

**Description of the Problem: What the client will be doing**

Financial institutions conduct stress tests to gauge the resilience of their balance sheets to substantial macroeconomic shocks. One way to measure the performance of a financial institution is by assessing the institution's loan portfolio loss under stressed scenarios. The first step in assessing loan loss is to estimate the probability of default (PD).

Understanding PD is necessary for the purpose of stress testing and risk management. Financial institutions may also find it beneficial as insights from default modeling can be incorporated to guide improvements on good underwriting practice and competitive mortgage pricing.

Most of the Models used are statistical models that predict default risk by estimating relationships between default risk and default determinants.

The recent financial crisis has highlighted the importance of understanding macroeconomic conditions in managing credit risk. A few studies have looked at the interaction between the macroeconomic environment and default risk both aggregated and at loan-level.

The proposed model attempts to calculate default probability as a function of macroeconomic variables.

There are various macroeconomic that has correlation with default probability , for this model , we are will use S&P index to get high level aggregate of market conditions.

**Client :**

Any big banks or mortgage holder that hold wide variety of home loans in their portfolio.

**Data Descriptions:(Number of Rows)**

1. In order to capture macroeconomic condition , we will use S&P indexes. The data is sourced from Yahoo Finance site .

Downloaded csv file has following columns

- Date
- Open
- High
- Low
- Close
- Volume
- Adjusted Close

This data will be used to capture market volatility monthly and quarterly.

2. Single Family loan level data from 1999-2016 . from Freddie Mae, one of the biggest mortgage holder.

This data is made available from Freddie Mae . Some of the important fields of the data includes

- Credit Score
- First Time buyer flag
- Mortgage Insurance Percentage
- Occupancy Status
- Debt to Income Ratio
- Original Loan to Value
- Product Type (fixed interest etc)
- Property Type (single family , condo etc)
- Loan Sequence Number.

3. Monthly Loan Performance Data from 1999-2016. Some of the important fields of the data include.

- Loan Sequence Number
- Monthly Reporting Period (YYYYMM)
- Current UPB
- Current Loan Delinquency Status (0,1,2)
- Loan Age
- Remaining months of maturity
- Repurchase Flag
- Zero Balance Code

**Approach :**

The Single family loan data fields and monthly mortgage data to be combined into one data set. The delinquent mortgage are will be compared with quarterly S&P indexes. The model will factor all the relevant mortgage parameter.

The model will try to estimate how default rates respond to shocks in macroeconomic environment over time.

The model will represents (a) the relationship between one variable in the current time period and itself in the past  $p$  time periods, and (b) the relationship between the variable and other variables in the current and past time periods.

**Deliverables**

- Code creating data set modeling and evaluation
- Data Story
- Slide Deck