

Appendix A: Econometric Analysis of Mortgages

This appendix describes the technical details of the econometric models used to estimate the historical and future performance of FHA single-family loans for the FY 2008 Review. The overall modeling approach remains consistent with that applied in previous years, with three significant model changes undertaken for the FY 2008 Review:

- Introduction of a measure of the initial level of house price for a property securing an FHA loan relative to the local area median house price,
- Introduction of an indicator to distinguish purchase versus refinance originations,
- Improvement in the modeling of the impacts of borrower credit scores on FHA claim and prepay terminations.

Each of these changes is described in greater detail below.

Section I of this appendix summarizes the model specification and estimation issues arising from the analysis of FHA claim and prepayment rates. We discuss issues related to differences in the timing of borrower default episodes and prepayment and claim terminations, followed by a review of the mathematical derivation of multinomial logit probabilities from the separate binomial logit estimates. We then turn to a description of the historical loan event history data needed for estimation and the future loan records required for forecasting future loan performance. Section II describes the specific explanatory variables used in the analysis and Section III presents the logit estimation results for the separate loan product models.

I. Model Specification and Estimation Issues

A. Specification of FHA Mortgage Termination Models

Competing risk models for mortgage prepayment and claim terminations were specified and estimated for the FY 2008 Review. Prepayment- and claim-rate estimates were based on a multinomial logit model for quarterly conditional probabilities of prepayment and claim terminations. The general approach is based on the multinomial logit models reported by Calhoun and Deng (2002) that were originally developed for application to OFHEO's risk-based capital adequacy test for Fannie Mae and Freddie Mac. The multinomial model recognizes the competing-risks nature of prepayment and claim terminations. The use of quarterly data aligns closely with key economic predictors of mortgage prepayment and claims such as changes in interest rates and housing values.

The loan performance analysis was undertaken at the loan level. Through the use of categorical explanatory variables and discrete indexing of mortgage age, it was possible to achieve considerable efficiency in data storage and reduced estimation times by collapsing the data into a much smaller number of loan strata (i.e., observations). In effect, the data were transformed into synthetic loan pools, but without loss of detail on individual loan characteristics beyond that implied by the original categorization of the explanatory variables, which were entirely under our control. Sampling weights were created to account for differences in the number of loans in each loan strata.

The present analysis extended the Calhoun-Deng (2002) study in two important ways. First, following the approach suggested by Begg and Gray (1984), we estimated separate binomial logit models for prepayment and claim terminations, and then mathematically recombined the parameter estimates to compute the corresponding multinomial logit probabilities. This approach allowed us to account for differences between the timing of claim terminations and the censoring of potential prepayment outcomes at the onset of default episodes that ultimately lead to claims. This issue is discussed in greater detail below.

A second extension of the Calhoun-Deng (2002) study was the treatment of the age of the mortgage in the models. The traditional models applied quadratic age functions for both mortgage default and prepayment terminations. While the quadratic age function fits reasonably well for estimating conventional mortgage defaults rates, it performed less well for prepayments, as it failed to capture the more rapid increase in conditional prepayment rates early in the life of the loans. FHA conditional claim and prepayment rates also show a more rapid increase than conventional mortgages during their early loan life. We found a quadratic specification not to be sufficiently flexible to capture the age patterns of conditional claim and prepayment rates observed in the FHA data. The approach we adopted was to apply piece-wise linear spline functions. This approach is sufficiently flexible to fit the relatively rapid increase in conditional claim and prepayment rates observed during the first three years following mortgage origination, while still providing a good fit over the later ages and still limiting the overall number of model parameters that have to be estimated.

The starting point for specification of the loan performance models was a multinomial logit model of quarterly conditional probabilities of prepayment and claim terminations. The corresponding mathematical expressions for the conditional probabilities of claim ($\pi_c(t)$), prepayment ($\pi_p(t)$), or remaining active ($\pi_A(t)$) over the time interval from t to $t+1$ are given by:

$$\pi_c(t) = \frac{e^{\alpha_c + X_c(t)\beta_c}}{1 + e^{\alpha_c + X_c(t)\beta_c} + e^{\alpha_p + X_p(t)\beta_p}} \quad (1)$$

$$\pi_p(t) = \frac{e^{\alpha_p + X_p(t)\beta_p}}{1 + e^{\alpha_c + X_c(t)\beta_c} + e^{\alpha_p + X_p(t)\beta_p}} \quad (2)$$

$$\pi_A(t) = \frac{1}{1 + e^{\alpha_c + X_c(t)\beta_c} + e^{\alpha_p + X_p(t)\beta_p}} \quad (3)$$

where the constant terms α_c and α_p and the coefficient vectors β_c and β_p are the unknown parameters to be estimated for the multinomial logit model. The subscripts “P” and “C” denote prepayments and claims. We denote by $X_c(t)$ the vector of explanatory variables for the conditional probability of a claim termination, and $X_p(t)$ is the vector of explanatory variables for the conditional probability of prepayment. Some components of $X_c(t)$ and $X_p(t)$ are constant over the life of the loan and therefore do not vary with t .

B. Differences in the Timing of Borrower Default Episodes and Claim Terminations

The primary events of interest in the present context are mortgage prepayments that result in termination of positive cash flows from mortgage premiums paid by borrowers, and claim terminations that result in direct payouts to lenders. For consistency with the available data on loss rates, the incidence and timing of mortgage default-related terminations is defined specifically according to FHA claim incidences, although these typically arise from earlier decisions by borrowers to cease payment on their mortgages. In recognition of the potential censoring of prepayment prior to the actual claim termination date, we used information on the timing of the initiation of default episodes leading to claim terminations to create a prepayment-censoring indicator that was applied when estimating the prepayment-rate model, in effect removing that observation from the sample at risk of prepayment whenever it was clear from the details of the delinquency/default/claim sequence that the probability of prepayment was zero. Implementation of this strategy required estimating the prepayment function separately from that for claims. The Begg-Gray method of estimating separate binomial logit models is particularly advantageous in dealing with this requirement while preserving consistency with the competing risks multinomial logit model outlined above.

To complete the model, a separate binomial logit claim-rate model was estimated accounting for censoring of potential claim terminations by observed prepayments, and the two sets of parameter estimates were recombined mathematically to produce the final multinomial model for conditional prepayment and claim probabilities. This approach facilitated unbiased estimation of the prepayment function, which would not be possible in a joint multinomial model of claim and prepayment terminations, since one could not simultaneously censor loans at the onset of default episodes and still retain the observations for estimating subsequent claim termination rates.

To summarize, estimation of the multinomial logit model for prepayment and claim terminations involved the following steps:

- Data on the start of a default episode that ultimately leads to an FHA claim was used to define a default-censoring indicator for prepayment.
- A binomial logit model for conditional prepayment probabilities was estimated using the default-censoring indicator to truncate individual loan event samples at the onset of any default episodes (and all subsequent quarters).
- A binomial logit model for conditional claim probabilities was estimated using observed prepayments to truncate individual loan event samples during the quarter of the prepayment event (and all subsequent quarters).
- The separate sets of binomial logit parameter estimates were recombined mathematically to derive the corresponding multinomial logit model for the joint probabilities of prepayment and claim terminations accounting for the competing risks.

