# 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 2007 Review. 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 specificed and estimated for the FY 2007 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 used to account for differences in the number of identical 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 apply 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 worked 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 flexible enough to capture the age patterns of conditional claims and prepayments observed in the FHA data. The approach we adopted was a series of 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 while limiting the overall number of model parameters that have to be estimated.

As indicated, 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:

$$\pi_C(t) = \frac{e^{\alpha_C + X_C(t)\beta_C}}{1 + e^{\alpha_C + X_C(t)\beta_C} + e^{\alpha_P + X_P(t)\beta_P}} \tag{1}$$

$$\pi_P(t) = \frac{e^{\alpha_P + X_P(t)\beta_P}}{1 + e^{\alpha_C + X_C(t)\beta_C} + e^{\alpha_P + X_P(t)\beta_P}} \tag{2}$$

$$\pi_A(t) = \frac{1}{1 + e^{\alpha_C + X_C(t)\beta_C} + e^{\alpha_P + X_P(t)\beta_P}} \tag{3}$$

where the constant terms $\alpha_C$ and $\alpha_P$ and the coefficient vectors $\beta_C$ and $\beta_P$ are the unknown parameters to be estimated. $X_C(t)$ is 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 variables of $X_C(t)$ and $X_P(t)$ are constant over the life of the loan and are not functions of $t$.

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

Since loans in delinquency status may prepay if there is sufficient equity in the home, but not prepay if there is not, we applied the Begg-Gray method after sufficiently separating delinquencies into those that go to claim and those that do not. Because prepayments are unlikely to occur for defaulting loans on their way to becoming claim terminations, censoring of prepayments actually occurs prior to the observed claim termination date. Failure to account for this particular form of censoring could result in biased estimates of the parameters of the prepayment model.

The claim-rate model is best viewed as a reduced-form of a more complicated structural model with two components: (1) an option-based model of borrower payment behavior that determines the incidence and timing of default events that ultimately lead to FHA claims and (2) a model for differences in the waiting time from borrower default until the claim is submitted to FHA. The second component can be properly addressed in conjunction with estimates of loss severity (or loss-given-default), and can vary significantly with differences in state laws on mortgage foreclosure procedures, differences in lender loss-mitigation policies, and with current economic conditions that affect the value and time-to-sale of collateral properties.

For projections in the FY 2007 Review, we apply average loss severity rates observed during FY 2006 stratified by six mortgage product types, whether borrowers received downpayment assistance from non-profit organizations, and whether the state imposes a judicial foreclosure. 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. The Begg-Gray method of estimating separate binomial logit models is particularly advantageous in dealing with this requirement. 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 delinquency episodes leading to claim terminations to create a default-censoring indicator that was applied when estimating the prepayment-rate model. The loan was censored—i.e., removed—upon the onset of a delinquency that lead to a claim without any intervening correction to a current-pay status.

A separate claim-rate model was estimated that accounted for the censoring of potential claim terminations by observed prepayments. Here, there is no prior indicator as there is for claims. The two sets of parameter estimates were recombined mathematically to produce the final multinomial model for prepayment and claim probabilities. The Begg-Gray methodology produces parameter estimates that are equivalent to those of the multinomial logit model. Failure to exclude defaulting loans from the sample of loans assumed to be at risk of prepayment would result in a downward bias in the estimates of the conditional probabilities of prepayment because loans with a zero chance of prepayment would be included in the sample in estimating conditional prepayment rates.

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

- Data on the start of a delinquency 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).
- Data of an observed prepayment was used to define a prepayment-censoring indicator for claim.
- A binomial logit model for conditional claim probabilities was estimated using the prepayment-censoring indicator 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 (according to the equations below) 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 $\alpha_C, \beta_C, \alpha_P$, and $\beta_P$ 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:

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

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

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

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

$$\pi^C_{MNL} = \frac{\dfrac{\pi^C_{BNL}}{(1-\pi^C_{BNL})}}{1+\dfrac{\pi^C_{BNL}}{(1-\pi^C_{BNL})}+\dfrac{\pi^P_{BNL}}{(1-\pi^P_{BNL})}}, \quad \pi^P_{MNL} = \frac{\dfrac{\pi^P_{BNL}}{(1-\pi^P_{BNL})}}{1+\dfrac{\pi^C_{BNL}}{(1-\pi^C_{BNL})}+\dfrac{\pi^P_{BNL}}{(1-\pi^P_{BNL})}}. \tag{6}$$

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

$$\pi^C_{MNL} = \frac{\pi^C_{BNL}\cdot(1-\pi^P_{BNL})}{(1-\pi^C_{BNL}\cdot\pi^P_{BNL})}, \quad \pi^P_{MNL} = \frac{\pi^P_{BNL}\cdot(1-\pi^C_{BNL})}{(1-\pi^C_{BNL}\cdot\pi^P_{BNL})}. \tag{7}$$

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 data set used in this Actuarial Review is based on an extract from FHA's database as of March 31, 2007. The data set was first filtered for loans with missing or abnormal values of key variables in our econometric model. In addition, lender information was not used in our econometric model and loans with missing lender/servicer information were also excluded from our analysis. 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.

