Counter-Party Credit Risk Modelling using PD,EAD and LGD model

June 27, 2021

1 Credit Risk

It is the risk of borrower not repaying loan, credit card or any other type of loan. Sometimes customers pay some installments of loan but don't repay the full amount which includes principal amount plus interest. It's a loss to the bank.

2 Basel Regulations

Basel III has incorporated several risk measures to counter issues which were identified and highlighted in 2008 financial crisis. It emphasis on revised capital standards (such as leverage ratios), stress testing and tangible equity capital which is the component with the greatest loss-absorbing capacity.

2.1 Internal Ratings Based (IRB) Approach

It has three credit risk components :

- Probability of Default (PD)
- Exposure at Default (EAD)
- Loss given Default (LGD)

2.1.1 Probability of Default (PD)

Probability of default means the likelihood that a borrower will default on debt (credit card, mortgage or non-mortgage loan) over a one-year period. In simple words, it returns the expected probability of customers fail to repay the loan. Probability is expressed in the form of percentage, lies between 0% and 100%. Higher the probability, higher the chance of default.

2.1.2 Exposure at Default (EAD)

It means how much should we expect the amount outstanding to be in the case of default. It is the amount that the borrower has to pay the bank at the time of default.

2.1.3 Loss given Default (LGD)

It means how much of the amount outstanding we expect to lose. It is a proportion of the total exposure when borrower defaults. It is calculated by (1 - Recovery Rate).

- LGD = (EAD - PV(recovery) - PV(cost)) / EAD
- PV (recovery) = Present value of recovery discounted till time of default.
- PV (cost) = Present value of cost discounted till time of default.

2.1.4 Expected Loss

Expected Loss is calculated by (PD * LGD * EAD).

3 What is Credit Risk Modelling?

Credit risk modeling refers to data driven risk models which calculates the chances of a borrower defaults on loan (or credit card). If a borrower fails to repay loan, how much amount he/she owes at the time of default and how much lender would lose from the outstanding amount. In other words, we need to build probability of default, loss given default and exposure at default models as per advanced IRB approach under Basel norms.

4 Probability of Default Modeling

4.1 Define Dependent Variable

Binary variable having values 1 and 0. 1 refers to bad customers and 0 refers to good customers.
Bad Customers : Customers who defaulted in payment. By 'default', it means if either or all of the following scenarios have taken place.

- Payment due more than 90 days. In some countries, it is 120 or 180 days.
- Borrower has filed for bankruptcy
- Loan is partially or fully written off

Indeterminates or rollovers : These customers fall into these 2 categories :

- Payment due 30 or max 60 days but paid after that. They are regular late payers.
- Inactive accounts

All the other customers are good customers.

Indeterminates should not be included as it would reduce the discrimination ability to distinguish between good and bad. It is important to note that we include these customers at the time of scoring.

We consider 12 months as performance window to flag defaults which means if a customer has defaulted any time in next 12 months, it would be flagged as 'Bad'

4.2 Methodologies for Estimating PD

There are two main methodologies for estimating Probability of Default.

- Judgmental Method
- Statistical Method

4.2.1 Judgmental Method

It relies on the knowledge of experienced credit professionals. It is generally based on five Cs of the applicant and loan.

- Character : Check credit history of borrower. If no credit history, bank can ask for referees who bank can contact to know about the reputation of borrower.
- Capital : Calculate difference between the borrower's assets (e.g., car, house, etc.) and liabilities (e.g., renting expenses, etc.)
- Collateral : Value of the collateral (security) provided in case borrower fails to repay
- Capacity : Assess borrower's ability to pay principal plus interest amount by checking job status, income etc.
- Conditions includes internal and external factors (e.g. economic recession, war, natural calamities etc.)

Judgmental methods have become past as Statistical methods are more popular these days. But it is still widely used when historical data is not available (especially new credit products).

4.2.2 Statistical Method

In today's world, nobody has time to wait for 1-2 months to know about the status of loan. Also many borrowers apply for loan through bank's website. Hence real-time credit decisions by bank is required to remain competitive in the digital world. The advantage of using statistical method is that it produces mathematical equation which is an automated and faster solution for making credit decisions. This method is unbiased and free from dishonest or fraudulent conduct by loan approval officer or manager.

This method also comes with higher accuracy as statistical and machine learning models considers hundreds of data points to identify defaulters.

4.3 Data Sources for PD Modeling

- Demographic Data : Applicant's age, income, employment status, marital status, no. of years at current address, no. of years at job, postal code

- Existing Relationship : Tenure, number of products, payment performance, previous claims
- Credit Bureau Variables : Default or Delinquency history, Bureau score, Amount of credits, Inquiries etc.

4.4 Steps of PD Modeling

- Data Preparation: pre-processing step in which data from one or more sources is cleaned and transformed to improve its quality prior to its use in business analytics.
- Variable Selection: selecting which variables to include in our model
- Model Development: an iterative process, in which many models are derived, tested and built upon until a model fitting the desired criteria is built
- Model Validation: the task of confirming that the outputs of a statistical model are acceptable with respect to the real data-generating process
- Independent Validation:the process of evaluating a trained model on test data set.
- Supervisory Approval
- caliberation
- Model Implementation : Roll out to users
- Periodic Monitoring
- Post Implementation Validation : Backtesting and Benchmarking
- Model Refinement (if any issue)

5 Our Problem Statement

LendingClub Corp. is an online lender who makes loans online to consumers and sells the loans to investors. The product is a fixed rate, 3 or 4-year fully amortizing term installment loan.

In this project, we analyze the loan data of LendingClub from the year 2007

to the second quarter in 2019. The data explores the demographic dimensions of personal loans and loan status (fully paid/ charged off), which is available on kaggle.

Our goal is to identify the risk of unsecured personal loans. Specifically, a machine learning model was developed to predict the probability of full payment and charge off. On top of that, we utilize the model to predict the expected loan loss from current borrowers, providing insights about risk control and loss reduction.

In our model, to calculate the expected loss, we consider that the $LGD = 100\%$, since if a customer defaults the club loses the entire amount and EAD as the amount lended as it loses the entire amount, here we are taking a simple case and focusing mainly on building the PD model using the process described above.

We first did a preliminary data analysis to look at the different features and then used logistic regression to find the probability of default. We also found a interesting linear relationship between interest rates and default rates. The code and the numerical results are explained in the jupyter notebook attached.