Appendix A: Econometric Analysis of Mortgages

This appendix describes technical details of the econometric models used to analyze the historical and future performance of FHA single-family loans for the FY 2004 Review. We first summarize the model specification and estimation issues arising from the analysis of FHA claim and prepayment rates. Then 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 provided thereafter.

I. Model Specification and Estimation Issues

A. Specification of FHA Mortgage Termination Models

For the FY 2004 Review, the TAC Team developed and estimated new models for mortgage prepayment and claim terminations based on a competing-risk specification. 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 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 more closely than annual data (as used previously) 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. 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. Sampling weights were used to account for differences in the number of identical loans in each loan strata.

The present analysis differs from 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 difference from the Calhoun-Deng (2002) study was the treatment of mortgage age 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 default 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 prepay rates also show a more rapid increase during the early part of the loan life. We found a quadratic specification to be insufficiently flexible 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 two to three years following mortgage origination, while still providing a good fit over the later ages and 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 of our final model estimates.

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 elements 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

As mentioned above, timing differences between borrower default episodes and actual FHA claims led us to apply the Begg-Gray method of estimating separate binomial logit models for FHA prepayment and claim terminations and then recombining the parameter estimates to derive the corresponding multinomial logit model. The issue in this case is the time lag between the time that a borrower decides to cease payment on a loan, *i.e.*, default, and when FHA actually receives the claim from the servicer. 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 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, differences in lender loss-mitigation policies, and with current economic conditions that affect the values and time-to-sale of collateral properties.

Because of the accelerated delivery schedule, TAC was required to apply the 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 claim-rate model was estimated that accounted for the censoring of potential claim terminations by observed prepayments. 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 cannot simultaneously censor loans at the onset of default episodes and retain the same observations for estimating subsequent claim termination rates.

The Begg-Gray methodology produces parameter estimates that are theoretically equivalent to those in the multinomial logit model. By estimating the prepayment and claim rate models separately, we can isolate the issues associated with the timing of claims from the estimation of the parameters of the prepayment function. Failure to exclude defaulting loans from the sample of loans assumed to be at risk of prepayment would result in downward bias in the estimates of conditional probabilities of prepayment because loans with 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:

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 at 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 logit parameter estimates were recombined mathematically to derive the corresponding multinomial logit model for the joint probabilities of prepayment and claim terminations.

B. Computation of Multinomial Logit Parameters from Binomial Logit Parameters

Once the separate binomial claim rate and prepayment rate models have been estimated, the parameter estimates must be combined to compute the multinomial probabilities. The theory underlying the Begg-Gray method is that the values of parameters α_C, β_C, α_P, and β_P from separate binomial logit (BNL) models are identical to those in the corresponding multinomial logit (MNL) model. 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 π_{BNL}^{C} and π_{BNL}^{P}, respectively,

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

Then we can substitute directly into the MNL probabilities for $e^{\alpha_C + X_C\beta_C}$ and $e^{\alpha_P + X_P\beta_P}$:

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

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

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

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

C. 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. 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.

II. Explanatory Variables

Three main categories of explanatory variables were developed:

1. Fixed 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. Dynamic variables based entirely on the loan information, such as mortgage age, season of the year, and scheduled amortization of the loan balance; and

3. 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 rates 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 specific 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. The specification of each variable is described in more detail as 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:

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

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

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

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

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

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 is given by:

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

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*, corresponding to average-sized loans plus or minus 10 percent, *i.e.*, 90 to 110 percent of the size of the average sized loan.

Loan-to-Value Ratio

Several original LTV ratios were recorded in FHA's data warehouse. The specific variable chosen for this year's analysis is based on the availability throughout the sample period and the consistency with previous Reviews. The selected LTV ratio variable may exceed 100 percent due to FHA's practice of allowing the financing of some closing costs.

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 2004.

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-Sheller (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, we required 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:

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

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 prepay due to the difficulty of qualifying and more likely to exercise the default option.

Pre-1986 Origination

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

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.

Exhibit A-1

Logit Model Explanatory Variables

Mortgage Age Function

Variable Name	FRM30	FRM15	ARM	FRM30_SR	FRM15_SR	ARM_SR	Description
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	> 12	16	16	16	> 12	
age6	20		20	20	20		
age7	24		24	24	> 20		
age8	28		28	28			
age9	32		32	32			
age10	40		40	40			
age11	60		> 40	60			
age12	100			80			
age13	> 100			> 80			

Loan Size

Variable Name	Values	Description
loancat_cat_1	0 < X ≤ 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 ≤ 90	
loancat_cat_3	90 < X ≤ 110	
loancat_cat_4	110 < X ≤ 140	
loancat_cat_5	X > 140	

Loan-to-Value

Variable Name	Values	Description
ltvcat_cat_1	0 < X ≤ 80	Loan-to-value at origination. 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 ≤ 90	
ltvcat_cat_3	90 < X ≤ 95	
ltvcat_cat_4	95 < X ≤ 97	
ltvcat_cat_5	97 < X ≤ 98	
ltvcat_cat_6	98 < X ≤ 100	
ltvcat_cat_7	X > 100	