C. Computation of Multinomial Logit Parameters from Binomial Logit Parameters

Begg and Gray applied Bayes Law for conditional probabilities to demonstrate that the values of parameters α_C , β_C , α_P , and β_P estimated from separate binomial logit (BNL) models of claims and prepayments are identical to those for the corresponding multinomial logit (MNL) model once the appropriate calculations are performed. Assume that conditional probabilities for claim and prepay terminations for separate BNL models are given, respectively, by:

$$\pi_{BNL}^C = \frac{e^{\alpha_C + X_C \beta_C}}{1 + e^{\alpha_C + X_C \beta_C}}, \quad \pi_{BNL}^P = \frac{e^{\alpha_P + X_P \beta_P}}{1 + e^{\alpha_P + X_P \beta_P}}. \quad (4)$$

We have suppressed the time index t to simplify the notation. We can rearrange terms to solve for components $e^{\alpha_C + X_C \beta_C}$ and $e^{\alpha_P + X_P \beta_P}$ of the multinomial model in terms of binomial probabilities π_{BNL}^C and π_{BNL}^P , respectively,

$$e^{\alpha_C + X_C \beta_C} = \frac{\pi_{BNL}^C}{(1 - \pi_{BNL}^C)}, \quad e^{\alpha_P + X_P \beta_P} = \frac{\pi_{BNL}^P}{(1 - \pi_{BNL}^P)}. \quad (5)$$

Then we can substitute directly into the MNL probabilities shown in equations (1) and (2) for $e^{\alpha_C + X_C \beta_C}$ and $e^{\alpha_P + X_P \beta_P}$:

$$\pi_{MNL}^C = \frac{\frac{\pi_{BNL}^C}{(1 - \pi_{BNL}^C)}}{1 + \frac{\pi_{BNL}^C}{(1 - \pi_{BNL}^C)} + \frac{\pi_{BNL}^P}{(1 - \pi_{BNL}^P)}}, \quad \pi_{MNL}^P = \frac{\frac{\pi_{BNL}^P}{(1 - \pi_{BNL}^P)}}{1 + \frac{\pi_{BNL}^C}{(1 - \pi_{BNL}^C)} + \frac{\pi_{BNL}^P}{(1 - \pi_{BNL}^P)}}. \quad (6)$$

These expressions for the MNL probabilities can be simplified algebraically to:

$$\pi_{MNL}^C = \frac{\pi_{BNL}^C \cdot (1 - \pi_{BNL}^P)}{(1 - \pi_{BNL}^C \cdot \pi_{BNL}^P)}, \quad \pi_{MNL}^P = \frac{\pi_{BNL}^P \cdot (1 - \pi_{BNL}^C)}{(1 - \pi_{BNL}^C \cdot \pi_{BNL}^P)}. \quad (7)$$

Equations (7) were used to derive the corresponding MNL probabilities directly from separately estimated BNL probabilities.

D. Loan Event Data

We used loan-level data to reconstruct quarterly loan event histories by combining mortgage origination information with contemporaneous values of time-dependent factors. In the process of creating quarterly event histories, each loan contributed an additional observed “transition” for every quarter from origination up to and including the period of mortgage termination, or until the last time period of the historical data sample. The term “transition” is used here to refer to any period in which a loan remains active, or in which claim or prepayment terminations are observed.

The FHA single-family data warehouse records each loan for which insurance was endorsed and includes additional data fields updating the timing of changes in the status of the loan. The historical data used in model estimation for this Actuarial Review is based on an extract from FHA’s database as of March 31, 2008. The data set was first filtered for loans with missing or invalid values of key variables in our econometric model. In addition, there is a subset of historical loans where the payoff status of the loans was never updated, to which FHA has assigned a special servicer identification code. Most of those loans were believed to have already been prepaid but the records were not yet updated. Since FY 2004, HUD has been investigating and updating the performance records of these loans. The remaining loans from these servicers were deleted from the sample used for model estimation following preliminary statistical analysis that confirmed there would be no material impact on the final econometric estimates.

A dynamic event history sample was constructed from the database of loan originations by creating additional observations for each quarter that the loan was active from the beginning amortization date up to and including the termination date for the loan, or the end of the second quarter of FY 2008 if the loan was not terminated prior to that date. Additional “future” observations were created for projecting the future performance of loans currently outstanding, and additional future cohorts and transition periods were created to enable simulation of the performance of future books of business. These aspects of data creation and simulation of future loan performance are discussed in greater detail in Appendix C.

E. Sampling Issues

A full 100-percent sample of loan-level data from the FHA single-family data warehouse was extracted for the FY 2008 analysis. This produced a very large sample of approximately 23 million single-family loans originated between the first quarter of FY 1975 and the first quarter of FY 2008. These data were used to generate loan-level event histories for up to 120 quarters (30 years) of loan life per loan or until the age at which the loan would mature based on the original term of the loan when the term is less than 30 years.

Estimation and forecasting was undertaken separately for each of the following six FHA mortgage product types:

1. FRM30 Fixed-rate 30-year fully-underwritten purchase and refinance mortgages
2. FRM15 Fixed-rate 15-year fully-underwritten purchase and refinance mortgages
3. ARM Adjustable-rate fully-underwritten purchase and refinance mortgages
4. FRM30_SR Fixed-rate 30-year streamlined refinance mortgages
5. FRM15_SR Fixed-rate 15-year streamlined refinance mortgages
6. ARM_SR Adjustable-rate streamlined refinance mortgages

We used a 20-percent random sample of FRM30 mortgages and 100-percent samples for all other product types for estimation. For forecasting future loan performance we used an 8-percent sample for FRM30 and a 20-percent sample for FRM30_SR mortgages.

II. Explanatory Variables

Four main categories of explanatory variables were developed:

1. Fixed initial loan characteristics; including mortgage product type, purpose of loan (home purchase or refinance), amortization term, origination year and quarter, original loan-to-

value (LTV) ratio, relative house price level, original loan amount, original mortgage interest rate, and geographic location (MSA, state, Census division);

2. Fixed initial borrower characteristics; including borrower credit scores and indicators of the source of downpayment assistance (additional discussion of borrower credit scores and downpayment assistance is provided below);
3. Dynamic variables based entirely on loan information; including mortgage age, season of the year, and scheduled amortization of the loan balance; and
4. Dynamic variables derived by combining loan information with external economic data; including interest rates and house price indexes.

In some cases the two types of dynamic variables are combined, as in the case of adjustable-rate mortgage (ARM) loans where external data on changes in Treasury yields are used to update the original coupon rates and payment amounts on ARM loans in accordance with standard FHA loan contract features. This in turn affects the amortization schedule of the loan.