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 first quarter of FY 2007 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 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 2007 analysis. This produced a starting sample of approximately 23 million single-family loans originated between FY 1975 and the first quarter of FY 2007. 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 scheduled age of maturity of the loan).

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. Loan-level information on borrower FICO scores was introduced for the FY 2007 Review and some historical information on borrower FICO scores was obtained through a choice-based sampling scheme that impacted the sampling of FRM30 loans. Further consideration of this issue is provided in the discussion of loan-level borrower credit score data in the following section.

## II. Explanatory Variables

Four main categories of explanatory variables were developed:

1. Fixed initial loan characteristics, such as mortgage product type, amortization term, origination year and quarter, original loan-to-value (LTV) ratio, original loan amount, original mortgage interest rate, and geographic location (MSA, state, Census division);

2. Fixed initial borrower characteristics, such as 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, such as 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, such as 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 home purchase mortgages.
2. FRM15 Fixed-rate 15-year home purchase mortgages.
3. ARM Adjustable-rate home purchase 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:

$$\text{age1} = \begin{Bmatrix} \text{AGE} & \text{if AGE} \le k_1 \\ k_1 & \text{if AGE} > k_1 \end{Bmatrix}$$

$$\text{age2} = \begin{Bmatrix} 0 & \text{if AGE} \le k_1 \\ \text{AGE - } k_1 & \text{if } k_1 < \text{AGE} \le k_2 \\ k_2 - k_1 & \text{if AGE} > k_2 \end{Bmatrix}$$

$$\text{age3} = \begin{Bmatrix} 0 & \text{if AGE} \le k_2 \\ \text{AGE - } k_2 & \text{if } k_2 < \text{AGE} \le k_3 \\ k_3 - k_2 & \text{if AGE} > k_3 \end{Bmatrix}$$

$$\text{age4} = \begin{Bmatrix} 0 & \text{if AGE} \le k_3 \\ \text{AGE - } k_3 & \text{if } k_3 < \text{AGE} \le k_4 \\ k_4 - k_3 & \text{if AGE} > k_4 \end{Bmatrix}$$

$$\text{age5} = \begin{Bmatrix} 0 & \text{if AGE} \le k_4 \\ \text{AGE - } k_4 & \text{if AGE} > k_4 \end{Bmatrix} \tag{8}$$

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.

*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. The LTV ratio variable may exceed 100 percent due to FHA's practice of allowing the financing of some closing costs, so a categorical outcome is included for this possibility. 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:

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

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:

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

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 are given 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:

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

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:

$$\begin{aligned} C(t) = max\{\, & min[\, Index(t-S) + Margin, \\ & C(t-1) + A(t)\cdot Period\_UpCap, C(0) + Life\_UpCap\,], \\ & C(t-1) - A(t)\cdot Period\_DownCap(t), max(C(0) - Life\_DownCap, Life\_Min)\} \end{aligned} \quad (13)$$

where $Index(t)$ is the underlying rate index value at time *t*, *S* is the "lookback" period, and *Margin* is the amount added to $Index(t-S)$ 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 2007 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*

As described in the FY 2006 Review, we previously undertook to develop 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 times 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 considered for the FY 2006 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 FY2007 Review.

*Source of Downpayment Assistance*

As documented in the FY 2006 Review, 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 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 have been included in the models estimated for the FY 2007 Review. FHA has relatively complete data on borrower FICO scores for loans originated since May 2004. In addition, FHA has 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 can be used for estimation.

Historical FICO score data was collected for FHA cases with application dates during FYs 1992, 1994, and 1996. FICO scores of the borrower and up to two 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 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 and FY 2003 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 9 categorical outcomes for scores in the following range of values: 300-499, 500-539, 540-579, 580-619, 620-639, 640-659, 660-679, 680-719, and 720-850. Dummy variables for these categories were included for estimation.

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 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 FY 1992 or later for which no attempt was made to obtain the FICO score.

Finally, an indicator for FHA FICO score data was defined to distinguish loans with FICO scores obtained through the normal FHA loan approval process from loans that have FICO scores from the retrospective historical sampling procedure. There are some months in FY 2004 for which both types of FICO scores are present in the data. This variable was included to detect any statistical differences in the performance of loans with different sources of FICO score data.