(continued on following page)

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 are 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	

(continued on following page)

Logit Model Explanatory Variables		
Variable Name	**Values**	**Description**
Burnout Factor		
		Burnout factor equal to number of quarters in preceeding 8 quarters in which the difference between the mortgage coupon rate and FRM average market rate exceeded 200 basis points.
burnout_cat_1	$0 \leq X < 2$	
burnout_cat_2	$2 \leq X$	
Pre-1986 Origination		
pre_fy86_cat_1	$X \geq 1986$	Post- or pre-FY1986 Q3 origination. Included to account for changes in FHA underwriting standards.
pre_fy86_cat_2	$X < 1986$	
Post-1995 Origination		
post_fy95_cat_1	$X \leq 1995$	Pre- or post-FY1995 origination. Included to account for changes in FHA underwriting standards.
post_fy95_cat_2	$X > 1995$	
Exposure Year/Quarter FRM Rate		
ey_ratecat_cat_1	$X \leq 6$	FRM average mortgage rate during exposure year and quarter. Included to distinguish high-rate and low-rate environments.
ey_ratecat_cat_2	$6 < X \leq 7$	
ey_ratecat_cat_3	$7 < X \leq 8$	
ey_ratecat_cat_4	$8 < X \leq 9$	
ey_ratecat_cat_5	$9 < X \leq 10$	
ey_ratecat_cat_6	$X > 10$	

III. Model Estimation Results

Exhibit 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

Variables	FRM 30	FRM 15	ARM	SR 30	SR 15	SR ARM
loancat_cat_2	-0.059127	-0.323259	-0.101461	0.1552599	-0.418874	0.030617 *
loancat_cat_3	-0.189945	-0.539008	-0.260933	0.2001736	-0.65505	0.2632803 *
loancat_cat_4	-0.251595	-0.601671	-0.352433	0.2289639	-0.862406	0.1273762 *
loancat_cat_5	-0.286181	-0.625955	-0.475125	0.0080371 *	-1.009584	-0.469389 *
ltvcat_cat_2	0.5415291	1.252786	0.3383154	0.4125113	1.106835	0.6402663
ltvcat_cat_3	0.479359	1.377117	0.4794186	-0.169367	0.5850388	0.8309671
ltvcat_cat_4	0.5513676	1.535668	0.4830865	-0.169972	0.1289049 *	0.2582303 *
ltvcat_cat_5	0.4162305	1.175339	0.379324	-0.400741	-0.621651 *	0.2213682 *
ltvcat_cat_6	0.4412736	1.145758	0.6612088	-0.761921	0.630049 *	-0.914661 *
ltvcat_cat_7	0.4273246	1.865465	0.6775264	-2.662702	-0.262932 *	-2.006375 *
season_cat_2	0.0032691 *	0.0152297 *	-0.004105 *	0.0014188 *	-0.028065 *	0.1908225 *
season_cat_3	0.0015523 *	-0.003789 *	-0.043223 *	0.0115247 *	-0.061399 *	0.0504427 *
season_cat_4	-0.000161 *	0.0453058 *	-0.043964 *	-0.015407 *	0.1423882 *	0.0515423 *
pneqcat_cat_2	0.4879898	0.5717242	0.3193929	0.8896564	0.8439571	0.3584491
pneqcat_cat_3	0.5870397	0.9630352	0.4398417	1.278723	1.216812	1.209338
pneqcat_cat_4	0.7319887	0.8259592	0.6974526	1.597493	1.56455	1.539744
pneqcat_cat_5	0.8886045	1.504057	0.8543484	1.670007	1.56455	1.735483
pneqcat_cat_6	0.9821404	0.6083473	0.9541634	1.820372	1.56455	1.900922
pneqcat_cat_7	1.405699	1.747764	1.497258	2.460649	1.56455	2.085184
ycslopecat_cat_2	-0.143477	-0.186729 *	-0.119456	-0.355404	-0.119157 *	-0.077363 *
ycslopecat_cat_3	-0.09393	-0.014518 *	-0.188235	-0.140319	0.463319 *	-0.282084 *
ycslopecat_cat_4	-0.241352	-0.210743 *	-0.204688	-0.389995	0.1256672 *	-0.084847 *
spreadcat_cat_2	0.5039708	-0.547187 *	0.4074651	-0.463786		0.5317923
spreadcat_cat_3	0.6596004	0.0249898 *	0.5737952	-0.582546		0.5453767
spreadcat_cat_4	0.8978683	0.0317503 *	0.5236798	-0.255914 *		0.6359308

(continued on following page)