Exhibit A-1 summarizes the explanatory variables that are used in the statistical modeling of loan performance. All of the variables except for mortgage age listed in Exhibit A-1 were entered as 0-1 dummy variables in the statistical models. For each set of categorical variables, one of the dummy variables is omitted during estimation and serves as the baseline category. The mortgage age variable was entered as a piecewise linear spline function. The specification of each variable is described in more detail below.

Mortgage Product Types

As described above, separate statistical models were estimated for the following six FHA mortgage product types:

1. FRM30 Fixed-rate 30-year fully-underwritten purchase and refinance mortgages
2. FRM15 Fixed-rate 15-year fully-underwritten purchase and refinance mortgages
3. ARM Adjustable-rate fully-underwritten purchase and refinance mortgages
4. FRM30_SR Fixed-rate 30-year streamlined refinance mortgages
5. FRM15_SR Fixed-rate 15-year streamlined refinance mortgages
6. ARM_SR Adjustable-rate streamlined refinance mortgages

Specification of Piece-Wise Linear Age Functions

Exhibit A-1 lists the series of piece-wise linear age functions that were used for the six different mortgage product types. For example, we created a piece-wise linear age function for FRM15

loans with knots (the k's) at 2, 4, 8, and 12 quarters by generating 5 new age variables age1 to age5 defined as follows:

$$\begin{aligned}
 \text{age1} &= \begin{cases} \text{AGE} & \text{if AGE} \leq k_1 \\ k_1 & \text{if AGE} > k_1 \end{cases} \\
 \text{age2} &= \begin{cases} 0 & \text{if AGE} \leq k_1 \\ \text{AGE} - k_1 & \text{if } k_1 < \text{AGE} \leq k_2 \\ k_2 - k_1 & \text{if AGE} > k_2 \end{cases} \\
 \text{age3} &= \begin{cases} 0 & \text{if AGE} \leq k_2 \\ \text{AGE} - k_2 & \text{if } k_2 < \text{AGE} \leq k_3 \\ k_3 - k_2 & \text{if AGE} > k_3 \end{cases} \\
 \text{age4} &= \begin{cases} 0 & \text{if AGE} \leq k_3 \\ \text{AGE} - k_3 & \text{if } k_3 < \text{AGE} \leq k_4 \\ k_4 - k_3 & \text{if AGE} > k_4 \end{cases} \\
 \text{age5} &= \begin{cases} 0 & \text{if AGE} \leq k_4 \\ \text{AGE} - k_4 & \text{if AGE} > k_4 \end{cases} \tag{8}
 \end{aligned}$$

Coefficient estimates corresponding to the slopes of the line segments between each knot point and for the last line segment are estimated and reported in Exhibit A-2. The overall AGE function (for this 5-age segment example) is given by:

$$\text{Age Function} = \beta_1 \cdot \text{age1} + \beta_2 \cdot \text{age2} + \beta_3 \cdot \text{age3} + \beta_4 \cdot \text{age4} + \beta_5 \cdot \text{age5} \tag{9}$$

Age functions with greater or fewer numbers of segments were developed in a similar manner. The number of segments and the selection of the knot points are determined by experimentation based on the in-sample fit for conditional claim and prepayment rates.

Relative House Price

In this Review we introduced a variable measuring the relative house price level within the local market. The relative house price variable was computed by comparing the original purchase price of the house underlying a particular mortgage with the median house value in the same time period and location. HUD provided us with Census median house price data at the county and metropolitan Core Based Statistical Area (CBSA) levels for the years 1980, 1990, 2000 and 2006. Quarterly median price estimates for all time periods from 1980 to 2008 were derived through linear interpolation, except that values back to 1975 were imputed by discounting 1980 values based on an assumption of 3-percent annualized growth in house prices from 1975 to 1980. The CBSA median prices estimates were applied to FHA loans with properties located in those metropolitan areas. We derived separate state-wide non-metro median house price estimates using the Census county-level median data for all non-metro counties within a state. The non-metro state values were computed by taking the median of the county (median) values.

Loan Size

Loan size is defined relative to the average-sized FHA loan originated in the same state during the same fiscal year. The resulting values were stratified into 5 categories based on direct examination of the data, with the middle category, *category 3*, centered on the average-sized loans plus or minus 10 percent, *i.e.*, 90 to 110 percent of the average loan size.

Loan-to-Value Ratio

Initial loan-to-value is recorded in FHA's data warehouse. Based on discussions with FHA, any LTV values recorded for streamline refinance products may refer to values recorded at the time of the original FHA loan and were considered unreliable for use in the analysis. We imputed original LTV values for these loans for the purpose of establishing the starting point for tracking the evolution of the probability of negative equity (see description of this variable below). The imputed values were based on the mean LTV values for non-streamlined products FRM30, FRM15, and ARM loans stratified by product, beginning amortization year and quarter, and geographic location (state and county). The imputed LTV values do not provide good fits for these streamline mortgages. However, the "probability of negative equity" variable discussed below, built upon these imputed initial LTV values, appeared to have good explanatory power.

Season

The season of an event observation quarter is defined as the season of the year corresponding to the calendar quarter, where 1 = Winter (January, February, March), 2 = Spring (April, May, June), 3 = Summer (July, August, September), and 4 = Fall (October, November, December).

Probability of Negative Equity

Following the approach applied by Deng, Quigley, and Van Order (2000), Calhoun and Deng (2002), and others, we computed the equity positions of individual borrowers using *ex ante* probabilities of negative equity. The probability of negative equity is a function of the current loan balance and the probability of individual house price outcomes that fall below this value during the quarter of observation. The distributions of individual housing values relative to the value at mortgage origination were computed using estimates of house price drift and volatility based on OFHEO House Price Indexes (HPIs).

The probability of negative equity is computed as follows:

$$PNEQ = \Phi \left\{ \frac{\ln(UPB(t)) - \ln(P(0) \cdot HPI(t))}{\sigma(t)} \right\} \quad (10)$$

where $\Phi(x)$ is the standard normal cumulative distribution function evaluated at x , $UPB(t)$ is the current unpaid mortgage balance based on scheduled amortization, $P(0)$ is the value of the borrower's property at mortgage origination, $HPI(t)$ is an index factor for the percentage change in housing prices in the local market since origination of the loan, and $\sigma(t)$ is a measure of the diffusion volatility for individual house price appreciation rates over the same period of time. The values of $HPI(t)$ are computed directly from the house price indexes published by OFHEO, while the diffusion volatility is computed from the following equation:

$$\sigma(t) = \sqrt{a \cdot t + b \cdot t^2} . \quad (11)$$

The parameters “ a ” and “ b ” in this expression are estimated by OFHEO when applying the three-stage weighted-repeat-sales methodology advanced by Case-Shiller (1987, 1989). Further details on the OFHEO HPI methodology can be found in Calhoun (1996).

The resulting values of PNEQ were stratified into seven levels ranging from less than 5-percent to more than 30-percent probability of negative equity as listed in Exhibit A-1. Further mathematical details are presented in Appendix C of this Review.

