Grace Deng

Fall 2016

Statistics Honor Thesis Proposal

Purpose

One of the major causes of the 2007 financial crisis was due to an overflow of subprime mortgages and resulting defaults. Before non-government agencies took over the mortgage-backed securities market, most mortgages adhered to the underwriting rules by agencies such as Fannie Mae and Freddie Mac. The goal of this project is to explore single-family loan performance data published by Fannie Mae and analyze the risk of default on mortgage loans using a range of statistical techniques. The Fannie Mae website contains 15 years of data (2000-2015) on the acquisition and performances of 30-year fixed rate mortgage loans, divided by quarters.

Simple Analysis

Data for each quarter is split into two files: Acquisitions and Performance, containing 24 and 29 variables respectively. First, a logit model can be created for predicting the probability of a default based on predictors such as Loan-to-Value ratio, Debt-to-Income ratio, Credit Score, Property Type, etc. Other models, such as Linear Discriminant Analysis and Nonparametric tree models could also be used. Since not all of the loans in the dataset have reached the maturity date, it’s uncertain if loans that are “good” currently will not default in the future. However, it is still possible to create models for a fixed time period and answer questions such as “What is the probability of a default in 5 years? 7 years?”

More Complicated Analysis

A sensitivity analysis could be performed for factors such as Loan-to-Value or Debt-to-Income ratio, since if the cost of retaining a home is too high borrowers may choose to switch to renting instead (signaled by a default). In addition, the dataset categorizes the loans based on property type such as Single-Family, Condominium, Planned Urban Development, etc. Different types of loans could be compared to test for the presence of an “Endowment Effect”, where borrowers may be more willing to pay to keep a residential property, holding other factors such as interest rate and principal constant. Finally, because the Performance dataset includes data for each loan from the starting date up to the current payment period (somewhat like individual time series data), a Monte Carlo simulation could be performed to analyze value at risk or survival time/failure rates.

Data Sources:

Fannie Mae Single-Family Loan Performance Data

<http://www.fanniemae.com/portal/funding-the-market/data/loan-performance-data.html>

Preliminary Research:

[1] <http://www.nber.org/papers/w17516.pdf>

JOHN Y. CAMPBELL & JOÃO F. COCCO, 2015. "[A Model of Mortgage Default,](http://doi.org/10.1111/jofi.12252)" The Journal of Finance, vol 70(4), pages 1495-1554.

[2] <http://www.nber.org/papers/w13882.pdf>

Carmen M. Reinhart & Kenneth S. Rogoff, 2014. "[This Time is Different: A Panoramic View of Eight Centuries of Financial Crises,](http://ideas.repec.org/a/cuf/journl/y2015v16i1reinhartrogoff.html)" Annals of Economics and Finance, Society for AEF, vol. 15(2), pages 1065-1188, November.

[3] <http://www.fic.gov.bc.ca/pdf/fid/14-0877-sup.pdf>

Mingxin Li, 2014. “Residential Mortgage Probability of Default Models and Methods”

[4] <http://www.hkma.gov.hk/media/eng/publication-and-research/quarterly-bulletin/qb200412/fa3.pdf>

Wong, Jim and Fung, Laurence and Fong, Tom and Sze, Angela, Residential Mortgage Default Risk in Hong Kong (November 2004).

[5] <http://www.anderson.ucla.edu/faculty/eduardo.schwartz/articles/53.pdf>

Schwartz, E. S. and Torous, W. N. (1993), Mortgage Prepayment and Default Decisions: A Poisson Regression Approach. Real Estate Economics, 21: 431–449. doi:10.1111/1540-6229.00619