Results for Conditional Claim Rate Model Estimation

Variables	FRM 30	FRM 15	ARM	SR 30	SR 15	SR ARM
spreadcat_cat_5	1.052506	0.1948536 *	0.6146395	0.0561017 *		0.6807195
spreadcat_cat_6	1.325193	0.3620021 *	0.6377594	0.4209794		0.6807195
spreadcat_cat_7	1.559481	0.4988952	0.6377594	0.7489561		0.6807195
spreadcat_cat_8	1.801157	0.7892237	0.6377594	0.8418429		0.6807195
burnout_cat_2	0.4342724	0.2038309	0.2213994	0.4184656	0.8162186	0.7673641 *
pre_fy86_cat_2	0.7053814	0.8363768				
post_fy95_cat_2	0.5092742	0.2582393	0.7652759	0.6133092	-0.241814 *	0.8906455
ey_ratecat_cat_2			0.0355879 *			0.0573336 *
ey_ratecat_cat_3			-0.294925			-0.252967 *
ey_ratecat_cat_4			-0.399472			-0.203672 *
ey_ratecat_cat_5			-0.286929			-0.284185 *
ey_ratecat_cat_6			-0.372539			-0.804066 *
age1	1.439529	1.068697	13.12894	1.484563	0.997534 *	1.489061 *
age2	0.7418603	0.8250221	1.19028	0.7552092	0.8923219	1.032595
age3	0.1884129	0.219924	0.302934	0.1217118	0.2548741	0.1624608
age4	0.0272172	0.1171457	0.1283762	0.0266551 *	0.0538606 *	0.1624583
age5	0.0068863	-0.051064	0.0222323 *	0.0100433 *	-0.002986 *	-0.028557
age6	-0.014583		-0.008917 *	-0.016416 *	-0.004506 *	
age7	-0.042699		-0.026618 *	0.0096334 *	-0.088591	
age8	-0.036284		-0.011951 *	-0.042384 *		
age9	-0.024483		-0.051546 *	-0.071162		
age10	-0.016956		-0.044713	0.006393 *		
age11	-0.042284		-0.049453	-0.029298		
age12	-0.06888			-0.091691 *		
age13	0.0099335 *					
constant	-12.52723	-12.42177	-36.27572	-11.80595	-12.64411	-13.39202

Statistics	FRM 30	FRM 15	ARM	SR 30	SR 15	SR ARM
Log likelihood	-650142.4	-10954.06	-50184.23	-46204.42	-3118.532	-4134.626
Number of obs	30237989	1209141	2019894	3193521	894824	211696
LR χ^2	101658.62	2033.66	7845.87	8214.68	330.32	907.54
Prob > χ^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

* Not significant for 0.05-level asymptotic normal test.

Exhibit A-3

Results for Conditional Prepayment Rate Model Estimation

Variables	FRM 30	FRM 15	ARM	SR 30	SR 15	SR ARM
loancat_cat_2	0.3774854	0.2405055	0.3366589	0.3461627	0.103895	0.3668389
loancat_cat_3	0.6580628	0.4026154	0.555617	0.5703086	0.165086	0.5436157
loancat_cat_4	0.8351934	0.5131098	0.6902548	0.7235828	0.212865	0.6423086
loancat_cat_5	0.959084	0.5989075	0.7681204	0.8157152	0.362059	0.7515396
ltvcat_cat_2	-0.12471	-0.08433	-0.125282	0.3324103	0.394324	0.1892471
ltvcat_cat_3	-0.112797	-0.07016	-0.085351	0.1564787	-0.35152	0.0903173
ltvcat_cat_4	-0.04906	-0.01723 *	0.0135105 *	0.2895874	0.02669 *	0.0883167 *
ltvcat_cat_5	0.0653258	0.087207	0.0237051 *	0.3836201	0.032344 *	0.154577
ltvcat_cat_6	-0.100444	-0.02719 *	-0.112238	0.1235851	-0.30984	-0.055141 *
ltvcat_cat_7	0.0965292	0.485906	0.1225901 *	-0.421833	-0.76923	-0.950815
season_cat_2	0.0264095	-0.03446	-0.011839 *	0.104865	0.01979 *	0.0414444 *
season_cat_3	-0.008876	-0.06641	-0.008338 *	0.0645908	-0.08068	0.0032928 *
season_cat_4	-0.167213	-0.18342	-0.144121	-0.144939	-0.19138	-0.167254
pneqcat_cat_2	-0.160575	-0.29206	-0.235888	-0.208523	-0.24959	-0.212186
pneqcat_cat_3	-0.202855	-0.41966	-0.372083	-0.303154	-0.60702	-0.341501
pneqcat_cat_4	-0.283199	-0.57994	-0.52437	-0.258651	-0.83828	-0.346199
pneqcat_cat_5	-0.456931	-0.66291	-0.655225	-0.437535	-0.83828	-0.449962
pneqcat_cat_6	-0.537249	-0.75426	-0.822628	-0.672057	-0.83828	-0.55469
pneqcat_cat_7	-0.622016	-0.87127	-1.05151	-1.173407	-0.83828	-1.183046
ycslopecat_cat_2	-0.077587	-0.13311	-0.324186	-0.127967	0.224906	-0.210546
ycslopecat_cat_3	-0.042031	-0.05497 *	-0.189047	-0.197584	0.104397	-0.133544
ycslopecat_cat_4	0.3762529	0.249806	-0.407626	0.41692	0.784726	-0.228049
spreadcat_cat_2	0.713544	0.123492 *	0.412165	-0.70669		0.2754713
spreadcat_cat_3	0.6468343	0.460037	0.5606156	-0.52494		0.53252
spreadcat_cat_4	0.7816162	0.71912	0.8412846	-0.227644		0.7846497
spreadcat_cat_5	1.270694	1.167638	1.212008	0.3401394		0.9596974
spreadcat_cat_6	2.09974	1.696313	1.316185	1.017943		1.097357
spreadcat_cat_7	2.573372	1.992827	1.316185	1.413977		1.097357