Mortgage Premium (Refinance Incentive)

The financial incentive of a borrower to refinance is measured using a variable for the relative spread between the current mortgage contract interest rate and the current market mortgage rate:

$$MP(t) = \left\{ \frac{C(t) - R(t)}{C(t)} \right\}. \quad (12)$$

Where $C(t)$ is the current note rate on the mortgage and $R(t)$ is the current market average fixed-rate mortgage rate. This variable is as an approximation to the call option value of the mortgage given by the difference between the present value of the “anticipated” future stream of mortgage payments discounted at the current market rate of interest, $R(t)$, and the present value of the mortgage evaluated at the current note rate, $C(t)$. Additional details are given in Deng, Quigley, and Van Order (2000) and Calhoun and Deng (2002).

The relative mortgage premium values for ARMs and FRMs are derived in exactly the same manner, except that the current coupon is always equal to the coupon at origination for FRMs, whereas ARM coupon rates are updated over the life of the mortgage as described below.

ARM Coupon Rate Dynamics

To estimate the current financial value of the prepayment option for ARM loans, and to compute amortization rates that vary over time, we needed to track the path of the coupon rate over the active life of individual ARM loans. The coupon rate resets periodically to a new level that depends on the underlying index, plus a fixed margin, subject to periodic and lifetime caps and floors that specify the maximum and minimum amounts by which the coupon can change on each adjustment date and over the life of the loan. Accordingly, the ARM coupon rate at time t , $C(t)$, was computed as follows:

$$C(t) = \max\{ \min[\text{Index}(t - S) + \text{Margin}, \\ C(t - 1) + A(t) \cdot \text{Period_UpCap}, C(0) + \text{Life_UpCap}], \\ C(t - 1) - A(t) \cdot \text{Period_DownCap}(t), \max(C(0) - \text{Life_DownCap}, \text{Life_Min}) \} \quad (13)$$

where $\text{Index}(t)$ is the underlying rate index value at time t , S is the “lookback” period, and Margin is the amount added to $\text{Index}(t - S)$ to obtain the “fully-indexed” coupon rate. The periodic adjustment caps are given by Period_UpCap and Period_DownCap , and are multiplied by dummy variable $A(t)$ which equals zero except during scheduled adjustment periods. Maximum lifetime adjustments are determined by Life_UpCap and Life_Down_Cap , and Life_Min is the overall minimum lifetime rate level. Any initial discounts in ARM coupon rates are reflected in the original interest rate represented by $C(0)$ in equation (13).

Yield Curve Slope

Expectations about future interest rates and differences in short-term and long-term borrowing rates associated with the slope of the Treasury yield curve influence the choice between ARM and FRM loans and the timing of refinancing. We use the ratio of the ten-year Constant Maturity Treasury (CMT) yield to the one-year CMT yield to measure the slope of the Treasury yield curve.

Burnout Factor

A burnout factor is included to identify borrowers who have foregone recent opportunities to refinance. The burnout factor is included to account for individual differences in propensity to prepay, often characterized as unobserved heterogeneity. In addition, unmeasured differences in borrower equity at the loan level may give rise to unobserved heterogeneity that can impact both prepayment and claim rates. Borrowers with negative equity are less likely to prepay due to the difficulty of qualifying and are more likely to exercise the default option.

Changes were introduced to the burnout factor for the FY 2006 Review and continue to be applied in the FY 2008 Review. The previous burnout factor, which was identical to that used in the OFHEO risk-based capital stress test model, took the value one if the mortgage note rate exceeds the market mortgage rate by 200 basis points or more in any two of the preceding eight quarters. Empirical evidence now suggests that borrowers who refinance tend to do so at much lower thresholds. The burnout factor is quantified as the moving average number of basis points the borrower was in the money, for all quarters during which the borrower was in the money, during the preceding 8 quarters. The resulting measure was categorized into 50 basis point categories corresponding to 0 (always out of the money) up to a category corresponding to a moving average value exceeding 200 basis points, for a total of 6 categories.

Exposure Year/Quarter FRM Rate

A variable measuring the market average FRM mortgage rate is included to distinguish high-rate and low-rate market environments. This variable was categorized into 100 basis point categories indicating market average FRM mortgage rates of 6 percent or less up to a category for market average FRM rates exceeding 10 percent.

Metropolitan Area Unemployment Rates

We previously developed a measure of changes in metropolitan area unemployment rates. Data on metropolitan area unemployment rates were obtained from the Bureau of Labor Statistics and converted into time series from which we computed a dynamic measure for the percentage change in the unemployment rate over the preceding year.

The unemployment rate variables did not perform well in any of the preliminary models that were estimated, and have not been included in the final model specifications. No consistent pattern was observed between mortgage claims and increases in local area unemployment rates, in contrast to the strong relationship between loan performance and borrower equity. This outcome is consistent with prior experience using this variable in loan-level models in which borrower behavior is more strongly linked to changes in the borrower's equity position or changes in the value of the mortgage instrument due to changes in interest rates. Changes in these variables have a direct impact on property and mortgage values, whereas the local area unemployment measure has a much weaker connection to individual borrowers.

ARM Payment Burden

Another variable previously considered for the Review was the ARM payment burden. This variable measured the percentage change in the monthly payment since origination. The percentage change was categorized into 5 levels ranging from no increase to more than a 30-percent increase.

The ARM payment burden variables did not perform well in the preliminary models that were estimated and were generally not statistically significant. This variable is highly collinear with the mortgage premium (spread) and burnout variables (for loans that do not prepay), particularly over the early years before there is substantial amortization of the loan balance. As a result, this variable contributes little to the explanation of loan performance once the other variables are included and is not included in the ARM product models for the FY 2008 Review.

Source of Downpayment assistance

As documented in the FY 2006 and FY 2007 Reviews, the FHA single-family program recently experienced a significant increase in the use of downpayment assistance from relatives, non-profit organizations, and government programs. Loans to borrowers utilizing downpayment assistance from non-profit organizations have been observed to generate significantly higher claim rates. Following the approach first applied for the FY 2006 Review, we have included in this year's Review a series of indicators to control for the use of different types of downpayment assistance by FHA borrowers.

Borrower Credit Scores

Borrower credit scores at the loan level were first included in the models estimated for the FY 2007 Review and continue to be an important predictor of claim and prepayment behavior. FHA has relatively complete data on borrower FICO scores for loans originated since May 2004. In addition, FHA retroactively obtained borrower credit history information for selected samples of FHA loan applications submitted as far back as FY 1992. These data provide an additional source of loan-level information on borrower FICO scores that are used for estimation. Historical FICO score data was collected for HUD by Unicon Corporation for FHA applications dated during FYs 1992, 1994, and 1996. FICO scores of the borrower and up to two co-applicants were collected from a single credit data repository for a random sample of approximately 20 percent of loan applications. A second set of sample data was collected for loan applications over the period from FY1997 to FY 2001. FICO scores for up to three co-applicants were collected from up to two credit data repositories for about 20 percent of the loans in each year, with over-sampling of loans defaulted by April 2003. A third and final set of data, similar to the second set, was collected for FY 2002 to FY 2005 applications, with over-sampling of loans defaulted by February 2005. The over-sampling of historical borrower credit scores for default outcomes introduces issues of choice-based sampling. These issues are addressed in a separate section below.

