

Title: Relationship between loan length and interest rate of peer-to-peer loans

Introduction:

Peer-to-peer lending is the practice of lending money to previously unrelated individuals or “peers” without the intermediation of traditional financial institutions [1]. The interest rates are set either by lenders who compete for the lowest rate on the reverse auction model, or are fixed by the intermediary company on the basis of their analysis of the borrower's credit [2]. Borrowers assessed as having a higher risk of default are assigned higher rates.

Lending Club is a US peer-to-peer lending company, headquartered in San Francisco, California. The interest rate of these loans is determined by the Lending Club on the basis of characteristics of the person asking for the loan such as their employment history, credit history, and creditworthiness scores. [3]

Exploratory analysis and standard multiple regression techniques were used in order to show that there is a significant relationship between loan length and the interest rate, in addition to the relationship between FICO Score and the interest rate. Our analysis suggests that increased loan length is associated with increased interest rate.

Methods:

Data Collection

For our analysis we used the data on 2,500 peer-to-peer loans issued through the Lending Club. The data were downloaded from Coursera Data Analysis course web page on February 6, 2013 using the R programming language [4].

Exploratory Analysis

Tables and plots of the observed data were examined in order to perform exploratory analysis. Missing values were identified and removed and the quality of data was verified. In addition, the terms used in the regression model relating FICO Score and loan length to the interest rate were determined.

Statistical Modeling

A standard multivariate linear regression model was performed in order to relate FICO Score and loan length to the interest rate [5]. Model selection was performed on the basis of our exploratory analysis and prior knowledge of the relationship between loan length and interest rate. Coefficients were estimated with ordinary least squares and standard errors were calculated using standard asymptotic approximations [6].

Reproducibility

Due to potential security issues, reproducible R code was not submitted. However, R code was written and saved for later use to meet the reproducibility requirements.

Results:

The peer-to-peer loans data used in this analysis contain the following variables that were used in the regression model: interest rate (IR), FICO score range (FICO) and loan length (LEN). In addition, the data contained the amount requested, the amount funded by investors, loan purpose, debt to income ratio, state, home ownership, monthly income, open credit lines, revolving credit balance, inquiries in the last 6 months and employment length.

Two cases with missing values were identified and removed from the data set. In order to fit the regression model to FICO score, FICO score ranges were transformed to integer values by taking the average value of each score range. Subsequent analyses focus on this transformed FICO score variable.

Interest rate density seemed to vary based on the loan length [Figure 1 Left Panel]. We fit a regression model relating FICO Score and loan length to the interest rate. The residuals showed patterns of non-random variation with the high values of interest rate. Our regression model was:

$$IR_i = b_0 + b_1 FICO_i + b_2 LEN_i + b_3 FICO_i \times LEN_i + e_i$$

where b_0 is an intercept, b_1 is change in interest rate for a single FICO score value, b_2 is coefficient term of loan length of 60 months and b_3 is the interaction term between FICO score and the length of the loan. The error term e represents all sources of unmeasured and unmodeled random variation in interest rate. In our regression model there seemed to be non-random pattern of variation in the residual with the values of high interest rate [Figure 1 Right Panel].

We observed a highly statistically significant ($P = 2.78e-12$) association between length of the loan and the interest rate [Figure 1 Middle Panel]. Loan length of 60 months compared to 36 months would increase the intercept term (interest rate) by 16 units of percentage (95% Confidence Interval: 11.7, 20.8).

Conclusions:

Our analysis suggests that increased loan length is associated with increased interest rate. Our analysis estimates the relationship using a linear model relating FICO Score and loan length to the interest rate.

Since non-random patterns of variation were shown in the final residuals, there is still work to be done in order to fully understand the relationship between different variables with the interest rate. A larger collection of representative peer-to-peer loans with additional variables would be desired. Nevertheless, our analysis was an interesting first step trying to understand the interest rate determination of peer-to-peer loans.

References

1. Wikipedia "Peer-to-peer lending" Page. URL: http://en.wikipedia.org/wiki/Peer-to-peer_lending. Accessed 2/6/2013.
2. Lepro, Sara (2010). "Prosper Ditches Auction Pricing for Model Like P-to-P Rival's". American Banker. URL: http://www.americanbanker.com/issues/175_243/prosper-lending-club-1030207-1.html . Accessed 2/6/2013.
3. Wikipedia "Lending Club" Page: URL: http://en.wikipedia.org/wiki/Lending_Club . Accessed 2/6/2013.
4. R Core Team (2012). "R: A language and environment for statistical computing." URL: <http://www.R-project.org>
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.