

Title: Monthly Income is a loan qualification factor that differs per State when determining Interest Rates offered to borrowers with similar FICO scores

Introduction:

Lenders use a variety of factors to qualify a borrower for a specific loan at a given Interest Rate. The Interest Rate is a key factor in the cost of borrowing money [1]. Typically, large financial institutions use a common set of factors to determine the Interest Rate they will offer borrowers for certain types of loans. These factors include, current economic rate levels, conditions of the loan, borrower-lender relationship and an assessment of the borrowers credit risk [2]. Lending Club, an online financial community that has been in business since 2007, provides investors and borrowers with an alternative to the traditional banking model so that both stakeholders can benefit financially [3]. Since Lending Club members are lending their own money, they can define their own criteria for qualifying borrowers and setting Interest Rates.

A key loan qualification factor for all lenders is the FICO score, which is used in more than 90% of lending decisions [4]. Lending Club members determine loan Interest Rates on the basis of individual borrower characteristics such as employment history, credit history, and creditworthiness scores [3]. The objective of our analysis was to identify and quantify associations between the Interest Rate of a loan and the other individual borrower characteristics (qualifications variables). In particular, we considered the impact a borrower's Monthly Income has on the Interest Rate offered by lenders in different U.S. States. Using exploratory analysis and standard bivariate correlation techniques we show that there is a significant relationship between Interest Rate and Monthly Income for borrowers with similar FICO scores.

Methods:

Data Collection

For our analysis we used a sample data set consisting of 2,500 peer-to-peer loans issued through the Lending Club. The data set was downloaded from Amazon EC2/S3 on 6 February 2013 using the R programming language [5].

Exploratory Analysis

Iterative examination of tables and plots coupled with the capturing of observations formed the basis of our exploratory analysis. Preliminary ETL [6] data processing techniques were used to (a) remove 2 loan observations (rows 367 and 1595) that had missing data, (b) transform certain qualitative factors into quantitative variables, and (c) create the structures and terms necessary to perform our correlation analysis.

Statistical Modeling

Using the Pearson correlation formula [7], our goal was to validate an assertion based on statistical inference that a relationship existed between a borrower's Monthly Income (y) and a lender's Interest Rate (x) offering.

$$r_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}.$$

The correlation coefficient (r) allows us to construct a confidence interval around (r) that has a given probability (p -value). We applied this correlation formula to two populations, (1) the entire corpus of loans (*the baseline*), and (2) segments of the population that had a minimum of 100 loan observations for the same mean FICO score (*the test data set*).

Reproducibility

This report represents a synthesized version of our research activity. Our analysis was performed using an R markdown file. The artifacts of this analysis have been organized into a reproducible research package, which is available upon request. Alternatively, a complete transcript of our data analysis workflow process is publically available [8].

Results:

Our first steps were to (a) check for Tidy Data, (b) clean up missing values, and (c) perform some data munging in order to get some of our qualitative data into quantitative format. Specifically, we created the following quantitative variables: Interest Rate ($ETL \cdot Rate$), Debit-to-Income Ratio ($ETL \cdot RatioDTI$), and Mean FICO Score ($ETL \cdot MeanFICO$). We also validated that our average FICO score for the baseline was equivalent to the published average that Lending Club has observed since their inception, namely a mean FICO score of 708 [9]. This was important as it added credibility to our sample set.

We proceeded to explore the existence of a relationship between Monthly Income and Interest Rates, by establishing two observation segments. Our baseline data set consisted of 2498 arbitrary loans. Our test data set consisted of 12 mean FICO scores where there was a minimum of 100 observations (loans) per score (FICO Group). Our loan distribution range for this test data set was between 112 and 171.

When applying the Pearson bivariate correlation analysis, we interpreted the results using the following four criteria [10]:

- The coefficient (r) strength with respect to the existence of a relationship
 - Absent ($r \approx 0$), Weak ($r \leq 0.5$), Strong ($r > 0.5$)
- The direction the correlation variables move based on the sign of the coefficient (r)
 - Positive ("same"), Negative ("opposite")

- The statistical significance in the confidence of a relationship (*p-value*)
 - Significant ($p < 0.05$), Strongly Significant ($p < 0.01$), Very Significant ($p < 0.001$)
- The ability of the relationship to predict variations between the correlation variables
 - 50/50 ($r^2 = 0.5$), Weak ($r^2 < 0.5$), Strong ($r^2 > 0.5$)

Using the baseline dataset (ignoring FICO scores), we compared the potential for an Interest Rate correlation with four loan factors (see *Figure 1 Panel A*). We observed that Monthly Income seems to present the least evidence ($r = 0.01$) of a relationship to Interest Rate as compared to correlations for Debit-to-Income ($r = 0.17$), Credit Balance ($r = 0.06$) and Quantity of Credit Lines ($r = 0.09$).