(continued on following page)

Exhibit A-3

Results for Conditional Prepayment Rate Model Estimation

Variables	FRM 30	FRM 15	ARM	SR 30	SR 15	SR ARM
spreadcat_cat_8	2.466031	1.874764	1.316185	1.494489		1.097357
burnout_cat_2	-0.197585	-0.22848	-0.231919	-0.196396	0.137534	-0.144323 *
pre_fy86_ cat_2	0.1196965	-0.00695 *			0.365661	
post_fy95_ cat_2	0.4117027	0.353255	0.4012693	0.5598229		0.5985623
ey_ratecat_cat_2			-0.158862			-0.195456
ey_ratecat_cat_3			-0.418605			-0.541294
ey_ratecat_cat_4			-0.722907			-0.848764
ey_ratecat_cat_5			-1.482059			-1.288601
ey_ratecat_cat_6			-2.061695			-1.8185
age1	0.604297	0.480799	0.8569482	0.4001098	0.473237	0.4600543
age2	0.2133237	0.352898	0.3041368	0.0499918	0.296033	0.0976496
age3	0.050678	0.071385	0.0417813	-0.029628	0.051674	-0.066118
age4	0.0225278	-0.00441 *	-0.020232	-8.13E-05 *	0.019053	-0.031749
age5	-0.019189	0.002711	-0.048453	-0.004526 *	0.019053	0.0006961 *
age6	-0.022818		-0.030281	0.0166283	0.019053	
age7	0.0005209 *		0.0032447 *	0.0069003 *	0.019053	
age8	0.0036652 *		0.0111045 *	0.0085807 *		
age9	0.0013484 *		0.0296337	-0.03104		
age10	-0.014379		-0.023353	-0.002095 *		
age11	-0.020593		-0.023095	-0.032013		
age12	0.0048498			-0.016292 *		
age13	-0.043324					
_cons	-7.346295	-6.82781	-5.828927	-5.416986	-5.98584	-4.530762

Statistics	FRM 30	FRM 15	ARM	SR 30	SR 15	SR ARM
Log likelihood	-3767605	-3790889	-348834.7	-539961.4	-116433	-42580.36
Number of obs	30512978	31179219	2056211	3296316	999306	219089
LR χ^2	787658.83	780900.23	55595.13	104644.06	9677.32	5089.8
Prob > χ^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

* Not significant for 0.05-level asymptotic normal test.

IV. Graphical Comparisons of Goodness-of-Fit by Age of Loan

Exhibits A-4 to A-15 present within-sample comparisons of the overall goodness-of-fit of the binomial logit models for claim and prepayment probabilities. Separate comparisons are given for each of the six mortgage product types. The graphs compare the average value by mortgage age of the observed and predicted conditional claim and prepayment probabilities. The large fluctuations in the observed probabilities at higher values of mortgage age are the result of sampling variation due to the small numbers of surviving loans. Exhibits A-16 to A-21 present comparisons of observed and estimated claim and prepayment frequencies for all mortgages by origination year, termination year, and the age of the loan.

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.