*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 sampling scheme was applied to assure adequate representation of loans with historical FICO score data. For each fiscal year loans with historical FICO scores were flagged and the total counts of loans with and without FICO scores were determined. Separate sampling rates for loans with and without FICO scores were derived to give as close to equal representation 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 loans with FICO scores, with the remainder of the 20 percent sample comprising FHA loans without FICO scores. In other years, this resulted in selecting a random subsample of loans with FICO scores and an equally-sized random sample of FHA loans without FICO scores. As described further below, our goal was to attain a mix of loans with and without FICO scores (for those years in which FICO scores were available) in order to analyze both 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.

Thus, observations used for estimation included a mix of randomly sampled FHA loan originations without FICO scores and a choice-based sample of loans with FICO scores. Estimation using choice-based samples can 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. Standard corrections for the bias in the intercept depend 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 such as comprise our population of loans. Furthermore, we are not able to access the original sampling weights applied to the population of loan applications.

As an alternative, we control directly for the differences in our two data sources by including a number of indicator variables that account for origination years during which historical FICO score data was available (albeit from choice-based sampling) versus FICO score data obtained directly for all loans from FHA. By estimating the conditional claim and prepay models across a larger sample that includes: (1) FHA loans prior to the years FICO score data were available from any source; (2) FHA loans during years when historical FICO scores were obtained for some loans through a choice-based sampling process; and (3) FHA loans for years during which FICO scores are available for all FHA loans, we are able to identify the direct impact of the choice-based samples on the intercept of the model, while obtaining unbiased estimates of other coefficients, including the FICO score categories.

*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 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 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**

<table>
<tr><th colspan="8">Logit Model Explanatory Variables</th></tr>
<tr><th colspan="4">Variable Name</th><th colspan="3">Values</th><th>Description</th></tr>
<tr><td colspan="8">Mortgage Age Function</td></tr>
<tr><td></td><td>FRM30</td><td>FRM15</td><td>ARM</td><td>FRM30_SR</td><td>FRM15_SR</td><td>ARM_SR</td><td rowspan="14">Piece-wise linear age functions for ages up to specified knot points (shown in this table as the number of quarters since origination).<br><br>Estimated parameters give the slope of the age function for each segment.<br><br>Functions differ by mortgage product type as indicated.</td></tr>
<tr><td>age1</td><td>2</td><td>4</td><td>2</td><td>2</td><td>4</td><td>2</td></tr>
<tr><td>age2</td><td>4</td><td>6</td><td>4</td><td>4</td><td>6</td><td>4</td></tr>
<tr><td>age3</td><td>8</td><td>8</td><td>8</td><td>8</td><td>8</td><td>8</td></tr>
<tr><td>age4</td><td>12</td><td>12</td><td>12</td><td>12</td><td>12</td><td>12</td></tr>
<tr><td>age5</td><td>16</td><td>16</td><td>16</td><td>> 12</td><td>16</td><td>16</td></tr>
<tr><td>age6</td><td>20</td><td>> 16</td><td>20</td><td></td><td>20</td><td>20</td></tr>
<tr><td>age7</td><td>24</td><td></td><td>24</td><td></td><td>24</td><td>24</td></tr>
<tr><td>age8</td><td>28</td><td></td><td>28</td><td></td><td>> 24</td><td>> 24</td></tr>
<tr><td>age9</td><td>32</td><td></td><td>32</td><td></td><td></td><td></td></tr>
<tr><td>age10</td><td>40</td><td></td><td>40</td><td></td><td></td><td></td></tr>
<tr><td>age11</td><td>60</td><td></td><td>> 40</td><td></td><td></td><td></td></tr>
<tr><td>age12</td><td>80</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>age13</td><td>> 80</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="8"></td></tr>
<tr><td colspan="8">Loan Size</td></tr>
<tr><td colspan="4">loancat_cat_1</td><td colspan="3">0 < X ≤ 60</td><td rowspan="5">Relative loan size measured as relative percentage of average size loan originated in same state in the same year.</td></tr>
<tr><td colspan="4">loancat_cat_2</td><td colspan="3">60 < X ≤ 90</td></tr>
<tr><td colspan="4">loancat_cat_3</td><td colspan="3">90 < X ≤ 110</td></tr>
<tr><td colspan="4">loancat_cat_4</td><td colspan="3">110 < X ≤ 140</td></tr>
<tr><td colspan="4">loancat_cat_5</td><td colspan="3">X > 140</td></tr>
<tr><td colspan="8"></td></tr>
<tr><td colspan="8">Loan-to-Value</td></tr>
<tr><td colspan="4">ltvcat_cat_1</td><td colspan="3">0 < X ≤ 80</td><td rowspan="5">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.</td></tr>
<tr><td colspan="4">ltvcat_cat_2</td><td colspan="3">80 < X ≤ 90</td></tr>
<tr><td colspan="4">ltvcat_cat_3</td><td colspan="3">90 < X < 95</td></tr>
<tr><td colspan="4">ltvcat_cat_4</td><td colspan="3">95 ≤ X < 97</td></tr>
<tr><td colspan="4">ltvcat_cat_5</td><td colspan="3">97 ≤ X</td></tr>
<tr><td colspan="8"></td></tr>
</table>