Using our test data set, we then compared loans that have the same FICO score (minimum of 100 observations per mean FICO score). We observed the presence of a relationship between Monthly Income and Interest Rate based on an estimated mean correlation coefficient ($r = 0.22$) for all FICO Group observations (see *Figure 1 Panel B*).

Using our inference criteria for Pearson correlations, we independently analyzed each of the 12 FICO Groups in our test data set and observed the following evidence for a correlation between Monthly Income and Interest Rate for borrowers with similar FICO scores:

- 100% (12 out of 12) of our observations yielded the presence of a weak relationship since they all had correlation coefficients below 0.5.
- 92% (11 out of 12) of our observations yielded a positive correlation implying that both variables move in the same direction. For example, if one increases so does the other.
- 67% (8 out of 12) yielded a *p-value* below 0.05, which implies the statistical significance of the presence of a relationship (see *Figure 1 Panel C*).
- 100% (12 out of 12) of the observations yielded an effective size (r^2) below 0.5, which indicates that this relationship can only predict <50% of the variations between Monthly Income and Interest Rate.

While our analysis resulted in evidence supporting the existence of an Income to Rate relationship within FICO Groups, we could not assume that the relationship was consistent across State boundaries. We extended our test data set to include States with the lowest and highest Interest Rates for each FICO Group as well as the incomes associated with those rates. Our results were that 9 out of the 48 represented States accounted for the highest Interest Rates for each of the 12 FICO Groups. AK, LA and OR each yielded the highest mean rates for 6 FICO Groups (see *Figure 1 Panel D*). We observed that State residence, another possible loan qualification factor, also has an effect on Interest Rate. This is consistent with our analysis, whereby a weak effective size (r^2) implies that relationship between Monthly Income and Interest Rate is not the only determining factor for Interest Rates and that a third confounding factor may be State of residence.

Conclusions:

Our reproducible research observed that Lending Club members use Monthly Income as a possible secondary factor for qualifying borrowers after they consider FICO scores. This observation is inline with published resources from Loan Club whereby they claim that the average borrower requires \$69,924 personal annual income [9]. In other words, the higher a borrowers income the more likely they are to qualify. However, this does not mean that income has a direct correlation to Interest Rate.

Our correlation analysis suggests that: (a) there is a significant probability of a relationship between Monthly Income and Interest Rates, (b) the relationship tends to exist when both variables (income and rate) move in the same direction, and (c) the impact of this relationship by itself on the offering of a specific Interest Rate is not very strong. In other words, there are other confounding variables, such as State of residence, that also have an impactful relationship on how a lender determines the Interest Rate to offer to a qualified borrower.

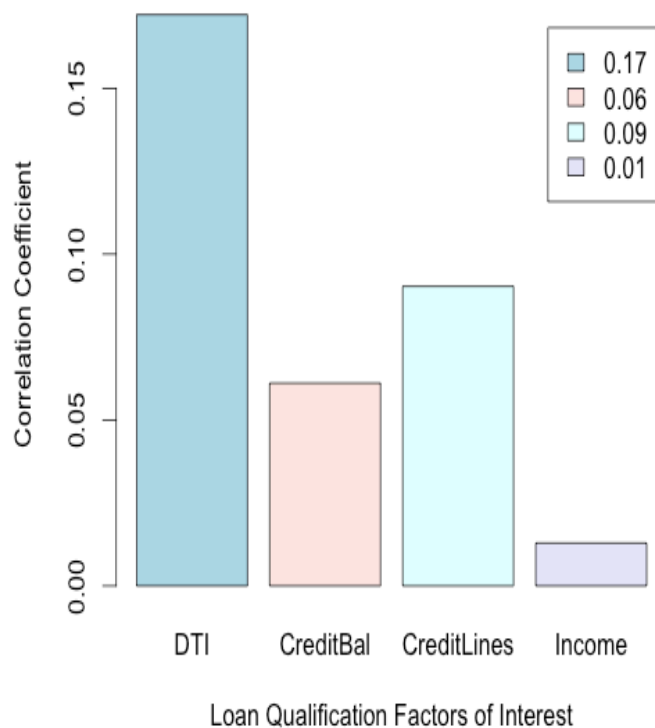
While statistical significance exists for a relationship between Monthly Income and Interest Rates, there is no evidence as to the exact qualification rules being applied by Lending Club members. For example, rates assigned to borrowers with similar FICO scores tend to differ based on State specific Income thresholds. All we can infer is that borrowers seeking loans from a lender in one State will be judged by an orthogonal set of rules by a lender in another State.