Exhibit A-4

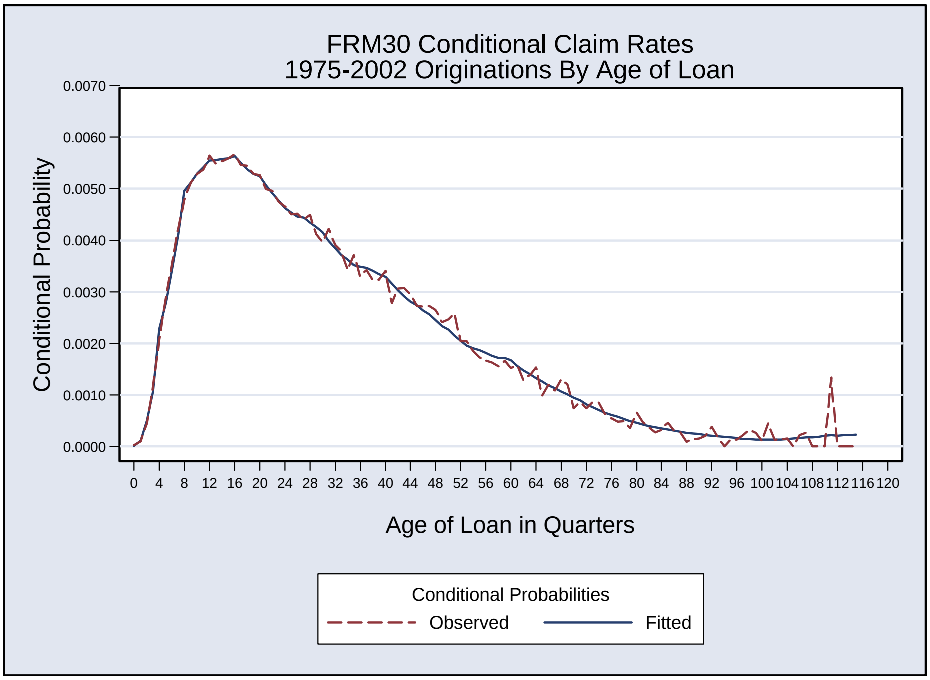

Exhibit A-5

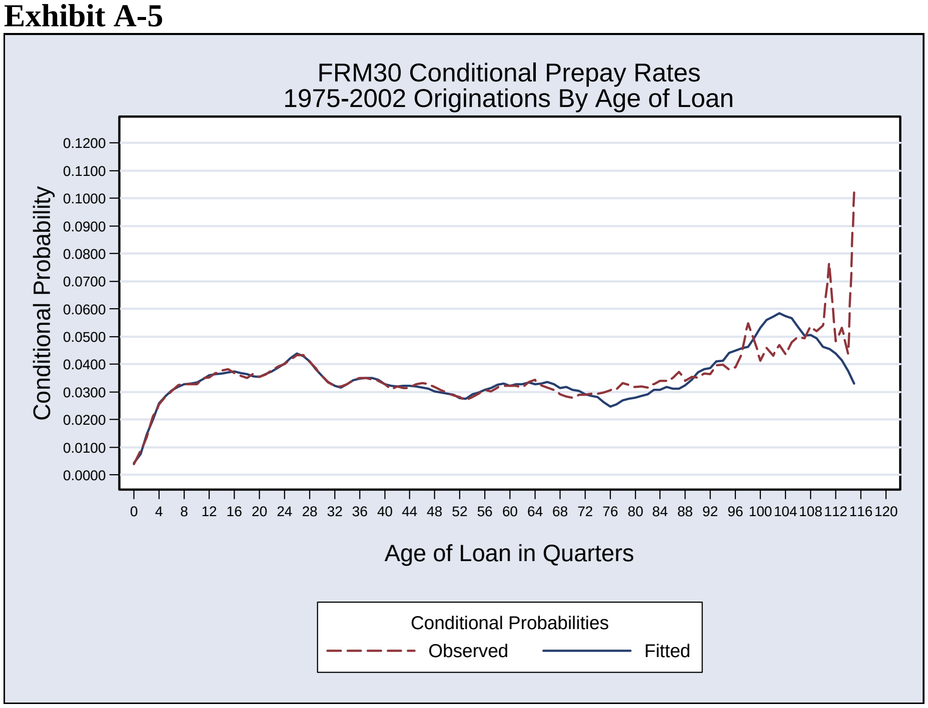

Exhibit A-6