(continued on following page)

## Exhibit A-1

| Logit Model Explanatory Variables | | |
|---|---|---|
| **Variable Name** | **Values** | **Description** |
| **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 ≤ X ≤ 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 ≤ 0.10 | |
| pneqcat_cat_3 | 0.10 < X ≤ 0.15 | |
| pneqcat_cat_4 | 0.15 < X ≤ 0.20 | |
| pneqcat_cat_5 | 0.20 < X ≤ 0.25 | |
| pneqcat_cat_6 | 0.25 < X ≤ 0.30 | |
| pneqcat_cat_7 | X > 0.30 | |
| | | |
| **Mortgage Premium (Spread)** | | |
| spreadcat_cat_1 | X ≤ -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 ≤ -20 | |
| spreadcat_cat_3 | -20 < X ≤ -10 | |
| spreadcat_cat_4 | -10 < X ≤ 0 | |
| spreadcat_cat_5 | 0 < X ≤ 10 | |
| spreadcat_cat_6 | 10 < X ≤ 20 | |
| spreadcat_cat_7 | 20 < X ≤ 30 | |
| spreadcat_cat_8 | X > 30 | |
| | | |
| **Yield Curve Slope** | | |
| ycslopecat_cat_1 | 0.0 ≤ X ≤ 1.0 | Yield curve slope measured as ratio of 10-year CMT to 1-year CMT rates. |
| ycslopecat_cat_2 | 1.0 < X ≤ 1.2 | |
| ycslopecat_cat_3 | 1.2 < X ≤ 1.5 | |
| ycslopecat_cat_4 | X > 1.5 | |
| | | |

## Exhibit A-1

| Logit Model Explanatory Variables | | |
|---|---|---|
| **Variable Name** | **Values** | **Description** |
| **Burnout Factor** | | |
| | | 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 during the preceding 8 quarters. |
| in_moneycat_cat_1 | $X \le 0$ | |
| in_moneycat_cat_2 | $0 < X \le 50$ | |
| in_moneycat_cat_3 | $50 < X \le 100$ | |
| in_moneycat_cat_4 | $100 < X \le 150$ | |
| in_moneycat_cat_5 | $150 < X \le 200$ | |
| in_moneycat_cat_6 | $X > 200$ | |
| | | |
| **1975-1986 Origination** | | |
| fy_1975_1986_cat_1 | $X \ge 1986$ | 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 $1992 \le X$ | 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 | $1986 \le X < 1992$ | |
| | | |
| **Post-1996 Origination** | | |
| Fy_1996_XXXX_1 | $X \le 1996$ | Post-1996 origination. After change in FHA underwriting requirements. For SR loan products with no credit score data. |
| Fy_1996_XXXX_2 | $X > 1996$ | |
| | | |
| **Exposure Year/Quarter FRM Rate** | | |
| ey_ratecat_cat_1 | $X \le 6$ | FRM average mortgage rate during exposure year and quarter. Included to distinguish high-rate and low-rate environments. |
| ey_ratecat_cat_2 | $6 < X \le 7$ | |
| ey_ratecat_cat_3 | $7 < X \le 8$ | |
| ey_ratecat_cat_4 | $8 < X \le 9$ | |

**Exhibit A-1**

| Logit Model Explanatory Variables | | |
|---|---|---|
| **Variable Name** | **Values** | **Description** |
| ey_ratecat_cat_5 | 9 < X ≤ 10 | |
| ey_ratecat_cat_6 | X > 10 | |
| | | |
| **Metropolitan Unemployment Rates** | | |
| uechngcat_1 | X ≤ -30 | Percent change over the preceding year in the metro-area unemployment rate. |
| uechngcat_2 | -30 < X ≤ -20 | |
| uechngcat_3 | -20 < X ≤ -10 | |
| uechngcat_4 | -10 < X ≤ 0 | |
| uechngcat_5 | 0 < X ≤ 10 | |
| uechngcat_6 | 10 < X ≤ 20 | |
| uechngcat_7 | 20 < X ≤ 30 | |
| uechngcat_8 | 30 < X ≤ 50 | |
| uechngcat_9 | 50 < X ≤ 100 | |
| uechngcat_10 | 100 < X ≤ 150 | |
| uechngcat_11 | X > 150 | |
| | | |
| **ARM Payment Burden** | | |
| arm_paymentcat_1 | X ≤ 0 | Percent increase in monthly payment since origination. |
| arm_paymentcat_2 | 0 < X ≤ 10 | |
| arm_paymentcat_3 | 10 < X ≤ 20 | |
| arm_paymentcat_4 | 0 < X ≤ 30 | |
| arm_paymentcat_5 | X > 30 | |
| | | |
| **Source of Down Payment Assistance** | | |
| gift_ltr_src_cat_1 | None Recorded | Source of down payment 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 | |
| | | |