As of January 7, 2013, over 47,648 Lending Club investment accounts have funded over \$1.3 trillion in loans and received over \$103,315,600 in interest payments since June 2007 [9]. Access to the entire corpus of loans funded to date would have been more ideal for this analysis. A larger corpus of loans coupled with a broader (>100) distribution of loans within FICO Groups may have yielded a higher confidence value for the existence of the relationship. Access to a larger historical data corpus may also provide more insight into combined correlation between Monthly Income and State of residence on Interest Rates.

References

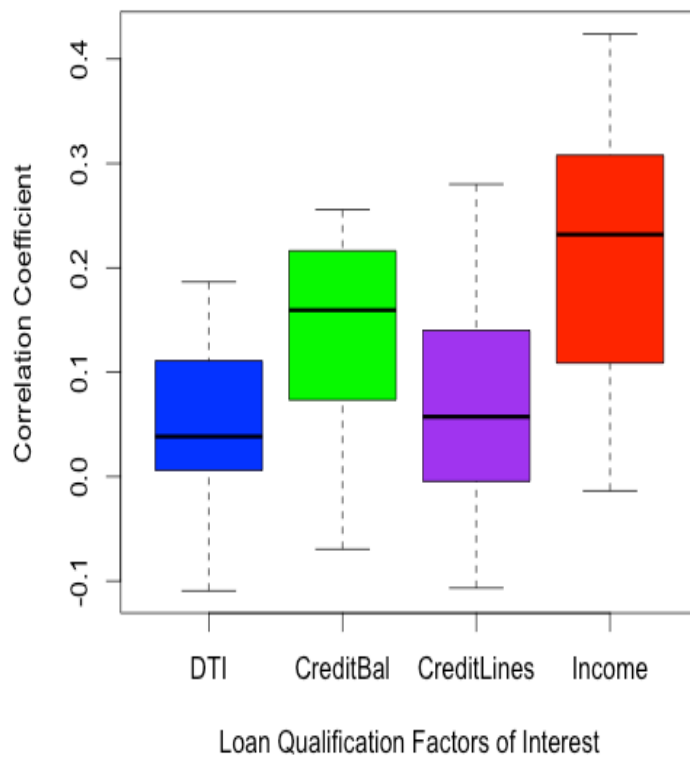
1. "Interest Rate Law & Legal Definition". URL: <http://definitions.uslegal.com/i/interest-rates/>. Accessed 2/13/2013.
2. "How Interest Rates are Determined". URL: https://www.wellsfargo.com/credit_center/use_credit/how_to_use/interest_rates. Accessed 2/13/2013.
3. "About Lending Club". URL: <https://www.lendingclub.com/public/about-us.action>. Accessed
4. "What is your FICO Score". URL: <http://www.myfico.com/crediteducation/whatsinyourscore.aspx>. Accessed 2/13/2013.
5. R Core Team (2012). "R: A language and environment for statistical computing." URL: <http://www.R-project.org>
6. Wikipedia "ETL" Page. URL: http://en.wikipedia.org/wiki/Extract,_transform,_load. Accessed 2/13/2013.
7. Wikipedia "Pearson Correlation Coefficient" Page. URL: http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient. Accessed 2/6/2013.
8. "Data Analysis Workflow Navigation: DAP1 Markdown Results". URL: [https://github.com/vinomaster/dawn/blob/master/guides/rant/sample/dap1/dap1_guide.p](https://github.com/vinomaster/dawn/blob/master/guides/rant/sample/dap1/dap1_guide.pdf)
[df](https://github.com/vinomaster/dawn/blob/master/guides/rant/sample/dap1/dap1_guide.pdf). Published 2/13/2013.
9. "Lending Club: Earn Solid Returns". URL: <http://www.lendingclub.com/public/steady-returns.action>. Accessed 2/6/13.
10. "Understanding Statistics Tutorial". URL: <http://www.ecs.org/html/educationissues/research/primer/understandingtutorial.asp>. Accessed 2/6/2013.