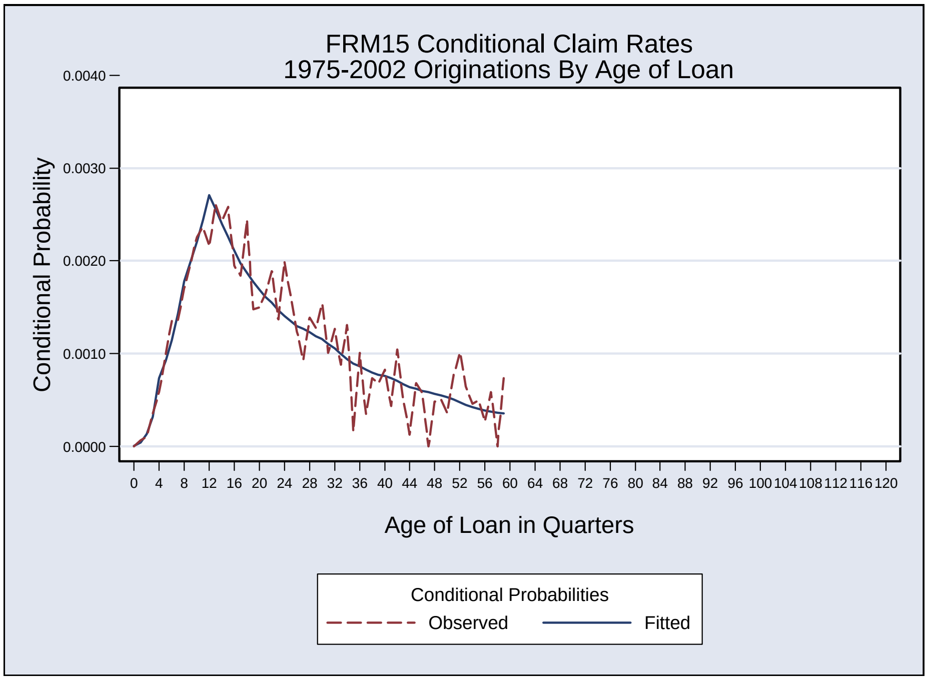

Exhibit A-7

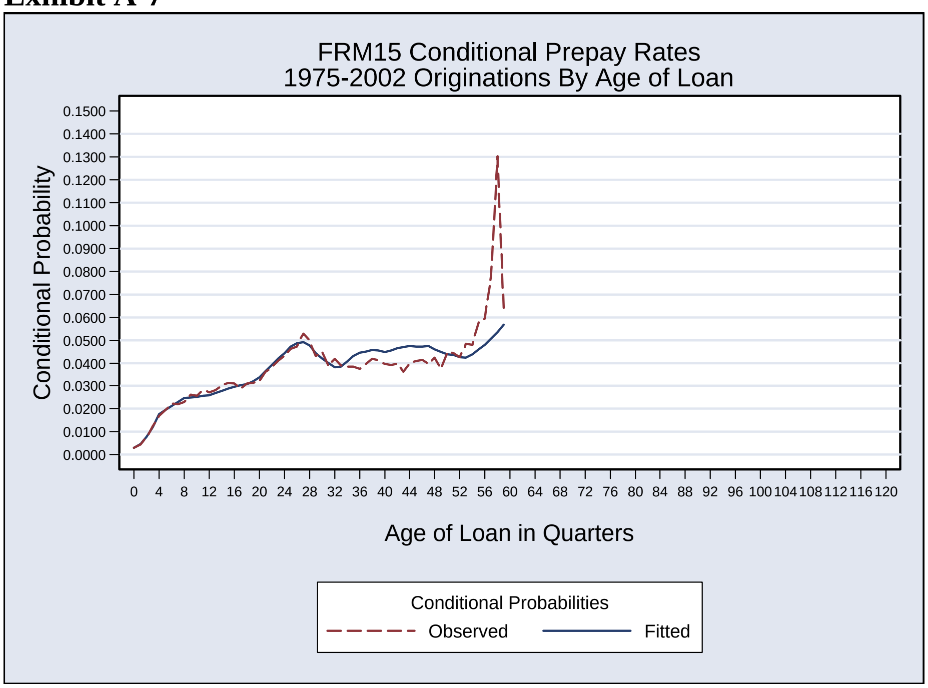

Exhibit A-8

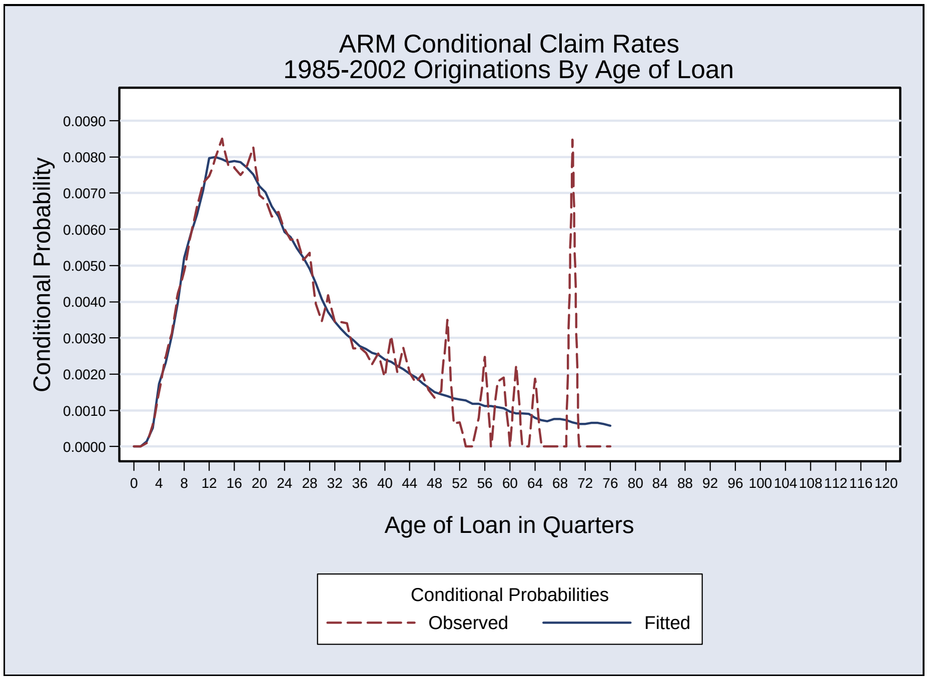

Exhibit A-9