These three sets of FICO data represent the most reliable sources of borrower credit history information available for historical FHA-endorsed loans. Following the methodology adopted by Freddie Mac and Fannie Mae, the FICO score of each individual borrower or co-borrower, respectively, is the median (of three) or minimum (of two) scores when scores are provided by multiple credit data repositories. The final FICO score assigned to a loan is the simple average of these individual FICO scores for the borrower and up to four co-borrowers. FICO scores derived in this manner were further stratified into categorical outcomes for use in the estimation models.

Additional indicator variables were specified to represent two particular forms of missing data on FICO scores. The categorical outcome 000 was defined corresponding to loans originated in FY 1992 or later that were known to have been submitted for scoring to one more credit data repository, but for which the borrower credit history was insufficient to generate a FICO score. The categorical outcome 999 was defined corresponding to loans originated in FY 1992 or later for which no attempt was made to obtain the FICO score.

Finally, an indicator was defined to distinguish loans with FICO scores obtained through the normal FHA loan approval process from loans for which FICO scores were obtained from the retrospective historical sampling procedure conducted by Unicon Corporation. There are some months in FY 2004 for which both types of FICO scores are present in the data. This variable

was included to control for the potential effect of choice-based sampling due to the oversampling of defaulted loans.

Choice-Based Sampling of Historical FICO Scores and Random Sampling of FRM30 Loans

As described in Section I of this Appendix, a 20-percent random sample of FRM30 loans was used for estimation and forecasting of claim and prepay rates. A stratified random sampling scheme was applied to assure adequate representation of loans with historical FICO score data. For each fiscal year the Unicon sample loans were flagged and the total counts of Unicon loans and other FHA loans were computed. Separate sampling rates for Unicon loans and other FHA loans were derived to give as close to equal representation of both loan samples as possible, while still achieving an overall sampling rate of 20 percent for the particular FY. Individual sampling weights were assigned to each loan based on the reciprocal of their probability of selection. In some years this resulted in selecting the entire sample of available Unicon sample loans, with the remainder of the 20 percent sample comprising FHA loans not included in the Unicon samples. In other years, this resulted in selecting a random subsample of Unicon sample loans and an equally-sized random sample of other FHA loans. Our goal was to attain a balanced mix of loans with and without FICO scores (for those years in which FICO scores were potentially available) in order to analyze the impact of credit scores on loan performance and to control for choice-based sampling of FICO scores by comparison to loan performance in a random sample of FHA loans. Under the approach outlined here the estimation data included a mix of randomly sampled FHA loan originations without FICO scores and a choice-based sample of loans with FICO scores prior to 2004, and randomly sampled FHA loans with FICO scores since late 2004.

Estimation using only observations from a choice-based sample is known to result in biased estimation of the constant terms of maximum-likelihood logit probability models, but still gives unbiased estimates of the coefficients of the explanatory variables. The standard correction for bias in the intercept terms depends on the relative population and sample proportions of the selected outcome (Costlett, 1981). It is not feasible to apply this type of correction in our case, as the original procedure was applied to a sample of FHA loan “applications,” not all of which resulted in originated loans endorsed for FHA insurance. Furthermore, we were not able to access the original sampling weights applied to the population of loan applications. However, we do benefit from the fact that we have available the full “population” of FHA at-risk insured loans, which allows us to directly estimate differences in performance among loans in the choice-based samples. We have controlled directly for the differences in loan performance across our two sources of FICO score information by including an indicator for whether the loan was included in the Unicon loan subsample, along with a series of indicator variables that account for the availability and source of FICO scores across different origination years.

Origination Year Indicators

The series of origination year indicators applied in past Reviews to account for changes in FHA underwriting requirements has been modified and extended to account for the periods during which loan-level credit score data were or were not available.

FY 1975-1986 Origination

An indicator for loans originated prior to FY 1986 Q3 is included to account for the period prior to tightening of FHA underwriting requirements.

FY 1986-1992 Origination

An indicator for loans originated between FY 1986 Q3 and FY 1991 Q4 is included to capture the condition that these loans were underwritten with more strict requirements but had no borrower credit history information. This variable also corresponds to the last period prior to the availability of borrower credit score data.

Post-FY 1996 Origination

An indicator for loans originated since FY 1996 Q1 is included to account for a loosening of FHA underwriting requirements. This variable is used in models for streamlined refinance loan products for which borrower credit scores are not available.

Exhibit A-1

Logit Model Explanatory Variables								
Variable Name				Values				Description
Mortgage Age Function								
	FRM30	FRM15		ARM	FRM30_SR	FRM15_SR	ARM_SR	Piece-wise linear age functions for ages up to specified knot points (shown in this table as the number of quarters since origination). Estimated parameters give the slope of the age function for each segment. Functions differ by mortgage product type and termination event as indicated (C-Claim, P-Prepay).
	C, P	C	P	C, P	C, P	C, P	C, P	
age1	2	4	4	2	2	4	2	
age2	6	6	6	4	4	6	4	
age3	8	8	8	8	8	8	8	
age4	10	12	12	12	12	12	12	
age5	12	16	>12	16	> 12	16	16	
age6	14	20		20		20	20	
age7	16	>20		24		24	24	
age8	18			28		> 24	> 24	
age9	36			32				
age10	> 36			40				
age11				> 40				
Relative House Price								
rel_hp_cat_1				0 < X ≤ 50				Relative house price measured as relative percentage of Census median house value during quarter of loan origination. Census median prices for CBSAs and non-metro counties were assigned to FHA loans based on locations and origination dates.
rel_hp_cat_2				50 < X ≤ 75				
rel_hp_cat_3				75 < X ≤ 100				
rel_hp_cat_4				100 < X ≤ 125				
rel_hp_cat_5				125 < X ≤ 150				
rel_hp_cat_6				X > 150				
Loan Size								
loancat_cat_1				0 < X ≤ 60				Relative loan size measured as relative percentage of average size loan originated in the same state in the same year.
loancat_cat_2				60 < X ≤ 90				
loancat_cat_3				90 < X ≤ 110				
loancat_cat_4				110 < X ≤ 140				
loancat_cat_5				X > 140				
(continued on following page)								