Baseline of Interest Rate Correlations for Corpus



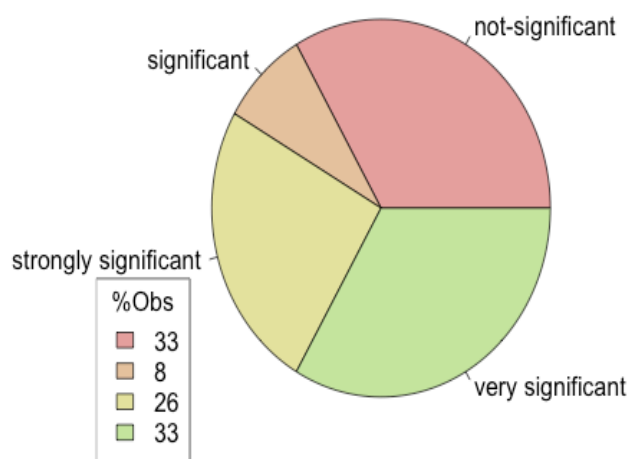
A

Rate Correlation Comparison for FICO Groups



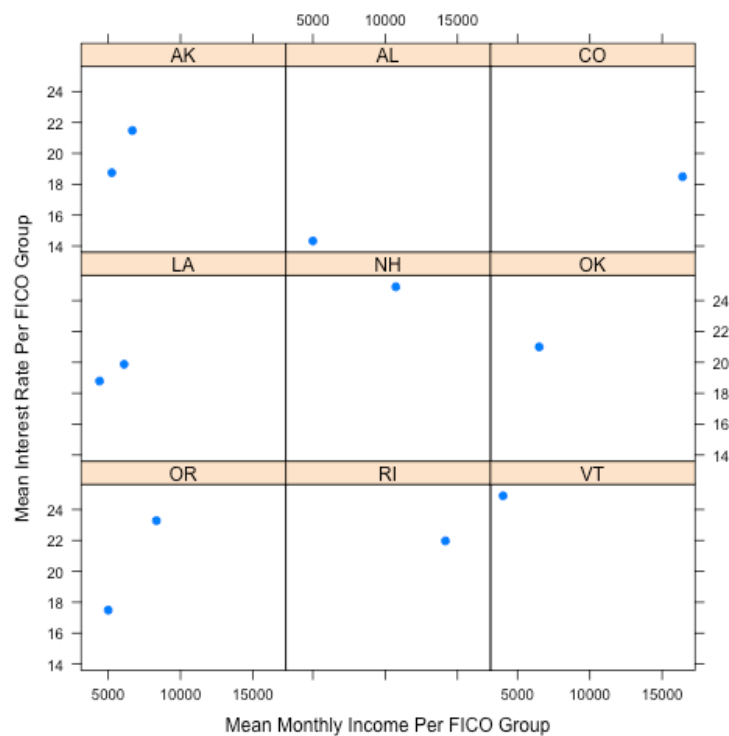
B

Income ~ Rate Relationship Confidence (P-Value)



C

Maximum Mean Rate ~ Income by State



D

Figure 1: (Panel A): A barplot of the correlation coefficients of four distinct loan qualification variables as compared to Interest Rates for all 2498 loan observations in the baseline corpus. Statistical significance of a relationship between Interest Rates and Monthly Income does not exist. **(Panel B):** A boxplot of the correlation coefficients of four distinct loan qualification variables as compared to Interest Rates for loans assigned to borrowers with the same mean FICO score. Based on an estimated mean correlation ($r=0.22$), we can infer that there is statistical significance of a relationship between Interest Rates and Monthly Income after FICO scores are considered. **(Panel C):** An analysis based on the Pearson correlation formula resulted in 67% of the observations having at least a significant confidence level ($p<0.05$). This implies a high degree of probability that a relationship between Interest Rates and Monthly Income exists after FICO scores are considered. **(Panel D):** In consideration of the existence of other confounding variables that may also have an impact on Interest Rates, the States with the highest Interest Rates per FICO Group were isolated. Half of the observations were associated with only 3 of the 48 represented States, implying that the relationship between Monthly Income and Interest Rate is not the only determining factor for Interest Rates and that a third confounding factor may be State of residence.