Kingdom. Research journal of finance and accounting, 5(2), 42-50.*

[5] *Kim, J. (2023). Stock market reaction to US interest rate hike: Evidence from an emerging market. Heliyon, 9(5). https://doi.org/10.1016/j.heliyon.2023.e15758*

[6] *Tang, H. (2019). Peer-to-peer lenders versus banks: Substitutes or complements? The Review of Financial Studies, 32(5), 1900–1938. https://doi.org/10.1093/rfs/hhy137*

[7] *De Roure, C., Pelizzon, L., & Thakor, A. (2021). P2P lenders versus banks: Cream skimming or bottom fishing? The Review of Corporate Finance Studies, 11(2), 213–262.*

[8] *Obstfeld, M., & Rogoff, K. (1995). The Mirage of fixed exchange rates. Journal of Economic Perspectives, 9(4), 73–96. https://doi.org/10.1257/jep.9.4.73*

[9] *Li, W., Ding, S., Chen, Y., & Yang, S. (2018). Heterogeneous ensemble for default prediction of peer-to-peer lending in China. IEEE Access, 6, 54396–54406. https://doi.org/10.1109/access.2018.2810864*

[10] *Herrero-Lopez, S. (2009). Social Interactions in P2P lending. Proceedings of the 3rd Workshop on Social Network Mining and Analysis. https://doi.org/10.1145/1731011.1731014*

[11] *Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of default in P2P lending. PLOS ONE, 10(10). https://doi.org/10.1371/journal.pone.0139427*

[12] *Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. Applied Economics, 47(1), 54-70.*

[13] *Berger, A. N., & Udell, G. F. (1998). The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. Journal of banking & finance, 22(6-8), 613-673.*

DOI: 10.54254/2754-1169/106/20241482

[14] *Moulton, L. (2007, September). Divining Value with Relational Proxies: How Moneylenders Balance Risk and Trust in the Quest for Good Borrowers 1. In Sociological Forum (Vol. 22, No. 3, pp. 300-330). Oxford, UK: Blackwell Publishing Ltd.*
[15] *Barron, J. M., & Staten, M. (2003). The value of comprehensive credit reports: Lessons from the US experience. Credit reporting systems and the international economy, 8, 273-310*
[16] *Jin, Y., & Zhu, Y. (2015, April). A data-driven approach to predict default risk of loan for online peer-to-peer (P2P) lending. In 2015 Fifth international conference on communication systems and network technologies (pp. 609-613). IEEE.*
[17] *Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.*
[18] *Li, Y., Spigt, R., & Swinkels, L. (2017). The impact of FinTech start-ups on incumbent retail banks' share prices. Financial Innovation, 3(1), 1-16.*