## Exhibit A-1

| Logit Model Explanatory Variables | | |
|---|---|---|
| **Variable Name** | **Values** | **Description** |
| **Borrower FICO Score** | | |
| fico_300_499 | 300 < X ≤ 499 | Borrower FICO scores obtained from sample data for FY 1992-2004 originations. Complete data on FHA FICO scores is available from FY 2004.<br><br>FICO category 000 represents loans submitted to credit data repository for scoring but not resulting in a credit score.<br><br>FICO category 999 represents loans for which FICO score not available from any source. |
| fico_500_539 | 500 < X ≤ 539 | |
| fico_540_579 | 540 < X ≤ 579 | |
| fico_580_619 | 580 < X ≤ 619 | |
| fico_620_639 | 620 < X ≤ 639 | |
| fico_640_659 | 640 < X ≤ 659 | |
| fico_660_679 | 660 < X ≤ 679 | |
| fico_680_719 | 680 < X ≤ 719 | |
| fico_720_850 | 720 < X ≤ 850 | |
| fico_000 | No FICO Score Generated | |
| fico_999 | Missing FICO Score | |
| fha_fico | FICO Score Source is FHA | 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 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| loancat_cat_2 | -0.0465 | | -0.2376 | | -0.1422 | 0.1868 | | -0.1409 | | 0.1412 | |
| loancat_cat_3 | -0.1389 | | -0.3762 | | -0.2509 | 0.2988 | | -0.2986 | | 0.2800 | |
| loancat_cat_4 | -0.2049 | | -0.5789 | | -0.3223 | 0.3868 | | -0.2504 | | 0.2732 | |
| loancat_cat_5 | -0.2396 | | -0.6505 | | -0.2249 | 0.3870 | | -0.0778 | * | 0.2275 | |
| ltvcat_cat_2 | 0.5592 | | 1.0884 | | 0.5027 | | | | | | |
| ltvcat_cat_3 | 0.5368 | | 1.3174 | | 0.6441 | | | | | | |
| ltvcat_cat_4 | 0.6471 | | 1.4579 | | 0.6793 | | | | | | |
| ltvcat_cat_5 | 0.5857 | | 1.3918 | | 0.6102 | | | | | | |
| season_cat_2 | 0.0342 | | 0.0217 | * | 0.0411 | 0.0168 | * | 0.0031 | * | 0.1233 | |
| season_cat_3 | 0.0049 | * | 0.0278 | * | -0.0405 | -0.0022 | * | 0.0103 | * | -0.0149 | * |
| season_cat_4 | -0.0120 | * | -0.0312 | * | -0.0366 | -0.0328 | * | -0.0349 | * | -0.0342 | * |
| pneqcat_cat_2 | 0.4345 | | 0.6454 | | 0.2484 | 0.6800 | | 0.7825 | | 0.6929 | |
| pneqcat_cat_3 | 0.5522 | | 0.8670 | | 0.4185 | 1.0606 | | 0.8265 | | 1.0005 | |
| pneqcat_cat_4 | 0.6820 | | 1.0050 | | 0.6143 | 1.2306 | | 0.8938 | | 1.1781 | |
| pneqcat_cat_5 | 0.8262 | | 1.2077 | | 0.8131 | 1.2941 | | 1.1599 | | 1.5014 | |
| pneqcat_cat_6 | 0.9897 | | 1.2981 | | 0.9992 | 1.5921 | | 1.7773 | | 2.0655 | |
| pneqcat_cat_7 | 1.4119 | | 1.6496 | | 1.5774 | 2.2881 | | 1.5383 | | 2.7358 | |
| ycslopecat_cat_2 | -0.0993 | | -0.0818 | | -0.0380 | -0.3448 | | -0.1814 | | -0.2200 | |
| ycslopecat_cat_3 | -0.0327 | | -0.0087 | * | -0.2051 | -0.1622 | | 0.0009 | * | -0.4300 | |
| ycslopecat_cat_4 | -0.1715 | | -0.1042 | | -0.1226 | -0.2468 | | -0.0983 | * | -0.4510 | |
| spreadcat_cat_2 | | | -0.1797 | * | 0.0649 | -0.4535 | | | | 0.0084 | * |
| spreadcat_cat_3 | | | -0.1423 | * | 0.1615 | -0.2742 | | | | 0.1254 | |
| spreadcat_cat_4 | 0.5043 | | 0.0586 | * | 0.1310 | 0.0495 | * | | | 0.2032 | |
| spreadcat_cat_5 | 0.5597 | | 0.1090 | * | 0.1518 | 0.2339 | | | | 0.2307 | |
| spreadcat_cat_6 | 0.6691 | | 0.1776 | | 0.1292 | 0.4173 | | | | 0.3194 | |
| spreadcat_cat_7 | 0.7986 | | 0.2510 | | | 0.4717 | | | | 1.7517 | * |
| spreadcat_cat_8 | 0.9727 | | 0.4303 | | | 0.5045 | | | | 1.7517 | * |
| inmoneycat_cat_2 | 0.1694 | | 0.0886 | | 0.4451 | -0.1410 | | 0.3968 | | | |
| inmoneycat_cat_3 | 0.3274 | | 0.2486 | | 0.7022 | 0.1819 | | 0.8034 | | | |
| inmoneycat_cat_4 | 0.5587 | | 0.3611 | | 0.9066 | 0.4936 | | 1.1219 | | | |
| inmoneycat_cat_5 | 0.7741 | | 0.5147 | | 0.9066 | 0.6644 | | 1.3360 | | | |
| inmoneycat_cat_6 | 0.9656 | | 0.6386 | | 0.9066 | 0.8555 | | 1.7117 | | | |
| gift_ltr_src_cat_2 | 0.3202 | | 0.2794 | | 0.3287 | | | | | | |
| gift_ltr_src_cat_3 | 0.9594 | | 1.6441 | | 0.8349 | | | | | | |
| gift_ltr_src_cat_4 | 0.5397 | | 1.2638 | | 0.6918 | | | | | | |
| ey_ratecat_cat_2 | | | | | -0.0782 | | | | | -0.2686 | |
| ey_ratecat_cat_3 | | | | | -0.2315 | | | | | -0.6937 | |
| ey_ratecat_cat_4 | | | | | -0.3119 | | | | | -0.7904 | |
| ey_ratecat_cat_5 | | | | | -0.1195 | | | | | -0.9540 | |
| ey_ratecat_cat_6 | | | | | -0.1195 | | | | | -0.9540 | |
| fy_1975_1986_cat_2 | 0.3610 | | 0.4192 | | | | | | | | |

