Mortgage-Backed Securities Prepayment Prediction: A Machine Learning Approach

Finance & Commerce

Yu Dong Zhang, lz927

October 2021

Mortgage-backed securities (MBS), the "vicious" origin of 2008 financial crisis, which have severed as a crucial lever for government in managing monetary policies for nearly 40 years, are among the largest financial sector in United States. U.S. mortgage finance market had experienced structural changes after the crisis. Especially during recent couple of years, MBSs are mostly guaranteed by agencies, which lead to huge decrease in credit risk of such product, making them mostly default-free. The investing public had always been trying to solve the world that has sufficient uncertainties to generate excessive economical returns, instead of risk-free entities. This time is no exception. The elimination of credit uncertainties of MBSs has encouraged Wall Street quants to seek profits in the other key factor that drives the value of MBSs, and that is prepayment. Prepayment modelling is among the most complex and novel areas of financial modelling. Its demand has surged significantly only in recent 4 years, and the complexities involve enormous amount of data and factors, as well as the difficulties in model specification and estimation. In this project, I tend to tackle this problem with prepayment modelling. The goal is to model and predict the prepayment risk of agency MBSs, using classification learning algorithms of logistic regression and multi-layer artificial neural networks.

The deep neural network has been applied and evaluated in prepayment modelling. In Zhang (2019) [2], neural network is designed and applied to predict conditional prepayment rate of 30-year mortgage and compared with an industry production model. The result is favorable towards neural network that it produces highly accurate results of not only prepayment rates but also nonlinear risk drivers. In a similar fashion, Amar (2020) [1], models prepayment rate on a loan-level instead of a portfolio-level and achieved wonderful results. However, there is no current research that studies prepayment modelling as both a classification problem and a regression problem. This project, on a general level, not only tends to model prepayment rates as a regression problem, but also aim to predict if a loan will be paid prematurely as a classification problem. In a combination of two results, a higher degree of accuracy should be resulted.

The project will be started off by conducting exploratory data analysis to perform data cleansing and preprocessing on mortgage data obtained from eMBS. The dataset contains 30 mortgage pool attributes and prepayment rates spanning from 2000 to 2020. Principal Component Analysis combined with domain knowledge will be conducted on selecting model features among 30 raw mortgage attributes plus 3 economical factors to achieve the dimensionality reduction of the feature space. Once a handful number of features are selected, learning algorithms can be designed. For a logistic algorithm, monthly mortgage and economical factor data will be taken as input, and whether prepayment will occur on a loan-level for next month is as output. For a regression algorithm, multi-layer deep neural network will be designed that take same input as the logistic model and next month prepayment rate as output. Certain percentage of dataset will be used to train these learning algorithms depending on the training speed. A non-overlapping dataset will be selected to conduct the validation process. During that, error tracking is used to record the prediction error and to measure the power of our prediction, so that the algorithms can be improved to some extent. At last stage, to interpret the model behavior, sensitivity analysis can be performed with a few features changed.

References

[1] Shlomo Amar. Modeling of mortgage loan prepayment risk with machine learning. 2020.

[2] J. "David" Zhang, X. "Jan" Zhao, J. Zhang, F. Teng, S. Lin, and H. "Henry" Li. Agency mbs prepayment model using neural networks. *The Journal of Structured Finance*, 24(4):17–33, 2019.