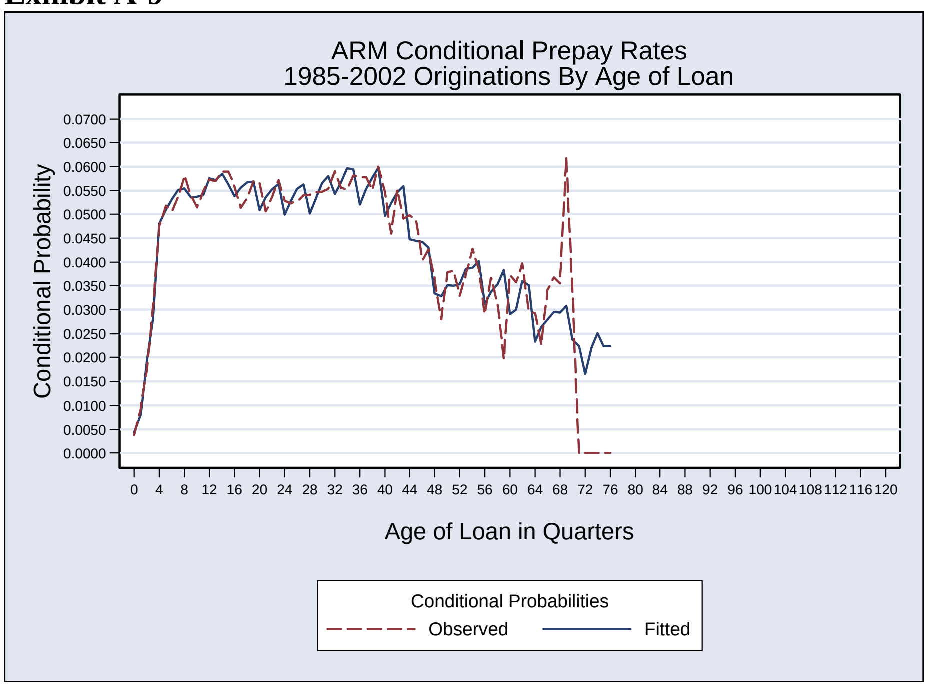

Exhibit A-10

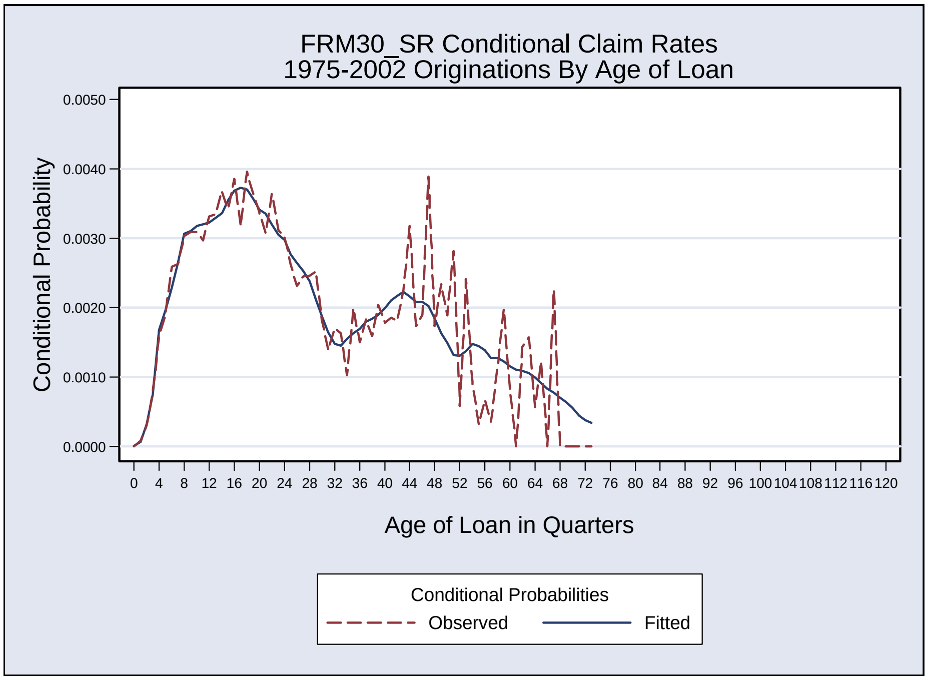

Exhibit A-11

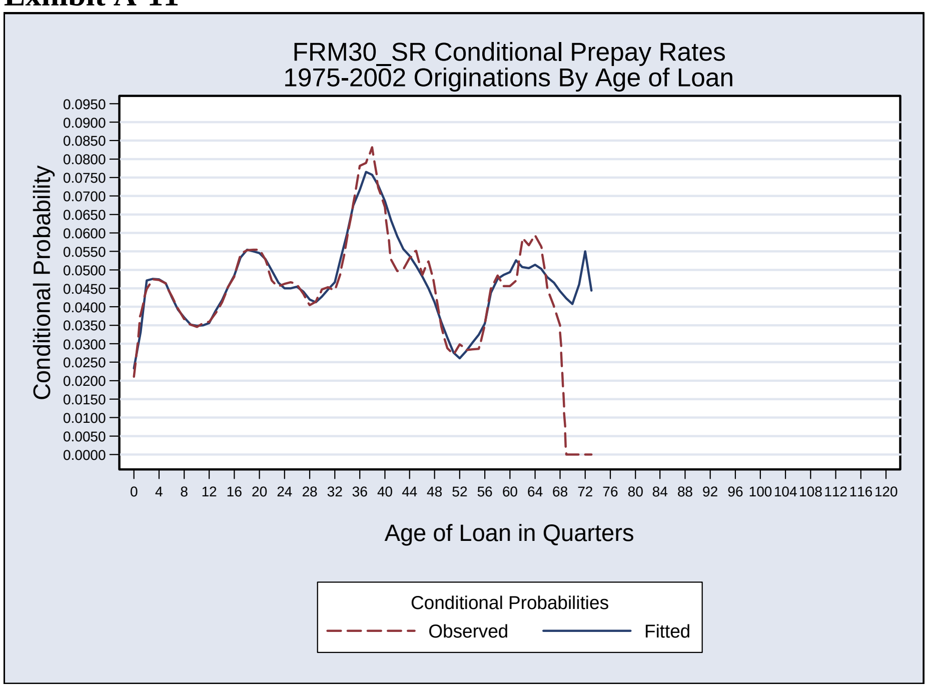

