Appendix A

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 2006 Review. We first summarize the model specification and estimation issues arising from the analysis of FHA claim and prepayment rates. Then we describe the specific explanatory variables used in the analysis. The model estimation statistics and graphical comparisons of the overall within-sample fit of the models are also provided. Finally, we show graphically the estimated age-of-loan distributions compared with their actual distributions.

I. Model Specification and Estimation Issues

A. Specification of FHA Mortgage Termination Models

For the FY 2006 Review, the TAC Team developed and estimated updated competing risk models for mortgage prepayment and claim terminations. 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. At the end of this Appendix we present graphical comparisons showing the goodness of fit by age.

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 α_C and α_P and the coefficient vectors β_C and β_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 delinquency 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 values and time-to-sale of collateral properties.

For the FY 2006 Review, we apply average loss severity rates observed during FYs 2005 stratified by six mortgage product types and whether borrowers received downpayment assistance from non-profit organizations. 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 default episodes leading to claim terminations to create a prepayment-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 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 (according to the above equations) 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

Once the separate binomial claim- and prepayment-rate models have been estimated by binomial logit estimation, the parameter estimates are combined to compute the appropriate multinomial probabilities. The theory underlying the Begg-Gray method is that the values of parameters $\alpha_C, \beta_C, \alpha_P$, and β_P from separate binomial logit (BNL) models of claims and prepayments are identical to those in 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 $\pi_{\mathrm{BNL}}^{\mathrm{C}}$ and $\pi_{\mathrm{BNL}}^{\mathrm{P}}$, respectively,

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

Then we can substitute directly into the MNL probabilities 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 February 28, 2006. 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 2006 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. Random Sampling

A full 100-percent sample of loan level data from the FHA single-family data warehouse was extracted for the FY 2006 analysis. This produced a starting sample of approximately 20 million single-family loans originated between FY 1975 and the first quarter of FY 2006. At the estimation stage a 10-percent random sample of loans was used to generate loan-level event histories for up to 120 quarters (30 years) of loan life per loan, or until the scheduled maturity date of the loan.

II. Explanatory Variables

Three 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 origination group characteristics, such as the distribution of borrower credit history (FICO scores) within a particular group of homogenous loans, i.e., with the same mortgage product type, amortization term, origination year and quarter, original LTV ratio, original loan amount, and the source of downpayments.

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 listed in Exhibit A-1 were entered as 0-1 dummy variables in the statistical models, with the exception of the mortgage age variables, which were entered directly; and the FICO score variables, which are percentages of loans in a homogenous loan group with FICO score within the specified range. The specification of each variable is described in more detail below.

Mortgage Product Types

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 create 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-age5* defined as follows:

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

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

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

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

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

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 are developed in a similar manner. The number of segments is 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 levels 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 the 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 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 FRM30, FRM15, and ARM loans stratified by product, beginning amortization year and quarter, and geographic location (state and county).

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) published in the first quarter of 2005.

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.

Mortgage Premium (Spread)

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. 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 any one adjustment 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.

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 yield to the one-year Constant Maturity Treasury 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. 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 new burnout factor measures the average of the number of basis points the borrower was in the money, for all quarters during which the borrower was in the money, anytime during the preceding 8 quarters. The resulting measure was categorized into smaller 50 basis point categories that provide a more refined measure of burnout.

Pre-FY 1986 Origination

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

FY 1986-1992 Origination

An indicator for loans originated between FY 1986 Q3 and 1992 Q1 to capture the condition that these loans were underwritten with more strict requirements but had no borrower's credit history information.

Post-1995 Origination

An indicator for loans originated after FY 1995 is included to account for a loosening of FHA underwriting requirements.

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.

C*hange in Metropolitan Area Unemployment Rates*

For the FY 20066 Review we 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 are not 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 new variable considered for the FY 2006 Review was the ARM payment burden. This variable measured the percentage change in the monthly payment since origination on ARM loans. The percentage change was categorized into 5 levels ranging from no increase to more than 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.

Source of Downpayment Assistance

The FHA single-family program has experienced a significant increase in the use of downpayment assistance from relatives, non-profit organizations, and government programs in the past two years, and loans to borrowers utilizing downpayment assistance have been observed to generate higher claim rates.
For the FY 2006 Review we included a series of indicators for the use of different types of downpayment assistance.

