

Title: Lending Club Interest Rates are closely linked with FICO scores and Loan Length

Introduction:

The Lending Club is a unique website that allows people to directly borrow money from other people [1]. Borrowers have the opportunity to borrow at lower rates than the average banks, and lenders, also called investors, have the potential to earn better returns by investing in very creditworthy borrowers [2]. Borrowers can receive interest rates as low as 6.78%, well below the national average, or as high as 27.99%, which is higher than many typical bank loans. Since borrowers are always seeking the lowest possible interest rates, it is important to understand what factors most significantly affect the rates a Lending Club member can obtain.

FICO scores are often a significant factor in determining a lender's interest rates. The FICO score is a statistical calculation made up of a person's payment history, credit history, credit utilization, types of credit used, and recent searches for credit [3]. However, the FICO score is usually not the only determinant of one's interest rates. Regardless, there are a number of confounding factors which are difficult to adjust for because they are closely linked to the FICO score. Our analysis examines what largely independent factors outside of the FICO score influence a lender's interest rates. Our results suggest that the second largest factor is the lender's loan length.

Methods:

Data Collection

To examine what impacts Lending Club's interest rates, we analyzed a sample of 2,500 loans from Lending Club's publicly available data provided by Professor Jeff Leek on Coursera [4]. Lending Club also has a more complete data set with 110,751 observations from their website [5].

Exploratory Analysis

Exploratory analysis was done using plots and correlations. The dataset was small enough that it was easy to make comparisons between Interest Rates and each of the other categories. Exploratory analysis was used to (1) identify missing values, (2) identify outliers in the data and assess whether they should be omitted, (3) determine the factors that appear to influence interest rates, and (4) determine how some values may need to be manipulated.

Statistics

We first used Pearson correlation tests between interest rates and the other variables to determine the most influential factors as well as confounding factors [6]. Using the corrgram package was a simple way to view all the correlations [7]. After determining the variables most likely to influence interest rates, we ran two linear models based on the two Loan Lengths.

Results:

The data in this analysis is a small subset of the data available from Lending Club's website. There were 2 rows of partially complete data, but it did not affect our analysis so we used all 2,500 observations. Correlations showed clear links between Interest Rate (Ra), FICO Range (FICO), and the Loan Length. The interest rate was converted to numeric values for statistical analysis. Because the FICO data was recorded as a range, we converted it to an integer which was rounded down to the lower end of the original FICO range. These transformations allowed us to make a more accurate model.

The interest rates on loans fell between 5.42-24.89% with the average being 13.07%. The interest rates and FICO range variables were both converted to numbers for statistical analysis. There was also a strong correlation ($R = -0.71$) between interest rates and the FICO, which included a few small outliers on the FICO. As well there was a strong correlation between the loan length and the interest rates ($R = 0.42$), which is important since there is not a strong link between FICO and loan length. Thus, the FICO and loan length are largely independent variables.

We created two regression models for the two loan length options—36-month loans (S), and 60-month loans (L). There was clearly non-random variation with both models. However, there was one outlier with a low interest rate, low FICO, and a 60-month loan duration which may have affected the model for 60-month loans. The following two regressions models were used in the final analysis:

$$\begin{aligned} RaS_i &= bs_0 + bs_1 FICO * S_i + es_i \\ RaL_i &= bl_0 + bl_1 FICO * L_i + el_i \end{aligned}$$

In both models the first term is an intercept and the second term represents the change in interest rates associated with a difference in the FICO depending on the loan length. The last term represents error from any unmeasured variation.

Figure 1 shows both of the models overlayed on top of a scatter plot of the interest rates with respect to the lender's FICO. There is a very strong negative relationship between interest rate and FICO ($P < 2e-16$) for both models, i.e., as FICO increases, interest rate decreases. For 36 month loans, a one point change in a lender's FICO score corresponds with a 0.08 percent change in interest rate (95% Confidence Interval: -.08, -.07). For 60 month loans, a one point change in FICO score corresponds to a 0.1 percent change in interest rate (95% Confidence Interval: -.10, -.09). These two models make it possible for us to estimate a lender's interest rate based on their FICO and loan length.

Conclusions:

Our analysis indicates that a lender can largely determine their interest rate based upon their FICO and the loan length. We used two linear models estimating the relationship between interest rate and FICO—one for each of the two loan lengths. We can see that the FICO has a much greater impact on the lender's interest rates for 60 month loans. In our exploratory analysis, we also found strong links between interest rates and the amount requested, the amount funded, and the debt to

income ratio, however, these are all confounding variables with the FICO and each other so modeling a regression using all of the terms would have been difficult.

From the models, we can see that the higher a lender's FICO score is, the lower their interest rate is. When loan length is shorter, the FICO score has less of an impact than when the loan length is longer.

The analysis with this small sample size give interesting insight, but ideally we could explore if there were any links between the interest rates and the factors that were not included in our data. In addition, further analysis should be done to see if our model using this sample still holds true for interest rates in the full set of data available from Lending Club. This model could be of great benefit to lenders, but further models relating this information to actual returns would probably be more beneficial for both Lending Club and borrowers.

References

1. LendingClub "Personal Loans & Investing with Peer Lending" Page. URL: <https://www.lendingclub.com/home.action>. Accessed 2/13/2013.
2. LendingClub "How Peer Lending Works" Page. URL: <https://www.lendingclub.com/public/how-peer-lending-works.action>. Accessed 2/13/2013.
3. Wikipedia "Credit score in the United States" Page. URL: http://en.wikipedia.org/wiki/Credit_score_in_the_United_States#FICO_score. Accessed 2/16/2013.
4. Coursera "Assessment Details | Data Analysis" Page. URL: https://class.coursera.org/dataanalysis-001/human_grading/view/courses/294/assessments/4/submissions. Accessed 2/13/2013.
5. LendingClub "Lending Club Statistics" Page. URL: <https://www.lendingclub.com/info/statistics.action>. Accessed 2/13/2013.
6. Wikipedia "Pearson product-moment correlation coefficient" Page. URL: http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient. Accessed 2/17/2013.
7. R-bloggers "CORRGRAM: Correlation Matrix (Constituents)" Page. URL: <http://www.r-bloggers.com/corrgram-correlation-matrix-constituents/>. Accessed 2/16/2013.