Linear Models Project Data Description

Joshua Ingram

3/25/2020

About LendingClub

The data I will be working with for this project is the LendingClub Loan Dataset. It is sourced directly from the LendingClub, a peer-to-peer lending platform that enables borrowers to have quick access to funds of up to $40,000. By partially funding loans, lenders are able to create diversified portfolios made up of any number of loans that are in accordance with their desired level of risk. Lenders are able to assess the risk of each loan by accessing information filled out by the potential borrower in the application, as well as additional information provided by LendingClub. Most of this information is found in the dataset, barring some information due to privacy concerns.

Motivations

The motivation behind this project comes from a desire to model the supply and demand for loans, enabling us to find the equilbrium price point (interest rate). Depending on what happens after data exploration, we may be able to find some level of credit rationing (the limiting of of the supply of additional loans by lenders to borrowers, even at a higher interest rate). Above all, I want to model the price of loans (interest rate) builty from the theory for the market for loans. I will also have to avoid the homogeneity bias that will come up due to both the quantity and price of loans being determined simultaneously. This dataset and project subject is of interest to me as I would like to turn this into my economics thesis.

Descriptions

There are over 2.25 million observations (loans) and 145 variables in this dataset, as well as an accompanying data dictionary. This dataset contains information on every loan issued from 2007 and 2019. I'll provide a link to the data dictionary at the bottom of the page, as 145 variable is quite a lot to list here. Here are some of the more important variables:

Response Variable

intRate - Interest rate on the loan

Explanatory Variables

Loan_Amount - the listed amount of the loan applied for by the borrower

Mortagage_Accounts - the number of mortgage accounts

Home_Ownership - the home ownership status provided by the borrower during registration (Rent, OWn, Mortgage, Other)

Debt_to_Incom - (debt-to-income ratio) a ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortage and the requested LendingClub loan, divided by the borrower's self-reported income

Annual_income - the self-reported annual income provided by the borrower during registration

Application_Type - indicates whether the loan is an indvidual application or a join application with two co-borrowers (Individual, Joint App)

Links

Data Dictionary Link: https://www.kaggle.com/wendykan/lending-club-loan-data#LCDataDictionary.xlsx

Note: I will not be using the data sourced from Kaggle (linked above), but rather the raw data from the LendingClub website... This will require a bit of data wrangling, but it is more up to date and I can get some more practice working with big data.