Borrower Credit History

Borrower credit history information has been collected by FHA over three separate periods. The first set of sample data was collected for FHA application cases during FYs 1992, 1994, and 1996. FICO scores of the borrower and up to two co-borrowers were collected from Experian for about a random 20 percent of the loans from the application population. The second set of sample data was collected for applications over FYs 1997 to 2001. FICO scores for up to three co-borrowers were collected from both Experian and Equifax for about 20 percent of the loans in each year with over-sampling of loans defaulted by April 2003. The third set of data is similar to the second set, for FYs 2002 and 2003 applications. Again, there was over-sampling of loans defaulted by February 2005 for the random sample.

These three sets of FICO data represent the most reliable sources of borrower credit history information available for FHA endorsed loans. Following the FHA methodology, one single FICO score is derived from up to three scores from the co-borrowers of a loan. The final score is defined as the lower of the two or the middle of the three should multiple scores be available. In order to keep the measure of this credit history information consistent, we choose to discard the scores obtained from Equifax in the second and third data sets.

Because the credit score information is available only in limited origination years and for a limited number of loans, these data are not adequate to support the loan-level categorization variables similar to other origination characteristics as discussed above. To overcome this limitation, we developed a set of categorical FICO score variables to capture the distribution of loans with similar origination characteristics among different range of scores. The credit scores are divided into 8 groups: 400-459, 460-509, 510-559, 560-609, 610-659, 660-709, 710-759, and above 750. A separate category of 000 was created for loans for which no FICO scores were returned from Experian. Another category of 999 is also created for loans that were not selected in the random sample. However, the credit history data set we received does not allow accurate allocation between the last two cases. As a result, the coefficients for these two categories should be interpreted as mixed impacts of the two different reasons for missing scores.

The value being assigned to each FICO score variable is the percentage of loans with similar origination characteristics with borrower FICO scores falling in a specific range indicated by the particular variable. An example will help explain how FICO variables were constructed. Assume there are 5 loans insured by MMI Fund that were originated in the following cohort: originated in second quarter of FY 2001, original loan size between 60 and 90 percent of the state median loan size, initial LTV ratio between 95 and 97, and receiving downpayment assistance from relatives. The single FICO score of these five loans, using the rules above when there are multiple scores for a loan, are 532, 619, no score returned from Experian, and two are not in the FICO collection sample. Then these five loans receive the same set of values for the FICO score variables: fico_510_550 = 0.2, fico_610_650 = 0.2, fico_000 = 0.2, fico_999 = 0.4, and all other FICO score variables are zero. For all loans originated prior to the second quarter of FY 1992, fico_999 is assigned a value equal to one and all other variables take the value of zero.

When simulating the composition of future books of business, all future loans are assumed to follow the same credit score distribution as the comparable loans in the FY 2003 book of business, which is the most recent book to have complete FICO sample data.

Exhibit A-1

Logit Model Explanatory Variables

Variable Name	FRM30	FRM15	ARM	FRM30_SR	FRM15_SR	ARM_SR	Description
Mortgage Age Function							
age1	2	2	2	2	2	2	Piece-wise linear age functions for ages up to specified knot points. Estimated parameters give the slope of the age function for each segment. Functions differ by mortgage product type as indicated.
age2	4	4	4	4	4	4	
age3	8	8	8	8	8	8	
age4	12	12	12	12	12	12	
age5	16	16	16	> 12	16	16	
age6	20	> 16	20		20	20	
age7	24		24		24	24	
age8	28		28		> 24	> 24	
age9	32		32				
age10	40		40				
age11	60		> 40				
age12	80						
age13	> 80						

Variable Name	Values	Description
Loan Size		
loancat_cat_1	$0 < X \le 60$	Relative loan size measured as percent difference from average size loan originated in same state in the same year.
loancat_cat_2	$60 < X \le 90$	
loancat_cat_3	$90 < X \le 110$	
loancat_cat_4	$110 < X \le 140$	
loancat_cat_5	$X > 140$	
Loan-to-Value		
ltvcat_cat_1	$0 < X \le 80$	Missing or zero values replaced with update file provided by FHA. Additional missing values imputed as mean LTV by state, origination FY, and product type.
ltvcat_cat_2	$80 < X \le 90$	
ltvcat_cat_3	$90 < X < 95$	
ltvcat_cat_4	$95 \le X < 97$	
ltvcat_cat_5	$97 \le X$	