Exhibit A-1

Logit Model Explanatory Variables		
Variable Name	Values	Description
Loan Purpose		
refinance_cat_1	Not a refinance loan	Indicates whether the loan purpose was for refinancing
refinance_cat_2	Refinance loan	
Loan-to-Value		
ltvcat_cat_1	$0 < X \leq 80$	Loan-to-value at origination. Missing LTV values for SR product types are replaced by mean LTV by state, origination FY, and corresponding non-SR product types.
ltvcat_cat_2	$80 < X \leq 90$	
ltvcat_cat_3	$90 < X < 95$	
ltvcat_cat_4	$95 \leq X < 97$	
ltvcat_cat_5	$97 \leq X$	
Season		
season_cat_1	$X = 1$	Calendar quarter of mortgage origination.
season_cat_2	$X = 2$	
season_cat_3	$X = 3$	
season_cat_4	$X = 4$	
Probability of Negative Equity		
pneqcat_cat_1	$0.00 \leq X \leq 0.05$	Probability of negative equity. Based on OFHEO house price drift and volatility estimates. MSA-level estimates used for selected MSAs; otherwise, Census Division level estimates were used.
pneqcat_cat_2	$0.05 < X \leq 0.10$	
pneqcat_cat_3	$0.10 < X \leq 0.15$	
pneqcat_cat_4	$0.15 < X \leq 0.20$	
pneqcat_cat_5	$0.20 < X \leq 0.25$	
pneqcat_cat_6	$0.25 < X \leq 0.30$	
pneqcat_cat_7	$X > 0.30$	
Mortgage Premium (Spread)		
spreadcat_cat_1	$X \leq -30$	Mortgage premium value measured as difference between current coupon rate and average FRM market rate, relative to current coupon rate.
spreadcat_cat_2	$-30 < X \leq -20$	
spreadcat_cat_3	$-20 < X \leq -10$	
spreadcat_cat_4	$-10 < X \leq 0$	
spreadcat_cat_5	$0 < X \leq 10$	
spreadcat_cat_6	$10 < X \leq 20$	
spreadcat_cat_7	$20 < X \leq 30$	
spreadcat_cat_8	$X > 30$	
(continued on following page)		

Exhibit A-1

Logit Model Explanatory Variables		
Variable Name	Values	Description
Yield Curve Slope		
yslopecat_cat_1	$0.0 \leq X \leq 1.0$	Yield curve slope measured as ratio of 10-year CMT to 1-year CMT rates.
yslopecat_cat_2	$1.0 < X \leq 1.2$	
yslopecat_cat_3	$1.2 < X \leq 1.5$	
yslopecat_cat_4	$X > 1.5$	
Burnout Factor		
in_moneycat_cat_1	$X \leq 0$	Burnout factor equal to the moving average number of basis points the prepayment option was in the money during those quarters the option was in the money over the preceding 8 quarters.
in_moneycat_cat_2	$0 < X \leq 50$	
in_moneycat_cat_3	$50 < X \leq 100$	
in_moneycat_cat_4	$100 < X \leq 150$	
in_moneycat_cat_5	$150 < X \leq 200$	
in_moneycat_cat_6	$X > 200$	
1975-1986 Origination		
fy_1975_1986_cat_1	$X \geq 1986$	Pre-FY1986 Q3 origination prior to changes in FHA underwriting requirements. Prior to availability of credit score data.
fy_1975_1986_cat_2	$X < 1986$	
1986-1992 Origination		
fy_1986_1992_cat_1	$1986 > X$ or $1992 \leq X$	Post-FY 1986 Q3 and pre-FY 1992 origination. After changes in FHA underwriting requirements. Prior to availability of sample credit score data.
fy_1986_1992_cat_2	$1986 \leq X < 1992$	
Post-1996 Origination		
Fy_1996_XXXX_1	$X \leq 1996$	Post-1996 origination. After changes in FHA underwriting requirements. For SR loan products with no credit score data.
Fy_1996_XXXX_2	$X > 1996$	
(continued on following page)		

Exhibit A-1

Logit Model Explanatory Variables		
Variable Name	Values	Description
Exposure Year/Quarter FRM Rate		
ey_ratecat_cat_1	$X \leq 6$	FRM average mortgage rate during exposure year and quarter. Included to distinguish high-rate and low-rate environments.
ey_ratecat_cat_2	$6 < X \leq 7$	
ey_ratecat_cat_3	$7 < X \leq 8$	
ey_ratecat_cat_4	$8 < X \leq 9$	
ey_ratecat_cat_5	$9 < X \leq 10$	
ey_ratecat_cat_6	$X > 10$	
Metropolitan Unemployment Rates		
uechngcat_1	$X \leq -30$	Percent change over the preceding year in the metro-area unemployment rate.
uechngcat_2	$-30 < X \leq -20$	
uechngcat_3	$-20 < X \leq -10$	
uechngcat_4	$-10 < X \leq 0$	
uechngcat_5	$0 < X \leq 10$	
uechngcat_6	$10 < X \leq 20$	
uechngcat_7	$20 < X \leq 30$	
uechngcat_8	$30 < X \leq 50$	
uechngcat_9	$50 < X \leq 100$	
uechngcat_10	$100 < X \leq 150$	
uechngcat_11	$X > 150$	
ARM Payment Burden		
arm_paymentcat_1	$X \leq 0$	Percent increase in monthly payment since origination.
arm_paymentcat_2	$0 < X \leq 10$	
arm_paymentcat_3	$10 < X \leq 20$	
arm_paymentcat_4	$0 < X \leq 30$	
arm_paymentcat_5	$X > 30$	
Source of Downpayment Assistance		
gift_ltr_src_cat_1	None Recorded	Source of downpayment assistance.
gift_ltr_src_cat_2	Relatives	
gift_ltr_src_cat_3	Non-Profit	
gift_ltr_src_cat_4	Government	
gift_ltr_src_cat_5	Other	
(continued on following page)		

Exhibit A-1

Logit Model Explanatory Variables		
Variable Name	Values	Description
Borrower FICO Score		
fico_300_499	$300 < X \leq 499$	Borrower FICO scores obtained from sample data for FY 1992-2004 originations. Complete data on FHA FICO scores is available from FY 2004.
fico_500_539	$500 < X \leq 539$	
fico_540_579	$540 < X \leq 579$	
fico_580_619	$580 < X \leq 619$	
fico_620_639	$620 < X \leq 639$	
fico_640_659	$640 < X \leq 659$	
fico_660_679	$660 < X \leq 679$	
fico_680_719	$680 < X \leq 719$	
fico_720_850	$720 < X \leq 850$	
Fico_000	No FICO Score Generated after 1992	FICO category 000 represents loans submitted to credit data repository for scoring but not resulting in a credit score.
Fico_999	Missing FICO Score after 1992	FICO category 999 represents loans for which FICO score not available from any source.
unicon_loan	Loan is member of Unicon sample	Source of FICO score is FHA, not the historical sampling of applications.

III. Model Estimation Results

Exhibits A-2 and A-3 present the coefficient estimates for the binomial logit models for conditional claim and prepayment probabilities.

