

Data Availability and Its Implications for Behavioral Models

Kyle G. Lundstedt, Ph.D.
Andrew Davidson & Co., Inc. (ADCo)

1. Introduction

Investors in the secondary mortgage market have a tremendous need for models of mortgage behaviors (“behavioral models”) such as prepayment and default. Typically, behavioral models are built by combining theoretical and practical knowledge of the mortgage markets with historical data on MBS or loan behavior. Thus, the development of behavioral models is critically dependent on the availability of this data to secondary market researchers.

This article traces the evolution of data availability in the secondary mortgage market, and details the implications of that data availability for the behavioral models used by investors. In particular, we discuss how the changing structure of the mortgage market leads to changes in data availability. We conclude by discussing prospective changes in data availability proposed by Freddie Mac for 2005 Q4, and discuss potential impacts on future behavioral models.

2. Macro Models: Pre-1980

Pre-1980, the mortgage termination experience of the Federal Housing Administration (FHA) provided the only reliable source of mortgage termination and default data. By tracking early termination by cohort (origination year), the FHA published tables that related termination rates to age of a mortgage.

Beyond mortgage age, however, early behavioral models of mortgage termination depended almost entirely on macroeconomic data, primarily prevailing mortgage rates. When current mortgage rates are below the existing mortgage contract rate, households have an incentive to replace their existing fixed rate mortgage with one at a lower rate, thereby reducing their monthly mortgage payments.

Among the earliest examples of such a behavioral model was Curley and Guttentag (1974).¹ Their prepayment model utilized ordinary least squares (OLS) on mortgage cohorts because information on the individual loans or pools was unavailable.

The dependent variable in such behavioral models typically was the proportion of a given cohort that prepaid in a given period. The independent variables were limited to age and interest-rate related variable for each cohort. Thus, limited data availability resulted in fairly simple models.

3. Pool-Level Models: 1980-1990

The birth of the secondary mortgage market in the 1980's provided the catalyst for better empirical models of mortgage termination. While Ginnie Mae issued its first MBS in 1970 and Freddie Mac its first MBS in 1971, the agency market remained small until Fannie Mae issued its

¹ Curley, A. J. and Guttentag, J. M. (1974). Explorations in Economic Research, “The Yield on Insured Residential Mortgages”, pages 114-161. National Bureau of Economic Research.

first security in 1981. At that point, the agency market, as defined by conforming loan size and underwriting standards set by the three agencies, began its explosive growth.

As originating lenders began to sell many of their mortgages into the burgeoning agency market, investors demanded better information on the termination-related factors that dictated the returns on their mortgage-related investments. Because the agencies' MBS were collateralized by pools of mortgages, the agencies began to disclose a basic set of attributes describing each pool at origination – weighted average characteristics such as coupon, maturity, etc. – and the overall termination factor for each pool on a monthly basis.

Peters, Pinkus, and Askin (1984) and Richard and Roll (1989) are good examples of the first behavioral models built upon this newly available agency data. Using Freddie Mac pool information, they utilized as dependent variables the aggregated characteristics of the mortgage pools, such as average loan size, average loan age, and weighted average coupon (WAC).

Like the macro models, the dependent variable generally was the proportion of the pool that terminated each month. However, pool-level models often utilized more sophisticated statistical techniques, such as logistic regression, that constrained the outcome to between 0 and 100%.

Given the continuing importance of the agency market, pool-level models remain in heavy use throughout the mortgage market. Because the available information is relatively sparse (compared to the large number of attributes of the underlying mortgages in a pool), researchers have devoted significant resources to using these data as effectively as possible.²

4. Non-Agency Loan-level Models: 1990 and beyond

The next advance in data availability came about during the resolution of the S&L crisis in the early 1990's. The Resolution Trust Corporation (RTC), in its role disposing of the assets of failed thrifts, set new standards of data disclosure, providing very detailed information on the collateral underlying its securities. As the non-agency market began to develop during this same period, non-agency issuers followed the RTC's lead in terms of data availability, disclosing loan-level data in order to induce investors to hold their securities.

Hence, "loan-level" analyses were the next mortgage modeling innovation resulting from changes in data availability. These loan-level behavioral models attempted to explain why, despite good performance by pool-level models, apparently identical pools still showed a significant level of unexplained variation in termination behavior.

For example, Stern (2002) looked jumbo, non-conforming fixed rate residential mortgages, characterized by both their large original balances and their lack of any private or government sponsored mortgage insurance.³ While the prepayment characteristics of jumbo fixed rate loans were similar to agency loans, the additional loan level variables used generally increased the precision of such behavioral models.

² See Davidson, A., et. al. (2003), Securitization: Structuring and Investment Analysis, John Wiley & Sons.

³ Stern, H. (June 2002), "Residential Jumbo Loan Level Prepayment Model", Andrew Davidson & Co., Inc.

In addition, rather than estimating a conditional termination rate (fraction of pool terminating), many researchers actually observed whether or not individual loans terminated in a given month. As a result, they took advantage of advanced statistical techniques, such as hazard models, that explained both "if" and "when" a loan would terminate.

Non-agency loan-level models have taken on additional importance as the non-agency market has grown tremendously in the last few years. For example, according to Inside MBS & ABS (Sept. 16, 2005), over \$1 trillion in agency MBS were issued by Fannie, Freddie, and Ginnie in 2004, while the non-agency market produced \$864 billion.⁴ However, as of the end of August 2005, the \$598 billion in agency MBS produced year-to-date is far exceeded by the \$739 billion in non-agency securitizations.

5. **Enhanced Pool-Level Models: 2003 and beyond**

As investors became familiar with the investment characteristics of both agency and non-agency markets, they began to pressure the agencies to provide levels of data disclosure similar to the data availability in the non-agency market. Consequently, important changes in agency data availability have occurred or are in the offing.

First, in June 2003, Fannie Mae and Freddie Mac both began making available additional pool-level characteristics for new issues. This data including weighted averages of important variables such as loan size, LTV, and credit score. Though this change was fairly recent, mortgage models have changed to take advantage of this "enhanced" pool-level data from the past several years.

For example, the ADCo Enhanced Prepayment Model augments the ADCo pool-level models by considering additional pool level characteristics that help to fine-tune forecasts. By incorporating weighted averages for loan size, LTV, FICO score, states of origination, pool percentage of multi-family vs. single family, and refi vs. purchase, this enhanced model captures prepayment differentials across individual pools.

Second, Freddie Mac has announced that, in addition to its existing pool-level data, it will soon make almost 40 loan-level characteristics at issuance available to investors. This loan-level data at issuance will include both fixed- and adjustable-rate mortgage characteristics, as well as critical attributes such as loan purpose, occupancy status, and prepayment penalty flag. Though this data will not be available until sometime in the fourth quarter of 2005, this enormous increase in data availability certainly will provide rich fodder for behavioral models in the near future.

⁴ Interestingly, Fannie and Freddie purchased almost \$212 billion of that non-agency production.