(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.
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		
in_moneycat_cat_1	X ≤ 0	Burnout factor equal to the 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_2	0 < X ≤ 50	
in_moneycat_cat_3	50 < X ≤ 100	
in_moneycat_cat_4	100 < X ≤ 150	
in_moneycat_cat_5	150 < X ≤ 200	
in_moneycat_cat_6	X > 200	
Pre-1986 Origination		
fy_1975_1986_cat_1	X ≥ 1986	Post- or pre-FY1986 Q3 origination. Included to account for changes in FHA underwriting standards.
fy_1975_1986_cat_2	X < 1986	
1986-1992 Origination		
fy_1986_1992_cat_1	1986 > X or 1992 ≤ X	Post 1985 origination with no credit history information.
fy_1986_1992_cat_2	1986 ≤ X < 1992	
Post-1995 Origination		
fy_1996_2005_cat_1	X < 1996	Pre-or post-FY1995 origination. Included to account for changes in FHA underwriting standards.
fy_1996_2005_cat_2	1996 ≤ X	
Exposure Year/Quarter FRM Rate		
ey_ratecat_cat_1	X ≤ 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 ≤ 7	
ey_ratecat_cat_3	7 < X ≤ 8	
ey_ratecat_cat_4	8 < X ≤ 9	
ey_ratecat_cat_5	9 < X ≤ 10	
ey_ratecat_cat_6	X > 10	
Metropolitan Unemployment Rates		
uechngcat_1	X ≤ -30	Percent change over the

Exhibit A-1

Logit Model Explanatory Variables		
Variable Name	**Values**	**Description**
uechngcat_2	-30 < X ≤ -20	preceding year in the metro-area unemployment rate.
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	
Distribution of Borrowers FICO Scores		
Fico_400_450	400 < X ≤ 459	Percentage of loans of the same origination quarter, loan type, loan size, and initial LTV category with initial FICO score in the range.
Fico_460_500	460 < X ≤ 509	
Fico_510_550	510 < X ≤ 559	
Fico_560_600	560 < X ≤ 609	
Fico_610_650	610 < X ≤ 659	
Fico_660_700	660 < X ≤ 709	

Exhibit A-1

Logit Model Explanatory Variables

Variable Name	Values	Description
Fico_710_750	710 < X ≤ 759	
Fico_760_800	760 < X ≤ 809	
Fico_000	No FICO Score Available	
Fico_999	Not in FHA's FICO Sample	

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	-0.0464		-0.2835		0.0407	0.2559		0.0940		0.4307	
loancat_cat_3	-0.1599		-0.6065		-0.1056	0.3683		-0.2895		0.3275	
loancat_cat_4	-0.2264		-0.5645		-0.1879	0.3643		-0.1480		0.4574	
loancat_cat_5	-0.2719		-0.9202		-0.3491	0.1781		-0.7535		0.2404	
ltvcat_cat_2	0.5393		0.9622		0.4641						
ltvcat_cat_3	0.4962		1.1170		0.6645						
ltvcat_cat_4	0.5941		1.3198		0.6985						
ltvcat_cat_5	0.5454		1.1792		0.6232						
season_cat_2	-0.0474		0.0309	*	-0.0443	-0.0654		0.0402	*	0.0005	*
season_cat_3	0.0040	*	-0.0638		-0.0508	0.0244		-0.2869		-0.0420	*
season_cat_4	0.0183		-0.0650		-0.0713	0.0042	*	-0.0150	*	-0.0234	*
pneqcat_cat_2	0.4809		0.5084		0.3387	0.6879		0.7465		0.6951	
pneqcat_cat_3	0.6082		0.8892		0.4143	0.9541		1.1619		1.0443	
pneqcat_cat_4	0.7467		0.9299		0.6160	1.2228		-0.0246	*	1.3801	
pneqcat_cat_5	0.8499		0.8815		0.8459	1.3825		0.9579		1.5785	
pneqcat_cat_6	1.0477		0.9682		0.8746	1.6272		1.5079		2.0035	
pneqcat_cat_7	1.3974		1.3261		1.5571	2.3841		1.5911		2.8526	
ycslopecat_cat_2	-0.0897		0.0457	*	-0.1550	-0.3797		-0.1343	*	0.0088	*
ycslopecat_cat_3	-0.0225		0.1394		-0.2120	-0.2557		0.1082	*	-0.0871	*
ycslopecat_cat_4	-0.1071		0.1217		-0.2551	-0.2184		0.2059		-0.2266	
spreadcat_cat_2	0.5450		0.3064		0.1930	-0.6491				0.3720	
spreadcat_cat_3	0.6823		0.1299	*	0.3910	-0.2369				0.1907	
spreadcat_cat_4	0.8914		0.2593		0.3161	0.1631				0.1175	
spreadcat_cat_5	1.0192		0.2605		0.3306	0.5401				0.2003	
spreadcat_cat_6	1.1633		0.4015		0.3516	0.7823					
spreadcat_cat_7	1.3263		0.2432			0.9486					
spreadcat_cat_8	1.4812		0.2776			1.0157					
inmoneycat_cat_2	-0.0770		0.0007	*	0.4450	-0.3588		0.3956		0.2035	
inmoneycat_cat_3	0.1018		0.0724	*	0.5676	-0.1237		0.8561		0.3053	
inmoneycat_cat_4	0.3375		0.3076			0.1106		1.1464		0.3053	
inmoneycat_cat_5	0.5450		0.4068			0.3661		1.4050		0.3053	
inmoneycat_cat_6	0.7594		0.5686			0.5164		1.1337			
gift_ltr_src_cat_2	0.2361		0.4099		0.1440						
gift_ltr_src_cat_3	0.8355		1.3288		0.7264						
gift_ltr_src_cat_4	0.1608		-26.4952	*	-0.3083						