**Exhibit A-2**

**Results for Conditional Claim Rate Model Estimation**

| Variable | FRM 30 | FRM 15 | ARM | SR FRM 30 | SR FRM 15 | SR ARM |
|---|---|---|---|---|---|---|
| fy_1986_1992_cat_2 | -0.3057 | -0.3895 | -0.8423 | | | |
| fy_1996_XXXX_cat_2 | 0.5603 | 0.3439 | 0.7755 | 0.7043 | -0.1513 | 1.1196 |
| age1 | 2.0953 | 1.1410 | 2.1363 | 2.1792 | 1.2358 | 1.6093 |
| age2 | 1.1556 | 0.4887 | 1.4626 | 1.0500 | 0.3050 | 1.0757 |
| age3 | 0.2793 | 0.2677 | 0.3874 | 0.2482 | 0.2952 | 0.3109 |
| age4 | 0.0546 | 0.0742 | 0.1353 | 0.0769 | 0.0743 | 0.0825 |
| age5 | 0.0061 * | 0.0226 | 0.0356 | -0.0252 | 0.0475 | 0.0682 |
| age6 | -0.0178 | -0.0583 | 0.0055 * | | -0.0510 | -0.0021 * |
| age7 | -0.0345 | | -0.0196 | | -0.0523 | 0.0135 * |
| age8 | -0.0448 | | -0.0319 | | -0.0942 | -0.0774 |
| age9 | -0.0290 | | -0.0370 | | | |
| age10 | -0.0151 | | -0.0364 | | | |
| age11 | -0.0417 | | -0.0439 | | | |
| age12 | -0.0659 | | | | | |
| age13 | -0.0586 | | | | | |
| fico_000 | 0.5566 | 0.0433 * | 0.0729 | | | |
| fico_999 | -0.8458 | -0.9512 | -1.1195 | | | |
| fico_300_499 | 1.2376 | 1.2141 | 0.9801 | | | |
| fico_500_539 | 0.9128 | 0.8840 | 0.6957 | | | |
| fico_540_579 | 0.6847 | 0.5424 | 0.4061 | | | |
| fico_580_619 | 0.5089 | 0.2357 | 0.1828 | | | |
| fico_640_659 | 0.1092 | -0.2146 | -0.1960 | | | |
| fico_660_679 | -0.1384 | -0.5388 | -0.4352 | | | |
| fico_680_719 | -0.5107 | -0.8965 | -0.8556 | | | |
| fico_720_850 | -1.1310 | -1.5029 | -1.4923 | | | |
| fha_fico | -0.7521 | 0.1766 * | -0.8581 | | | |
| _cons | -14.5671 | -13.4036 | -14.7711 | -14.5425 | -14.0305 | -12.9860 |
| **Statistics** | **FRM 30** | **FRM 15** | **ARM** | **SR FRM 30** | **SR FRM 15** | **SR ARM** |
| Log likelihood | -208487 | -13059 | -75132 | -36883 | -3508 | -11795 |
| Number of obs | 9,997,020 | 1,546,861 | 3,256,078 | 2,662,272 | 1,103,811 | 590,162 |
| Prob > $\chi^2$ | 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 Prepay Rate Model Estimation**

