# Estimating Ginnie Mae Prepayment Levels

Charles Naylor & Steven Williams

March 24, 2017

## 1   Introduction

While Ginnie Mae (GNM) securities enjoy the guarantee of the US government, they do not provide a guaranteed schedule of coupons and principle repayment to the investor, as would, for example, a US Treasury bond. Mortgage-backed Security (MBS) pools comprise bundles of individual loans collateralized against American homes. Homeowners enjoy the right, at any time, to pay off the full value of their loan, exposing the investor to an unscheduled return of capital. As homeowners are most likely to do this in credit environments favorable to borrowers, investors may struggle to find debt of similar yield to replace their former investment. It is therefore important, when evaluating MBS pools, to have a good idea of which pools will be more likely to experience prepayments in the near future.

In any given month, these mortgage prepayments represent the aggregate decisions of many homeowners as they react to their own idiosyncratic environment, as well as to general market conditions. The problem of predicting these decisions grows more complex when one considers that the standards of loan underwriters have evolved over time. The ability of individual borrowers to obtain loans at the time of their loans' origination, and those same borrowers' prospects in sourcing new loans to refinance their mortgages, may be subject to differing relationships to the basic forces propelling prepayments as these standards change.

Viewed on a monthly basis, prepayment is a rare event, and thus difficult to predict consistently. Roughly 70% of all pools will experience no prepayment in any specific month. At Nikko, we have attempted to meet these challenges by implementing a Hierarchical Poisson hurdle model.

## 2   Differences Between Ginnie Mae and Fannie Mae Borrowers

Unlike Fannie Mae, which enjoys an implicit guarantee from the federal government, and serves the general American public, Ginnie Mae enjoys an explicit guarantee. GNM finances housing mortgage programs run by various federal
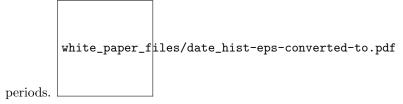
agencies with the stated mission of increasing home ownership among marginalized or otherwise disadvantaged groups. As a result, GNM pools typically contain riskier loans than Fannie Mae pools. Mortgages in GNM pools may include mortgage insurance, both upfront and ongoing payments.

# 3 Modeling

## 3.1 Data

In the aftermath of the credit crisis, Ginnie Mae began providing detailed data on its securities at both the pool-level and on individual loans. These consist of a large number of records providing summary and profile statistics of each pool's underlying borrowers, such as their geographical area, the distributino of average loan sizes, breakdowns of types of loans, etc., as well as basic pool pricing information, such as the pool's security interest rate, issue date, and unpaid principal balance. The data format is ambitious in the breadth of information it provides, but actual records are often incomplete, containing only a subset of the possible data fields.

Another major limitation is the scope of the data. GNM has released records on all extant pools, but only since February, 2012. If we wish to examine prepayments from before then, we need another source of data. Nikko solved this constraint by combining the GNM disclosure data with historical CPR data from Bloomberg. As many pool details are relatively constant, and many of the inputs into our model rely on historical market conditions, it was possible to infer pool characteristics for earlier periods, and thus bulk out our data set to include more of the credit cycle. However, as Bloomberg only provides historical data for investable securities, we have no information on pools that were fully prepaid before February, 2012. There will therefore be some systematic bias in our earlier estimates. However, we are primarily concerned with future prepayments, and consider the addition of data points in rising rate environments to be worth the potential distortions. Sampling remains heavily weighted towards more recent


white_paper_files/date_hist-eps-converted-to.pdf

periods.

- Bayes

  All machine learning is bayesian these days. We have a generative model instead of using straight-up MLE, and that's better because a. we can answer peripheral questions, b. we have more intuitive validation techniques than mere LOO cross-validation.

- Poisson vs Beta (proportional Hazard)

- – Popova model
- 

# 4  Where to Go From Here

- Loan-level
- Issuer-level