Title: Interest rate in peer-to-peer loans and its association with borrower's characteristics

Introduction

Peer-to-peer loans are becoming an increasingly common method of borrowing and lending money. With the advent of internet, the offer from lenders and the demand from borrowers can now meet online, without a bank acting as intermediate[1]. Potential borrowers simply sign up to one of the companies offering this service and post a request for a loan. Other users who are willing to finance this loan can accept it and lend the money applying a certain interest rate. In general financial terms, this rate measures the loan's risk: lenders ask for a higher interest rate if they think the borrower has higher chances of not returning the money. In the framework of this project, we were given the characteristics of 2500 loans issued through the company Lending Club[2]. The interest rate of these loans is determined by the company according to some characteristics of the borrower such as employment history, monthly income, creditworthiness and so on. We analysed these records to find the best predictors of the interest rate. One term that influences the interest rate, not surprisingly, is the creditworthiness score FICO[3]. Other variables appear to have a significant effect on the rate applied. We argue that, among these variables, the ratio between the amount requested and the monthly income of the borrower plays an important role in determining the interest rate applied.

Methods

Data

We analysed 2500 loans issued through the Lending Club. The data were downloaded from the assignment webpage in the form of an R data file, loaded and analysed with the R programming language[4]. For each transaction, 14 variables are reported. Some variables characterise the loan, for example, the amount requested and funded, the length of the loan and its purpose. Other variables are informative of the borrower, like the monthly income, the FICO score for creditworthiness, the state, the home ownership and so on.

Data cleaning, exploratory analysis

For one observation, monthly income and other variables are not reported: we removed this observation from our data.
In the original data, FICO score of the borrower is reported as a range of five points, rather than as actual score. In order to simplify the analysis, we replaced the range with its lowest end, so that range 640-644 corresponds to 640, for example, 645-649 to 645 and so on.
Based on the first exploratory analysis, and knowledge of the common criteria that might influence the interest rate applied, we defined a new variable that is the ratio between the amount requested for loan and the monthly income of the borrower. Exploratory analysis was conducted by producing plots, performing analysis of variance and linear fits using the R programming language. The figure was produced with the "effects" package[5].

Results

As expected, the interest rate applied goes down with the FICO score. Interest rate for borrowers with score between 700 and 704 is on average 13.4% (131 observations), while for borrowers with score between 750 and 754 it is on average 8.5% (61 observations). Some features do not seem to have an effect on the interest rate applied: for example, the employment length and the state of the borrower. This was observed by plotting histograms of the interest rate and confirmed by analysis of variance (anova). There is a small, although significant, effect of home ownership on the interest rate: borrowers who are on rent pay on average 0.7% interest more than those who own their residence or have a mortgage. The statistical significance of this effect was assessed with analysis of variance, and quantified to be at approximately 0.01%.

We explored the association of the interest rate with the monthly income. Since this is distributed in very skewed way, we consider its logarithm, that
is more symmetric. The linear fit of the interest rate vs. the logarithm of the income shows, quite surprisingly, a positive effect. This is not strongly significant, and the R-squared of the fit is very low (below 0.2%). The amount requested shows a stronger association with the interest rate, with an R-squared of 11% and a positive, strongly significant coefficient for the association.

The fact that the income is correlated to both the amount requested and the interest rate make suggest that it is better treated as a confounder.

We finally performed a linear fit including as covariates the FICO score, the amount requested and their interaction, as follows:

(1) IR = a + b_A * Amount + b_F * FICO_score + b_x * Amount * FICO_score,

where IR is interest rate in percent, Amount is the amount requested, measured in dollars, and FICO_score is the lower end of the FICO score range (see Methods). The estimated parameters, the intercept a, b_A, b_F and b_x, are all significantly different from zero (the highest p-value is 5E-9). Their estimates are

a = 64.2 +/- 1.8
b_A = 9.3 +/- 1.2 (E-4)
b_F = -7.6 +/- 0.2 (E-2)
b_x = -1.0 +/- 0.17 (E-6)

We immediately see that the interest rate goes up with the amount requested and down with the FICO score. The negative interaction term means that, for borrowers with higher score, the interest rate increases more slowly with the amount requested. This is clear if one writes the model (1) as follows:

IR = a + (b_A + b_x * FICO_score) * Amount + b_F * FICO_score.

Figure 1 reports the fitted interest rate as a function of the requested amount, plotted for three different values of the FICO score (the lowest, the median and the highest score observed). The effect of the negative interaction
term is in the fact that the solid black line (lowest score) is steeper than the dashed

red one (median score), which is steeper then the dotted green one (highest score).

Conclusions

We have identified several variables influencing the interest rate of loans granted through the peer-to-peer lending company Lending Club. While the largest effect is given by the FICO score, as expected, other factors like the home ownership are associated to a different interest rate: borrowers renting their residence paying on average 0.7% more. Factors as employment length or state do not show to play a role in the quantification of the interest rate.

We have observed that another predictor strongly associated with the interest rate is the amount requested. By looking at the interaction with the FICO score, we also observed that the interest rate grows more slowly as a function of the amount requested for individuals with higher creditworthiness.

A possible extension of this analysis might include a more extensive model selection. Rather than performing based on common sense, one could perform a cross validation.

[1]: http://en.wikipedia.org/wiki/Peer_to_peer_lending
[2]: http://www.lendingclub.com/home.action
[3]: http://en.wikipedia.org/wiki/Credit_score_in_the_United_States
[4]: http://www.R-project.org
[5]: http://cran.r-project.org/web/packages/effects/index.html