| Variable | FRM 30 | | FRM 15 | | ARM | | SR FRM 30 | | SR FRM 15 | | SR ARM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| loancat_cat_2 | 0.3796 | | 0.2090 | | 0.3372 | | 0.3445 | | 0.1053 | | 0.3027 | |
| loancat_cat_3 | 0.6504 | | 0.3749 | | 0.5408 | | 0.5716 | | 0.1856 | | 0.4816 | |
| loancat_cat_4 | 0.8178 | | 0.4931 | | 0.6441 | | 0.7127 | | 0.2653 | | 0.5928 | |
| loancat_cat_5 | 0.9401 | | 0.6099 | | 0.6617 | | 0.8336 | | 0.4164 | | 0.7116 | |
| ltvcat_cat_2 | -0.1297 | | -0.0738 | | -0.1192 | | | | | | | |
| ltvcat_cat_3 | -0.1367 | | -0.0997 | | -0.1070 | | | | | | | |
| ltvcat_cat_4 | -0.0969 | | -0.0769 | | -0.0220 | * | | | | | | |
| ltvcat_cat_5 | -0.0536 | | -0.0622 | | -0.0152 | * | | | | | | |
| season_cat_2 | 0.2044 | | 0.2178 | | 0.1992 | | 0.1960 | | 0.1694 | | 0.1909 | |
| season_cat_3 | 0.1052 | | 0.1104 | | 0.0349 | | 0.1191 | | 0.1112 | | 0.1245 | |
| season_cat_4 | 0.0951 | | 0.0733 | | 0.0247 | * | 0.0791 | | 0.0428 | * | 0.0255 | * |
| pneqcat_cat_2 | -0.2228 | | -0.3363 | | -0.3166 | | -0.3322 | | -0.2565 | | -0.3922 | |
| pneqcat_cat_3 | -0.2929 | | -0.5141 | | -0.4591 | | -0.3661 | | -0.5364 | | -0.5206 | |
| pneqcat_cat_4 | -0.4014 | | -0.6312 | | -0.5738 | | -0.5220 | | -0.5055 | | -0.6058 | |
| pneqcat_cat_5 | -0.5593 | | -0.6545 | | -0.7647 | | -0.8246 | | -0.6583 | | -0.9704 | |
| pneqcat_cat_6 | -0.6755 | | -0.6732 | | -1.0173 | | -0.9204 | | -0.8134 | | -1.0774 | |
| pneqcat_cat_7 | -0.7169 | | -1.0550 | | -1.2381 | | -1.1853 | | -0.8717 | | -1.5507 | |
| ycslopecat_cat_2 | 0.1015 | | -0.0155 | * | -0.0201 | * | 0.0894 | | 0.1447 | | 0.2727 | |
| ycslopecat_cat_3 | 0.3002 | | 0.1575 | | 0.2091 | | 0.1586 | | 0.2012 | | 0.2515 | |
| ycslopecat_cat_4 | 0.6096 | | 0.4199 | | -0.0589 | | 0.6325 | | 0.6023 | | 0.2550 | |
| spreadcat_cat_2 | 0.6219 | | 0.0611 | * | 0.0993 | | -0.8371 | | | | 0.1171 | |
| spreadcat_cat_3 | 0.4873 | | 0.3152 | | 0.1710 | | -0.6758 | | | | 0.2371 | |
| spreadcat_cat_4 | 0.5281 | | 0.5086 | | 0.2889 | | -0.4450 | | | | 0.4053 | |
| spreadcat_cat_5 | 0.7515 | | 0.7284 | | 0.4863 | | -0.2028 | | | | 0.6103 | |
| spreadcat_cat_6 | 1.3460 | | 1.0678 | | 0.7248 | | 0.3343 | | | | 0.7065 | |
| spreadcat_cat_7 | 1.7161 | | 1.3047 | | 1.8778 | | 0.6125 | | | | 1.5220 | |
| spreadcat_cat_8 | 1.6559 | | 1.2748 | | 2.3680 | | 0.6035 | | | | 1.5220 | |
| inmoneycat_cat_2 | 0.2726 | | 0.2217 | | 0.2903 | | 0.3648 | | 0.3976 | | | |
| inmoneycat_cat_3 | 0.5750 | | 0.3929 | | 0.3858 | | 0.5973 | | 0.6066 | | | |
| inmoneycat_cat_4 | 0.6106 | | 0.3607 | | | | 0.6123 | | 0.6619 | | | |
| inmoneycat_cat_5 | 0.5360 | | 0.2859 | | | | 0.5299 | | 0.6239 | | | |
| inmoneycat_cat_6 | 0.4593 | | 0.1742 | | | | 0.4523 | | 0.5440 | | | |
| gift_ltr_src_cat_2 | 0.0686 | | 0.0008 | * | 0.0396 | | | | | | | |
| gift_ltr_src_cat_3 | 0.1028 | | 0.5139 | | -0.1064 | | | | | | | |
| gift_ltr_src_cat_4 | -0.1319 | | -0.0029 | * | -0.1308 | | | | | | | |
| ey_ratecat_cat_2 | | | | | -0.0394 | | | | | | 0.0203 | * |
| ey_ratecat_cat_3 | | | | | -0.4820 | | | | | | -0.3010 | |
| ey_ratecat_cat_4 | | | | | -0.8638 | | | | | | -0.5040 | |
| ey_ratecat_cat_5 | | | | | -1.3336 | | | | | | -0.7683 | |
| ey_ratecat_cat_6 | | | | | -1.7983 | | | | | | -0.7683 | |
| fy_1975_1986_cat_2 | -0.1222 | | -0.0385 | * | | | | | | | | |
| fy_1986_1992_cat_2 | -0.2884 | | -0.1293 | | -0.0836 | | | | | | | |
| fy_1996_XXXX_cat_2 | 0.1755 | | 0.2105 | | 0.3783 | | 0.5183 | | 0.2291 | | 0.6747 | |
| age1 | 1.0669 | | 0.5342 | | 1.2733 | | 1.0025 | | 0.4844 | | 1.1930 | |
| age2 | 0.3597 | | 0.1489 | | 0.5218 | | 0.0857 | | 0.0192 | * | 0.2167 | |
| age3 | 0.0907 | | 0.0533 | | 0.0834 | | 0.0009 | * | 0.0573 | | -0.0154 | * |
| age4 | 0.0035 | * | 0.0353 | | -0.0247 | | -0.0293 | | 0.0033 | * | -0.0163 | * |
| age5 | -0.0018 | * | -0.0354 | | -0.0358 | | -0.0092 | | 0.0223 | * | -0.0290 | |
| age6 | -0.0288 | | 0.0073 | | -0.0419 | | | | 0.0518 | | -0.0241 | * |
| age7 | -0.0112 | | | | -0.0014 | * | | | -0.0516 | | 0.0295 | |
| age8 | 0.0176 | | | | -0.0043 | * | | | 0.0136 | | -0.0009 | * |