Exhibit A-2							
Results for Conditional Claim Rate Model Estimation							
Variable	FRM 30	FRM 15	ARM	SR FRM 30	SR FRM 15	SR ARM	
rel_hp_cat_2	-0.1312	-0.2288	-0.1593				
rel_hp_cat_3	-0.2136	-0.4081	-0.3745				
rel_hp_cat_4	-0.2368	-0.4656	-0.4169				
rel_hp_cat_5	-0.2276	-0.5024	-0.4487				
rel_hp_cat_6	-0.1884	-0.6299	-0.4116				
ltvcat_cat_2	0.5019	0.9479	0.3859				
ltvcat_cat_3	0.4697	1.0803	0.4482				
ltvcat_cat_4	0.5802	1.1738	0.4556				
ltvcat_cat_5	0.5707	1.1033	0.3965				
loancat_cat_2				0.1777	-0.1640	0.1685	
loancat_cat_3				0.2930	-0.3039	0.2981	
loancat_cat_4				0.3749	-0.2781	0.3276	
loancat_cat_5				0.3252	-0.1038	0.2378	*
refinance_cat_2	-0.0971	-0.4571	-0.0848				
season_cat_2	0.0241	0.0283	0.0358	0.0051	-0.0124	0.0724	*
season_cat_3	0.0095	0.0372	-0.0176	-0.0024	0.0203	0.0126	*
season_cat_4	-0.0179	-0.0384	-0.0308	-0.0277	-0.0511	-0.0509	*
pneqcat_cat_2	0.5393	0.7565	0.4691	0.8505	0.8760	0.7932	
pneqcat_cat_3	0.6566	0.9683	0.6307	1.2232	0.8974	1.2168	
pneqcat_cat_4	0.7903	1.2064	0.7814	1.4280	1.1899	1.3397	
pneqcat_cat_5	0.9150	1.3677	0.9477	1.4846	1.4140	1.4410	
pneqcat_cat_6	1.0579	1.5340	1.0611	1.6772	1.7347	1.9021	
pneqcat_cat_7	1.4861	1.8084	1.5427	2.3628	1.5269	2.5963	
ycslopecat_cat_2	-0.0997	-0.1059	-0.0696	-0.2746	-0.0887	-0.0521	*
ycslopecat_cat_3	-0.0678	-0.0347	-0.1778	-0.0978	0.0756	-0.2511	*
ycslopecat_cat_4	-0.1522	-0.1113	-0.0832	-0.1672	-0.0175	-0.2339	*
spreadcat_cat_2	0.4399	-0.1131	0.1142	-0.4693		0.0492	*
spreadcat_cat_3	0.5949	-0.0411	0.2098	-0.3595		0.1623	
spreadcat_cat_4	0.7832	0.1394	0.1525	-0.0812	*	0.2086	
spreadcat_cat_5	0.8718	0.1755	0.1356	0.1051		0.1991	
spreadcat_cat_6	1.0076	0.2347	0.1653	0.2770		0.2462	
spreadcat_cat_7	1.1278	0.2982	(dropped)	0.3414		1.6879	*
spreadcat_cat_8	1.2586	0.4563	(dropped)	0.3900		1.6879	*
in_moneycat_cat_2	0.0415	0.0974	0.5037	-0.1810	0.3797		
in_moneycat_cat_3	0.2014	0.2571	0.9042	0.1329	0.7673		
in_moneycat_cat_4	0.4384	0.3672	0.9953	0.4316	1.1057		
in_moneycat_cat_5	0.6537	0.5319	0.9953	0.5893	1.2980		
in_moneycat_cat_6	0.8428	0.6488	0.9953	0.7837	1.6995		
gift_ltr_src_cat_2	0.2693	0.1910	0.2844				
gift_ltr_src_cat_3	0.8406	1.6129	0.6704				
gift_ltr_src_cat_4	0.5149	1.3405	0.6457				

Exhibit A-2						
Results for Conditional Claim Rate Model Estimation						
Variable	FRM 30	FRM 15	ARM	SR FRM 30	SR FRM 15	SR ARM
gift_ltr_src_cat_5	0.4478	0.2920 *	0.4084 *			
ey_ratecat_cat_2			-0.1331			-0.2461
ey_ratecat_cat_3			-0.2769			-0.5598
ey_ratecat_cat_4			-0.4012			-0.6697
ey_ratecat_cat_5			-0.1625			-0.7780
ey_ratecat_cat_6			-0.1625			-0.7780
fy_1975_1985_cat_2	0.5454	0.3606				
fy_1986_1991_cat_2	-0.0900	-0.3600	-0.0293 *			
fy_1996_XXXX_cat_2	0.3561	0.2372	0.5772	0.6593	-0.1617	0.9959
age1	2.8549	1.1926	2.0770	2.0695	1.1782	1.5167
age2	0.5975	0.4979	1.4841	1.0475	0.3412	1.1004
age3	0.1468	0.2485	0.3817	0.2508	0.2935	0.3125
age4	0.0933	0.0839	0.1320	0.0859	0.0762	0.0891
age5	0.0464	0.0051 *	0.0392	-0.0229	0.0358	0.0655
age6	0.0058 *	-0.0254	0.0057 *		-0.0360 *	-0.0057 *
age7	0.0102 *	-0.0595	-0.0131		-0.0534	0.0182 *
age8	-0.0027 *		-0.0353		-0.0921	-0.0726
age9	-0.0261		-0.0306			
age10	-0.0455		-0.0433			
age11			-0.0459			
fico_000	0.2192	0.0711 *	0.0474			
fico_999	-0.4481	-0.8527	-0.2081			
fico_300_499	0.8987	1.4814	0.9395			
fico_500_539	0.5836	1.1189	0.6540			
fico_540_579	0.3568	0.9125	0.3766			
fico_580_619	0.1759	0.4752	0.1502			
fico_640_659	-0.2758	-0.0300 *	-0.2277			
fico_660_679	-0.5271	-0.3546	-0.4823			
fico_680_719	-0.9103	-0.7025	-0.8911			
fico_720_850	-1.4925	-1.3909	-1.5288			
unicon_loan	0.6750	-0.0295 *	0.9083			
cons	-15.7587	-13.4219	-15.2222	-14.3098	-13.8951	-13.1402
Statistics	FRM 30	FRM 15	ARM	SR FRM 30	SR FRM 15	SR ARM
Log likelihood	-700080	-63872	-211081	-167317	-15374	-36367
Number of obs	33065069	7405294	9053698	11457026	4764117	1583517
LR χ^2	7113151	653459	.	2442662	254332	.
Prob > χ^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