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	-0.0562 *					
fy_1975_1986_cat_2	0.7112	0.7589				
fy_1986_1992_cat_2	-0.0055 *	-0.2225	-0.0639			
fy_1996_2005_cat_2	0.4193	0.0615 *	0.5192	0.6794	-0.2055	1.1050
ey_ratecat_cat_2			-0.1264			-0.3409
ey_ratecat_cat_3			-0.4023			-0.7628
ey_ratecat_cat_4			-0.5013			-0.9100
ey_ratecat_cat_5			-0.4039			-1.0031
ey_ratecat_cat_6			-0.0918			0.1627 *
age1	1.2944	15.0692	1.6927	1.4815	-0.1488 *	14.7032
age2	0.7156	0.5147	0.9457	0.5733	1.3175	0.9491
age3	0.1898	0.3198	0.2622	0.1592	0.2183	0.1693
age4	0.0419	0.0383	0.1060	0.0539	0.0341 *	0.1282
age5	-0.0037	-0.0107 *	0.0122	-0.0268	0.0645	0.0505
age6	-0.0168	-0.0500	-0.0169		-0.0479	-0.1245
age7	-0.0338		-0.0145		-0.0823	0.0368
age8	-0.0503		-0.0123 *		-0.0864	-0.0317
age9	-0.0354		-0.0660			
age10	-0.0133		-0.0389			
age11	-0.0441		-0.0710			
age12	-0.0568					
age13	-0.0611					
fico_400_450	0.4691		0.4846			
fico_460_500	0.1159	0.8933	0.2303			
fico_560_600	-0.0612	-0.7553	-0.3016			
fico_610_650	-0.2499	-0.9456	-0.6180			
fico_660_700	-1.0157	-1.4406	-1.3694			
fico_710_750	-3.5383	-7.9083	-3.6481			
fico_760_850	-6.1236	-1.8927	-6.5589			
fico_000	-3.2308	-2.3869	-2.9930			
fico_999	-2.7450	-1.9971	-2.6361			
_cons	-9.5640	-38.0053	-10.2588	-11.9512	-11.3681	-38.9583
Statistics	**FRM 30**	**FRM 15**	**ARM**	**SR FRM 30**	**SR FRM 15**	**SR ARM**
Log likelihood	-6984510	-111465	-536552	-537873	-31801	-54734
Number of obs	318151000	12421630	21611310	37086480	10180950	2666290
LR χ^2	1311640	20266	107898	82865	3195	11739
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 Prepay Rate Model Estimation

