PROJECT PROGRESS REPORT
Prepayment Modeling for Small Business Loan Asset-Backed Securities

July 4, 2023

Samuel Kang
400177775

Contents

0.1 Abstract

The issue of prepayment for MBS (Mortgage-Backed Securities) or other forms of Asset-Backed Securities poses a challenge in terms of modeling. It requires a rigorous approach that involves complex interest rate modeling and consideration of various factors. In this project, we propose an alternative analytical approach to address this challenge, which can be divided into two key parts.

Firstly, we aim to reduce the number of existing MBS tickers through clustering techniques. By applying clustering analysis, we can group similar MBS together, leading to a smaller number of larger groups. This approach allows us to analyze the prepayment speed for these groups, providing valuable insights into their consistent behaviors. We anticipate that this clustering analysis will help distinguish potentially four groups exhibiting differing prepayment characteristics.

Secondly, we aim to identify statistically significant attributes that can predict prepayment speed. By exploring various factors, including macroeconomic indicators such as interest rates and inflation, we can determine which attributes have the most significant impact on prepayment. Based on this analysis, we will model these attributes as functions to predict prepayment for each group identified in the clustering analysis.

By implementing this alternative approach to prepayment modeling, our objective is to streamline the modeling process while still capturing the essential factors that influence prepayment behavior. This will enhance our understanding of prepayment dynamics for small business loan ABS and improve our predictive capabilities.

Through this project, we seek to contribute to the advancement of prepayment modeling techniques for small business loan ABS, providing valuable insights to investors and industry professionals. Our goal is to enhance the understanding of prepayment dynamics and improve the accuracy of prepayment predictions for these types of asset-backed securities.

0.2 Project Progress

1. June 9

    - Read through the standard prepayment modeling method, and have a high-level idea.

2. June 12

    - When people aggregated these loans, to generate polls of loans. How are pools generated? Not sure they aggregated by asset type? (ie. asset base on - equipment loan? real estate loan? car loan or mix it all together) Investigated through the SBA website, to find out how the pools were generated.

    - The SBA website indicates The loan pool together by 1. the floating rate they are referring 2. the level of the spread that is below/above the floating rate for example, prime rate + 30/60 bp. One of the characteristics for pooling is the range of the spread (prime + spread in between 30-60 bp), then we can calculate the weighted average coupon.
      Base Rate and Adjustment: For variable rate pools, each underlying Loan must use the same base rate, either the Wall Street Journal Prime Rate, the LIBOR Base Rate2, or the SBA Optional Peg Rate, and float on the same accrual basis, either monthly or calendar quarterly

    - Need to be done: to find if there is a clear indication of the type of underlying loans. investigated from the loan assemblers' side (they will take on SBA loans and make them into pools and securitize them and sell to investors), for example, BMO, FHN

3. June 20

    - Go through sample83164lZJ3.xlsx, understand the CPR (Conditional Prepayment Rate)
      write model for 1. define the prepayment speed 2. rank the speed 3. build initial logistics regression for fast/slow(1,0) speed

4. June 29

    - Go through Bloomberg data, to find which variable is useful for our model

0.3 Results

Main Results related to the purpose of the research or things to disccuss

0.4 Model Building

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix,
    accuracy_score

SBA_data = pd.read_excel("SBA_data.xlsx")

x=SBA_data.drop(["CPR_indicator","CPR"],axis=1)
y=SBA_data["CPR_indicator"]

x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.3,
    random_state=1)
logmodel=LogisticRegression()
logmodel.fit(x_train,y_train)
predictions= logmodel.predict(x_test)
classification_report(y_test,predictions)
confusion_matrix(y_test,predictions)
accuracy_score(y_test,predictions)
```

Listing 1: Python example

References