Title: Interest rates on loans affected both by amount requested and loan length besides applicant FICO Score

Introduction:

The best-known and most widely used credit score model in the United States, the FICO Score[1], is calculated[2] on applicant's payment history, credit utilization, length of credit history and other components. It provides a snapshot of risk that banks and other institutions use to help make lending decisions, so applicants with higher FICO Scores might be offered better interest rates.

Not only FICO Score is used for setting the interest rate of a loan. Understanding which other components are taken into account will reveal the internals of the risk assurance at the financial panorama. Our results suggest a strong correlation between interest rate, amount requested and loan length besides the applicant's FICO Score

Methods:

Data Collection

For our analysis we used a a sample of 2,500 peer-to-peer loans issued through the Lending Club (https://www.lendingclub.com/home.action). The data were downloaded on Saturday 16 Feb 2013 using the R programming language[3].

Exploratory Analysis

After some cleaning, simple linear regression was performed for several variables against interest rate. By comparing the resulting coefficients of determination some variables appeared to fit significantly better: loan length and amount requested.

Statistical Modeling

Multivariate linear regression model[4] was applied using loan length and amount requested variables, as they were revealed in our exploratory analysis. Coefficients were estimated with ordinary least squares and standard errors were calculated using standard asymptotic approximations [5].

Reproducibility

Performed with R programming language, the analysis were structured using ProjectTemplate package[6] and it has been published on a Github repository[7]. The *src/reproducible_code.html* file reproduces the whole exploratory analysis and results.

Results:

A exploratory analysis determines great differences on interest rate for 36 and 60 months loan length(1). Even though the amount requested slightly affects interest rate (0.000178% per extra dollar requested), there appears to be a trend.

A separated linear regression(1) on interest rate(%) vs. FICO range(score) for both 36 and 60 months loan's length shows the clear trend of higher interest rate on 60 month. A 36 months loan request receives an approximated increment of 5.66% on interest rate for a 60 months loan for the same applicant. Adding a the amount requested(2) to the model, as a relation with loan length, improves the coefficient of determination, and that drives us to think it's a better way to model it. So, taking into account potential confounders, our final regression model was:

$$IR = b_0 + b_1 (NFR) + b_2 (AR/LL) + e$$

where b_0 is an intercept term, b_1 represents a change of -0.4341% interest associated with each higher value for FICO range and b2 it's b_2 he coefficient for the relation between the amount requested (AR) and loan length (LL). The error term e represents all sources of unmeasured and unmodeled random variation.

We observed a statistically significant (P = 2e-16) association for this model, with a coefficient of determination (r-squared) of 0.7331.

Conclusions:

Our analysis suggests that there is a significant association between interest rates and the amount of money the applicant would pay each month, for applicants with same FICO score. Our model is best fitted between 700-820 FICO Ranges, regarding the original sample lacks of enough data outside this limits.

As conclusion, as FICO score itself reflects some of the rest of variables, it's reasonable to think the lender doesn't takes gives such importance to other components again on calculation of interest rate.

References

1. myFico site: https://www.myfico.com.
2. Credit score in the United States: http://en.wikipedia.org/wiki/Credit_score_in_the_United_States. Accessed 16/02/2013.
3. R Core Team (2012). ”R: A language and environment for statistical computing.” URL: http://www.R-project.org
4. Seber, George AF, and Alan J. Lee. *Linear regression analysis*. Vol. 936. Wiley, 2012.

5. Ferguson, Thomas S. *A Course in Large Sample Theory: Texts in Statistical Science*. Vol. 38. Chapman & Hall/CRC, 1996.
6. ProjectTemplate Page. URL: http://projecttemplate.net.
7. Analysis project page at Github. URL: https://github.com/fontanon/interestrates-vs-ficoscore-analysis