Variable	FRM 30	FRM 15	ARM	SR FRM 30	SR FRM 15	SR ARM
loancat_cat_2	0.3743	0.2246	0.3444	0.3102	0.1406	0.2487
loancat_cat_3	0.6551	0.3911	0.5641	0.5290	0.1956	0.4351
loancat_cat_4	0.8354	0.4927	0.6888	0.6490	0.2608	0.5470
loancat_cat_5	0.9547	0.5743	0.7623	0.7225	0.3890	0.7117
ltvcat_cat_2	-0.1317	-0.0653	-0.0722			
ltvcat_cat_3	-0.1315	-0.0651	-0.0073 *			
ltvcat_cat_4	-0.0818	-0.0345	0.0942			
ltvcat_cat_5	-0.0308	-0.0141	0.0913			
season_cat_2	-0.0541	-0.0663	0.0101	0.0117	0.0222	0.0981
season_cat_3	0.0110	-0.0290	0.0750	0.0662	-0.0082 *	-0.0380
season_cat_4	-0.1518	-0.1891	-0.0689	-0.1227	-0.1682	-0.1359
pneqcat_cat_2	-0.2022	-0.1973	-0.2745	-0.2929	-0.2539	-0.3223
pneqcat_cat_3	-0.2606	-0.3690	-0.4101	-0.3624	-0.4016	-0.4886
pneqcat_cat_4	-0.3310	-0.6532	-0.5240	-0.4923	-0.4926	-0.6183
pneqcat_cat_5	-0.4860	-0.6513	-0.6120	-0.6607	-0.5699	-0.7158
pneqcat_cat_6	-0.5813	-0.7868	-0.8186	-0.8422	-0.6001	-0.9159
pneqcat_cat_7	-0.6945	-0.9863	-1.1001	-1.0847	-0.6779	-1.3006
ycslopecat_cat_2	-0.0332	-0.0685	-0.4540	-0.0603	0.2808	-0.1827
ycslopecat_cat_3	-0.0230	-0.0550	-0.3201	-0.0903	0.1727	-0.3319
ycslopecat_cat_4	0.5009	0.4158	-0.5166	0.5237	0.6921	-0.3345
spreadcat_cat_2	0.6472	0.1419	0.2412	-0.7728		0.2723
spreadcat_cat_3	0.6125	0.4807	0.4803	-0.7098		0.4419
spreadcat_cat_4	0.7597	0.7742	0.7325	-0.3885		0.6419
spreadcat_cat_5	1.2717	1.1750	1.0643	0.0830		0.9325
spreadcat_cat_6	1.8672	1.4867	1.1814	0.5563		
spreadcat_cat_7	2.0906	1.5614		0.7763		
spreadcat_cat_8	2.0943	1.5098		0.8025		
inmoneycat_cat_2	0.0962	0.0208	-0.0017 *	0.2031	0.3814	-0.1563
inmoneycat_cat_3	0.1963	0.0936	-0.1900	0.2560	0.4859	-0.2680
inmoneycat_cat_4	0.2685	0.1440		0.2051	0.5679	-0.5465
inmoneycat_cat_5	0.1788	0.0516		0.1296	0.5457	-0.5465
inmoneycat_cat_6	0.0295	-0.0864		0.0269	0.4933	
gift_ltr_src_cat_2	0.0462	-0.1299	-0.0238			
gift_ltr_src_cat_3	0.0579	0.5320	-0.0965			
gift_ltr_src_cat_4	-0.1595	0.0738 *	-0.1886			
gift_ltr_src_cat_5	0.1187					
fy_1975_1986_cat_2	-0.0477	-0.0456				
fy_1986_1992_cat_2	-0.2649	-0.1826	-0.2016			
fy_1996_2005_cat_2	0.2283	0.2911	0.3263	0.5221	0.3541	0.5608
ey_ratecat_cat_2			-0.3904			-0.2737
ey_ratecat_cat_3			-0.6648			-0.6303

Exhibit A-3

Results for Conditional Prepay Rate Model Estimation

Variable	FRM 30		FRM 15	ARM		SR FRM 30	SR FRM 15	SR ARM
ey_ratecat_cat_4				-1.1123				-0.9290
ey_ratecat_cat_5				-1.6545				-1.1372
ey_ratecat_cat_6				-2.1337				-1.9640
age1	0.5811		0.5253	0.8368		0.3296	0.4956	0.5326
age2	0.2421		0.2868	0.3101		0.0743	0.1044	0.1080
age3	0.0387		0.0713	0.0372		-0.0104	0.0702	-0.0356
age4	0.0147		0.0384	-0.0218		-0.0150	-0.0369	-0.0257
age5	-0.0061		-0.0473	-0.0301		-0.0073	0.0605	-0.0199
age6	-0.0312		0.0083	-0.0313			0.0284	-0.0483
age7	0.0090			-0.0153			-0.0470	0.0590
age8	0.0000	*		0.0102			0.0160	-0.0092
age9	-0.0148			0.0112				
age10	0.0023			0.0001	*			
age11	-0.0226			-0.0195				
age12	-0.0004							
age13	0.0062							
_cons	-7.0643		-6.7001	-5.2721		-4.7223	-5.7689	-3.9987

Statistics	FRM 30	FRM 15	ARM	SR FRM 30	SR FRM 15	SR ARM
Log likelihood	-41937930	-1635049	-3876545	-6964694	-1392756	-601591
Number of obs	322526440	12730850	22129560	38568750	10483320	2789830
LR χ^2	8160861	189434	583302	1052865	99612	65696
Prob > χ^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

* Not significant for 0.05-level asymptotic normal test

Literature Cited

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.

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