VI. Summary of Methodology

A. Specification of FHA Mortgage Termination Models

Prepayment and claim rates estimates were based on a multinomial logit model for quarterly conditional probabilities of prepayment and claim terminations. The general approach is similar to 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, while 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 starting point for specification of the loan performance models was a multinomial logit model of quarterly conditional probabilities of prepayment and claim terminations. The corresponding mathematical expressions for the conditional probabilities of claim $(\pi_C(t))$, prepayment $(\pi_P(t))$, or remaining active $(\pi_A(t))$ over the time interval from t to $t+1$ are given by:

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

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

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

Constant terms α_C and α_P, and 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 (versus remaining active), and $X_P(t)$ is the vector of explanatory variables for the conditional probability of prepaying (versus remaining active). Some elements of $X_C(t)$ and $X_P(t)$ are constant over the life of the loan and other are functions of mortgage age.

Following an 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 FHA claim terminations and the censoring of potential prepayment outcomes at the onset of default episodes that ultimately lead to claims.

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. 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 control. Sampling weights were used to account for differences in the number of identical loans in each loan strata.

FHA conditional claim and prepay rates increase relatively quickly during the first two years following mortgage origination before peaking and descending more slowly over the remaining life of the loan. We applied a series of piece-wise linear spline functions to model the impact of mortgage age on conditional claim and prepayment probabilities. This approach is sufficiently flexible to provide a close fit during the first two to three years following mortgage origination, including the peak years of claim or prepayment risk, while limiting the overall number of model parameters that have to be estimated.

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

Because of an accelerated delivery schedule, we were required to apply loss severity rates provided by FHA for the FY 2004 Review. For consistency with the available data on loss rates, the incidence and timing of mortgage default-related terminations must be 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.

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

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

1. 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.

2. 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 the default episodes (and all subsequent quarters).

3. 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).

4. The separate sets of binomial parameter estimates were recombined mathematically to derive the corresponding multinomial logit model for the joint probabilities of prepayment and claim terminations.

C. Loan Event Data

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. 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 first quarter of FY 2004 if the loan has 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.

D. Random Sampling

A 10-percent random sample of loan level data from the FHA single-family data warehouse was extracted for the FY 2004 analysis. This produced a starting sample of approximately 1.8 million single-family loans originated between FY 1975 and the first quarter of FY 2004.

E. Cash Flow Model

After the future claim and prepayment rates are projected by the econometric models, the corresponding cash flows can be computed. The cash flows modeled include: 1) upfront mortgage insurance premia, 2) annual mortgage insurance premia, 3) claim losses, and 4) premium refunds. Two other cash flows were modeled in previous reviews but are not included in our analyses. The administrative expense was discontinued with the FY 2002 Actuarial Review according to federal credit reform requirements, and distributive shares were suspended in 1990. There is no indication that either of these will be resumed in the foreseeable future. We discount the future cash flows back to the end of FY 2004 to determine the present value of future cash flows component of the economic value.