* Not significant for 0.05-level asymptotic normal test

Exhibit A-3						
Results for Conditional Prepayment Rate Model Estimation						
Variable	FRM 30	FRM 15	ARM	SR FRM 30	SR FRM 15	SR ARM
rel_hp_cat_2	0.1061	0.1519	0.0922			
rel_hp_cat_3	0.2345	0.2637	0.1642			
rel_hp_cat_4	0.3274	0.3392	0.1813			
rel_hp_cat_5	0.3613	0.3642	0.1791			
rel_hp_cat_6	0.3045	0.3676	0.1573			
ltvcat_cat_2	-0.0093	0.0129	0.0097	*		
ltvcat_cat_3	0.0666	0.0428	0.0940			
ltvcat_cat_4	0.1305	0.0829	0.1954			
ltvcat_cat_5	0.1348	0.0806	0.1664			
loancat_cat_2				0.3149	0.0907	0.2764
loancat_cat_3				0.5276	0.1641	0.4387
loancat_cat_4				0.6634	0.2426	0.5448
loancat_cat_5				0.7874	0.3924	0.6607
refinance_cat_2	0.1573	0.1219	0.1427			
season_cat_2	0.1991	0.2145	0.2046	0.1969	0.1681	0.1930
season_cat_3	0.1051	0.1116	0.0549	0.1190	0.1024	0.1250
season_cat_4	0.0785	0.0723	0.0319	0.0770	0.0414	0.0167
pneqcat_cat_2	-0.3311	-0.4060	-0.3480	-0.3736	-0.3031	-0.4115
pneqcat_cat_3	-0.4165	-0.5795	-0.5228	-0.4334	-0.5997	-0.5576
pneqcat_cat_4	-0.5145	-0.7755	-0.6598	-0.5298	-0.7484	-0.6559
pneqcat_cat_5	-0.6172	-0.9117	-0.7580	-0.7836	-0.7943	-0.9233
pneqcat_cat_6	-0.7294	-0.9720	-0.9669	-0.9996	-0.8436	-1.1661
pneqcat_cat_7	-0.8333	-1.1597	-1.2914	-1.2871	-0.9699	-1.5515
yslopecat_cat_2	0.1081	-0.0357	-0.0075	*	0.0617	0.2232
yslopecat_cat_3	0.3014	0.1301	0.2032		0.1451	0.2304
yslopecat_cat_4	0.6145	0.3879	-0.0443		0.6264	0.2602
spreadcat_cat_2	0.5908	0.0351	*	0.1092	-0.8113	0.1319
spreadcat_cat_3	0.4555	0.2713		0.1889	-0.6322	0.2563
spreadcat_cat_4	0.4620	0.4297		0.3107	-0.4528	0.4129
spreadcat_cat_5	0.6568	0.6388		0.4816	-0.2266	0.5557
spreadcat_cat_6	1.2235	0.9676		0.7146	0.2968	0.5974
spreadcat_cat_7	1.5819	1.2096		1.9183	0.5743	1.4362
spreadcat_cat_8	1.4540	1.1852		2.4016	0.5790	1.4362
in_moneycat_cat_2	0.2817	0.2083		0.3179	0.3359	0.3989
in_moneycat_cat_3	0.5673	0.3642		0.4354	0.5688	0.6076
in_moneycat_cat_4	0.5961	0.3243	(dropped)		0.5836	0.6598
in_moneycat_cat_5	0.5215	0.2419	(dropped)		0.5047	0.6140
in_moneycat_cat_6	0.4169	0.1253	(dropped)		0.4288	0.5275
gift_ltr_src_cat_2	0.0608	0.0178	*	0.0134	*	
gift_ltr_src_cat_3	0.0867	0.5970		-0.1124		
gift_ltr_src_cat_4	-0.1981	0.0110	*	-0.1733		
gift_ltr_src_cat_5	0.1076	-0.1890	*	0.1384		
ey_ratecat_cat_2				-0.0757		-0.0049
ey_ratecat_cat_3				-0.4877		-0.2656
ey_ratecat_cat_4				-0.8754		-0.5129
ey_ratecat_cat_5				-1.3932		-0.8200

Exhibit A-3						
Results for Conditional Prepayment Rate Model Estimation						
Variable	FRM 30	FRM 15	ARM	SR FRM 30	SR FRM 15	SR ARM
ey_ratecat_cat_6			-1.7798			-0.8200
fy_1975_1985_cat_2	-0.1678	-0.0311				
fy_1986_1991_cat_2	-0.2439	-0.1593	0.2168			
fy_1996_XXXX_cat_2	0.2070	0.2620	0.3358	0.5116	0.2162	0.6358
age1	1.1492	0.5211	1.2801	1.0237	0.4846	1.2137
age2	0.2015	0.1391	0.5146	0.0827	0.0177	0.2367
age3	0.0314	0.0709	0.0755	0.0015	*	0.0525
age4	0.0018	* -0.0039	* -0.0214	-0.0291		0.0027
age5	0.0084	0.0061	-0.0397	-0.0099		0.0215
age6	0.0005	*	-0.0479			0.0483
age7	-0.0102		-0.0015	*		-0.0563
age8	-0.0534		-0.0092			0.0185
age9	-0.0091		0.0033	*		
age10	-0.0063		-0.0049			
age11			-0.0147			
fico_000	-0.0841	0.0340	* 0.0176	*		
fico_999	0.0898	0.1034	0.4292			
fico_300_499	-0.2558	0.2253	-0.4221			
fico_500_539	-0.1570	0.1976	-0.2759			
fico_540_579	-0.0316	0.2598	-0.1623			
fico_580_619	0.0474	0.2329	-0.0815			
fico_640_659	0.1564	0.1442	0.0712			
fico_660_679	0.1847	0.1280	0.1070			
fico_680_719	0.2453	0.0993	0.1872			
fico_720_850	0.2875	0.0669	0.2464			
unicon_loan	-0.1088	-0.0454	0.2912			
cons	-8.2567	-7.4018	-7.3047	-6.4039	-6.4585	-6.4145
Statistics	FRM 30	FRM 15	ARM	SR FRM 30	SR FRM 15	SR ARM
Log likelihood	-700080	-63872	-211081	-167317	-15374	-36367
Number of obs	33065069	7405294	9053698	11457026	4764117	1583517
LR χ^2	7113151	653459	.	2442662	254332	.
Prob > χ^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

* Not significant for 0.05-level asymptotic normal test

References

Begg, C.B. and R. Gray, "Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions," *Biometrika*, 71(1):11-18, 1984.

Calhoun, C.A. and Y. Deng, "A Dynamic Analysis of Fixed- and Adjustable-Rate Mortgage Terminations," *Journal of Real Estate Finance and Economics*, 24(1/2):9-33, 2002.

Calhoun, C.A., "OFHEO House Price Indexes: Technical Description," Washington, D.C., Office of Federal Housing Enterprise Oversight, April 1996.

Case, K.E. and Shiller, R.J., "Prices of Single Family Real Estate Prices," *New England Economic Review*, 45-56, 1987.

Case, K.E. and Shiller, R.J., "The Efficiency of the Market for Single-Family Homes," *The American Economic Review*, 79:125-137, 1989.

Costlett, S., "Efficient Estimation of Discrete Choice Models," pp. 51-111 in C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, 1981.

Deng, Y., J. M. Quigley and R. Van Order, "Mortgage Termination, Heterogeneity, and the Exercise of Mortgage Options," *Econometrica*, 68(2):275-307, 2000.