Higher FICO scores and shorter loan lengths are associated with lower interest rates in moneylending

Introduction:

Lending capital assets is one of the primary drivers of activity in modern economies. Lower interest rates in a loan decrease the overall costs to the borrower thereby stimulating greater economic activities [1]. Two variables that affect the interest rate for a particular loan are the borrower's FICO score and the length of the loan being requested. The FICO score is a proprietary score designed to evaluate the risk of credit default [2]. Loan length is usually measured in months and often comes in 36 or 60 month durations as indicated in our data set.

Understanding the variables that go into an interest rate determination can help borrowers position themselves to receive more favorable rates and decrease the total cost of these financial services. Here, we performed exploratory analysis and standard linear regression techniques to show that there is a significant relationship between a loan's interest rate, the FICO score, and the loan length. We found that the amount requested and the final amount funded were additional confounding variables in our model. Our analysis suggests that—in addition to a higher FICO score—a decreased loan length was associated with lower interest rates.

Methods:

Data Collection

For our analysis we used a sample of 2500 loans issued through the Lending Club [3], a peer-to-peer lending institution. The data were downloaded through the Data Analysis course website on 16 February 2013 using the R programming language [4].

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 measured variables. Exploratory analysis was used to (1) identify missing values, (2) verify the quality of the data, and (3) determine the terms used in the regression model relating FICO Score and loan length to interest rate.

Statistical Modeling

To relate FICO Score and loan length to interest rate, we performed a standard multivariate linear regression model [5]. Model selection was performed on the basis of our exploratory analysis and knowledge of standard lending practices. Coefficients were estimated with ordinary least squares and standard errors were calculated using standard asymptotic approximations [6].

Reproducibility

All analyses performed in this manuscript are reproduced in the R markdown file Project 1.Rmd [7]. To reproduce the exact results presented in this manuscript, we recommend using a cached version of the data, as the data available for download may be different in the future.

Results:

The lending data used in this analysis contains 14 variables characterizing some aspect of the loan. These included the amount requested by the borrower (R_i), the amount actually funded by the lenders (F_i), the interest of the loan (IR_i), the length of the loan in months (L_i), and the range into which borrowers' FICO score was grouped (S_i). The remaining nine variables contained information about such things as the debt-to-income ratio, monthly income, and other characteristics of the borrower; however these variables did not appear to have any significant relationship to the interest rate. During the course of our exploratory analysis, we identified missing values in two of the variables; however neither variable appeared to have any meaningful relationship to the interest rate and so we left those values undefined. All other values fell within normal expectations. The distribution of interest rates across all observations in the sample did not need to be transformed in a logarithmic fashion (Figure 1, Left). Several variables were stored as factors and had to be converted into numeric classes. In particular, we converted FICO scores into a numeric by taking the first three characters and coercing them into a numeric class. This allowed us greater flexibility when performing regression analysis. We found that there were two variables in particular that strongly suggested a relationship to the interest rate, and an additional two variables which weakly suggested a relationship. Further analysis focused on these four variables in order to explain the variation in interest rates.

We first fit a regression model relating FICO score to interest rate. However, the residuals showed clear additional patterns of non-random variation (Figure 1, Middle). We attempted to explain those patterns by fitting models including potential confounders. Our final regression model was:

$$IR_i = b_0 + b_1 S_i + b_2 L_i + b_3 R_i + b_4 F_i + e_i$$

- IR_i is the interest rate for borrower *i*
- b_0 is an intercept term
- $b_1 S_i$ represents the change in interest rate based on borrower *i*'s FICO Score
- $b_2 L_i$ represents the change in interest rate based on borrower *i*'s loan duration
- $b_3 R_i$ represents the change in interest rate based on borrower *i*'s requested amount
- $b_4 F_i$ represents the change in interest rate based on borrower *i*'s funded amount
- e_i represents all sources of unmeasured and unmodeled random variation in interest rate determinations.

Our final regression model appeared to remove most of the non-random patterns of variation in the residuals (Figure 1, Right).

We observed a very statistically significant ($P < 2 \times 10^{-16}$) association between interest rate and FICO Score. An increase of one point in a borrower's FICO score corresponded to a change of $b_1 = -8.75 \times 10^{-2}$ percentage points for the interest rate (95% Confidence Interval: -.090, -.085). For example, two borrowers seeking the same amount in loans, funded at the same level, for the same length of time, and whose FICO scores differed would expect the interest rate to be 8.75 points lower for every 100 points increase in the score.

We also found that duration of the loan had a very statistically significant ($P < 2 \times 10^{-16}$) association with the interest rate. An increase of 1 month in the duration of the loan corresponded to a change of $b_2 = .14$ percentage points for the interest rate (95% Confidence Interval: 0.128, 0.146). It is important to note, however, that loan durations are typically only given in 36 or 60 month durations for non-mortgage lending. In practice, a borrower would see the interest rate increase by 3.36 percentage points when moving from a 36-month to a 60-month loan.

Conclusions:

Our analysis suggests that loan interest rates have a very significant positive association with borrowers' FICO score and a very significant negative association with the loan duration. We also observed that the loan amount requested and amount funded have a lesser effect on the interest rate. Including these other variables in the regression model relating improves the fit, but does not remove the significant relationship between the variables.

While our analysis yields interesting results, it is based on a sample that is unlikely to be representative of the wider lending market. Peer-to-peer lending may differ in key ways from traditional financial institutions, or payday-loan type businesses. It is not clear if the associations uncovered in this analysis are a characteristic of these and other types of lending establishments. Thus, our analysis provides more value to a prospective consumer attempting to secure favorable loan terms than it does to the policy-maker attempting to alter the existing regulatory environment. Future research using a more broad-based sample including a variety of different lending institutions would be instrumental to inform policy-makers.

References

1. Wikipedia “Debt” Page. URL: http://en.wikipedia.org/wiki/Debt. Accessed 16 Feb 2013.

2. Wikipedia “Credit Score in the United States” Page. URL: http://en.wikipedia.org/wiki/Credit_score_in_the_United_States. Accessed 16 Feb 2013.

3. Lending Club Home Page. URL: http://www.lendingclub.com Accessed 16 Feb 2013

4. Johns Hopkins via Coursera. URL: https://class.coursera.org/dataanalysis-001/human_grading/view/courses/294/assessments/4/submissions Accessed 16 Feb 2013

5. Seber, George AF, and Alan J. Lee. *Linear regression analysis*. Vol. 936. Wiley, 2012.

6. Ferguson, Thomas S. *A Course in Large Sample Theory: Texts in Statistical Science*. Vol. 38. Chapman & Hall/CRC, 1996.

7. R Markdown Page. URL: http://www.rstudio.com/ide/docs/authoring/using_markdown. Accessed 16 Feb 2013