**Exhibit A-3**

**Results for Conditional Prepay Rate Model Estimation**

| Variable | FRM 30 | FRM 15 | ARM | SR FRM 30 | SR FRM 15 | SR ARM |
|---|---|---|---|---|---|---|
| age9 | -0.0188 | | 0.0036 * | | | |
| age10 | -0.0037 * | | 0.0080 * | | | |
| age11 | -0.0162 | | -0.0151 | | | |
| age12 | -0.0007 * | | | | | |
| age13 | 0.0065 | | | | | |
| fico_000 | -0.0944 | 0.0163 * | 0.0291 | | | |
| fico_999 | 0.0540 | 0.1226 | 0.1018 | | | |
| fico_300_499 | -0.4086 | 0.1283 * | -0.4148 | | | |
| fico_500_539 | -0.2472 | 0.1842 | -0.2572 | | | |
| fico_540_579 | -0.1394 | 0.2537 | -0.1441 | | | |
| fico_580_619 | -0.0734 | 0.2304 | -0.0747 | | | |
| fico_640_659 | 0.0469 | 0.1495 | 0.0745 | | | |
| fico_660_679 | 0.0839 | 0.1368 | 0.1061 | | | |
| fico_680_719 | 0.1373 | 0.1035 | 0.1905 | | | |
| fico_720_850 | 0.1867 | 0.0772 | 0.2553 | | | |
| fha_fico | 0.4619 | 0.6626 | -0.0263 * | | | |
| _cons | -8.6419 | -7.6061 | -7.2593 | -6.5185 | -6.5620 | -6.5103 |
| **Statistics** | **FRM 30** | **FRM 15** | **ARM** | **SR FRM 30** | **SR FRM 15** | **SR ARM** |
| Log likelihood | -1192907 | -204857 | -518577 | -452187 | -152383 | -123714 |
| Number of obs | 9,518,355 | 1,638,878 | 3,092,248 | 2,690,072 | 1,199,787 | 614,994 |
| Prob > $\chi^2$ | 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.