

Increased interest rate is associated with increased the requested amount, loan length and debt-to-income ratio

1 Introduction

An interest rate is the rate at which interest is paid by borrowers for the use of money that they borrow from a lender [1] and typically expressed as an annual percentage rate. Interest rates rise and fall, and these changes or anticipated changes in interest rates can have an effect on businesses' decisions with regard to their borrowing requirements. Interest rates changes are important because they affect individuals by, businesses, financial markets and government policy making [2].

When lenders allow borrowers use their money, it is in their best interests to ensure that the money are safe and will be returned back, and they use measures of creditworthiness, such as FICO score, for it. A person's FICO score will range between 300 and 850. In general, a FICO score above 650 indicates that the individual has a very good credit history. People with scores below 620 will often find it substantially more difficult to obtain financing at a favorable rate [3].

Understanding the relationship of interest rate with other variables can help us characterize loans and what variables contribute to the final interest rate number. Here we performed an analysis to determine if there was a significant association between *interest rate* and other variables such as *requested amount*, *granted amount*, *debt-to-income ratio*, *state*, *open credit lines*, *revolving credit balance*, *inquiries in the last six months* and *employment length*.

Using exploratory analysis and standard multiple regression techniques we show that there is a relationship between some of these variables. Our results suggests that there is a strong association between *interest rate* and *requested amount*, *loan length*, *loan purpose* and *debt-to-income ratio*. And finally, we analyze the actual FICO score in order to check our model.

2 Methods

2.1 Data Collection

For our analysis we used a data sample consisting of 2,500 peer-to-peer loans issued through the Leading Club¹. The data was already prepared for the "Data Analysis" course by the tutor and was downloaded using the provided link² on February 10, 2013 with the R programming language [4].

2.2 Exploratory Analysis

Exploratory analysis was performed by examining tables and plots of the observed data. We identified transformations to perform on the raw data on the basis of plots and knowledge of the scale of measured variables. Exploratory analysis was used to

1. identify and remove missing values,
2. verify the quality of the data, and
3. determine the terms used in the regression model relating *interest rate* to *amount requested*, *debt-to-income ratio* and *loan length*.

¹<https://www.lendingclub.com/home.action>

²<https://spark-public.s3.amazonaws.com/dataanalysis/loansData.csv>

2.3 Statistical Modeling

To relate *interest rate* to other variables we performed a standard multivariate linear regression model [6]. Coefficients were estimated with ordinary least squares and standard errors were calculated using standard asymptotic approximations [7].

2.4 Reproducibility

All analyses performed in this manuscript are reproduced in the R markdown file `loansFinal.Rmd` [5]. To reproduce the exact results presented in this manuscript the cached version of the analysis must be performed, as the data available from the “Data Analysis” coursera website might change.

3 Results

The loans data used in this analysis contains information on the source network that measured: *amount requested* – the requested amount (in dollars), *amount funded by investors* – the loaned amount (in dollars), *interest rate* (in percents), *loan length* – the length of time of the loan (in months), *loan purpose*, *debt-to-income ratio*³ (in percents), *state*, *home ownership*, *monthly income* (in dollars), *FICO range* [3], *open credit lines* – the number of open lines of credit at the time of application, *revolving credit balance* – the total amount of debt (in dollars), *inquiries in the last 6 months* and *employment length* (in years).

We identified that in two observations there were missing values in the data set we collected and they therefore were removed. Data columns with percent character (“%”) needed to be preprocessed, transformed to numeric variables and divided by 100 to represent percents to make it possible to work with it in R. All measured variables were observed to be inside the standard ranges. Potential confounders such as *amount funded by investors* (close to 1 correlation with amount requested) and *FICO score* (we know a-priori that it correlates with interest rate) were excluded from the model.

Interest rate when first seen did not seem to show major patterns against *requested amount*, but it was possible to fit a straight line with correlation coefficient 0.33 between them. When *requested amount* was cut into five interval and compared with interest rate, it was obvious that it did follow a pattern [figure 1a]: the more money requested the higher interest rate. Also, there was found an apparent correlation between *interest rate* and *loan length*: the longer the period the higher the rate. Also, relationships with *loan purpose* (renewable energy, education and car tend to have the lowest interest rate and dept consolidation is with the highest) and *debt-to-income ratio* (the increase in one leads to increase in other) were discovered.

We first fit a regression model relating *interest rate* to *amount requested*. The residuals showed patterns of non-random variation [figure 1b]. We attempted to explain those patterns by fitting models including potential confounders and non-relevant variables. Our final regression model was:

$$IR = b_0 + b_1 * AR + f_1(Len) + f_2(P) + b_3 * D/I + e$$

where b_0 is an intercept term and b_1 represents the change in *interest rate* IR associated with a change of one unit in *amount requested*. The term $f_1(Len)$ for *loan length* represents a factor model with two levels and $f_2(P)$ for *loan purpose* represent a factor model with 14 levels. The term b_3 represents a change in *interest rate* associated with a change of one unit in *debt-to-income ratio* D/I . The error term e represents all sources of unmeasured and unmodeled

³http://en.wikipedia.org/wiki/Debt-to-income_ratio

random variation in interest rate. Our final regression model appeared to remove a lot of the non-random patterns of variation in the residuals [figure 1c].

We observed statistically significant associations between *interest rate* and *amount requested* ($P < 2\text{e-}16$) and between *interest rate* and *debt-to-income rate* ($P = 3.3\text{e-}12$). A change of one unit in amount requested corresponded to a change of $b_1 = 8.76\text{e-}07$ (the value is very low since $IR \in [0..1]$ and values of amount requested are high (up to 35000) with 95% confidence interval ($6.681\text{e-}07 \dots 1.084\text{e-}06$). Likewise, a change of one unit in *debt-to-income* ratio resulted in a change of $b_2 = 6.96\text{e-}02$ with 95% confidence interval ($5.009\text{e-}02 \dots 8.908\text{e-}02$). The correlation coefficient between the model and interest rate is 0.51.

4 Conclusions

There appears to be a strong relationship between the variables in the regression model: *amount requested*, *debt-to-income ratio*, *loan purpose* and *loan length* are associated with interest rate.

However, if taking *FICO score* in consideration, a correlation coefficient between it and *interest rate* is 0.7, which suggests that there is room for further improvement in the performed analysis.

While our analysis is an interesting first step it is based on a limited sample of data. A larger collection of representative loans may be more appropriate for understanding the relationship between interest rate and other variables. Our analysis may be of interest to scientists seeking to better understand the principles of financial markets.

References

- [1] Wikipedia "Interest rate" Page. URL: http://en.wikipedia.org/wiki/Interest_rate. Accessed 2013/02/14.
- [2] "Importance of Interest Rates" web page. URL: http://www.ehow.com/info_8155955_importance-interest-rates.html. Accessed 2013/02/14.
- [3] "FICO Score" web page at Investopedia. URL: <http://www.investopedia.com/terms/f/ficoscore.asp> Accessed 2013/02/14.
- [4] R Core Team (2012). "R: A language and environment for statistical computing." URL: <http://www.R-project.org>. Accessed 2013/02/13.
- [5] R Markdown Page. URL: http://www.rstudio.com/ide/docs/authoring/using_markdown. Accessed 2013/02/14
- [6] Seber, George AF, and Alan J. Lee. Linear regression analysis. Vol. 936. Wiley, 2012.
- [7] Ferguson, Thomas S. A Course in Large Sample Theory: Texts in Statistical Science. Vol. 38. Chapman & Hall/CRC, 1996.