Using Machine Learning Algorithms to test Asymmetric Information: Application to the U.S. Mortgage Servicing Market

Helmi Jedidi[1] and Georges Dionne

Canada Research Chair in Risk Management
Finance Department, HEC Montreal

28 November 2019

Abstract

The main objective of this paper is to use Machine Learning (ML) to test for the presence of asymmetric information in the U.S. mortgage servicing market. Our paper is the first to consider ML algorithms to predict the likelihood of mortgage default in the U.S. non-agency market and the first to use ML methods to test for asymmetric information. Machine Learning models are ideally suited for mortgage default predictions given their ability to process big data sets and to handle possible nonlinear relationships among large feature sets. In the first part, we evaluate and compare the out-of-sample performance of ML algorithms in predicting mortgage default based on a large sample of U.S. mortgages securitized through the non-agency channel during the period from January 2000 to December 2013. The selected Machine Learning models are: Decision Trees, Naïve Bayes, k-Nearest Neighbors, Support Vector Machines, and Random Forests. Each classification algorithm has a unique approach to process information contained in the feature set as well as a distinct decision-making path. The results show that ML models constantly outperform the standard logistic regression model, regardless the evaluation metric, the study period, or output class imbalances. The results also reveal that tree-based algorithms (Decision Trees and Random Forests) outperform other candidate ML models. In the second part, we utilize feature importance evaluation techniques to highlight how the decision of U.S. lenders to sell the mortgage servicing rights (MSRs) to another financial institution represents a crucial information for predicting mortgage default. Finally, we propose a two-stage estimation procedure to address endogeneity and simultaneity while testing for asymmetric information in a principal-agent context. Our empirical results provide strong support for the presence of second-stage asymmetric information in the market for mortgage servicing rights.

Keywords: Mortgage servicing rights, mortgage default, information asymmetry test, machine learning, classification algorithms, non-parametric tests.

JEL codes: C14, C15, C52, C53, C63, G21, G33.

[1] Corresponding author: *helmi.jedidi@hec.ca*