Exhibit A-12

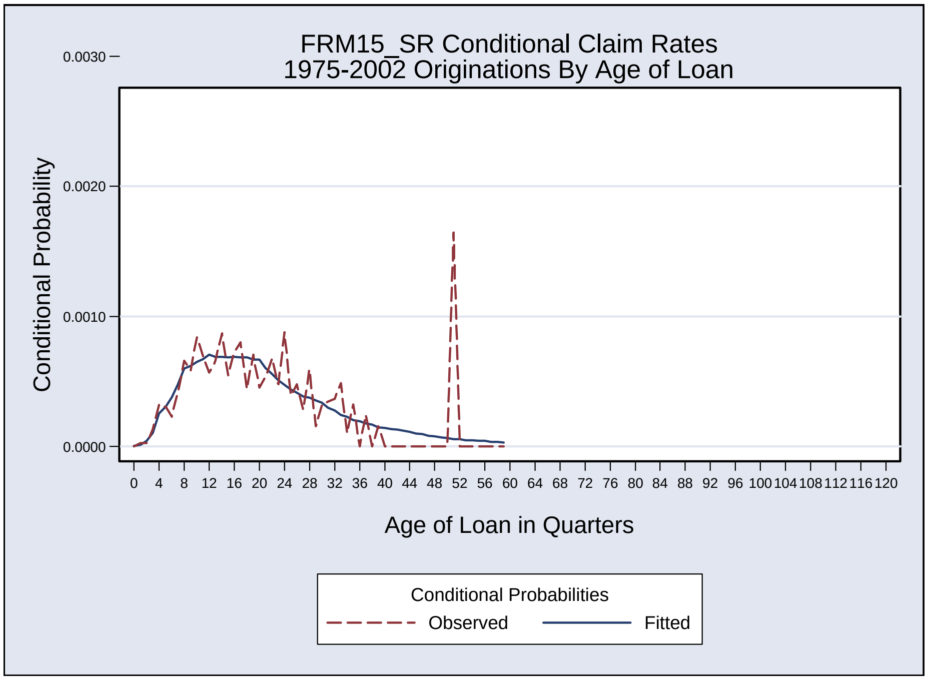

Exhibit A-13

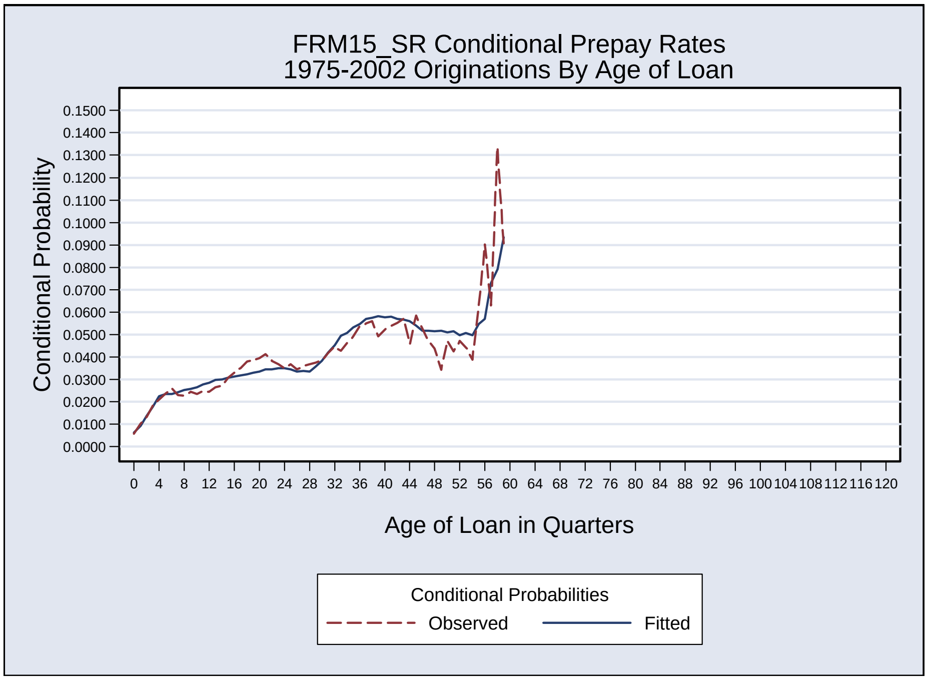

Exhibit A-14

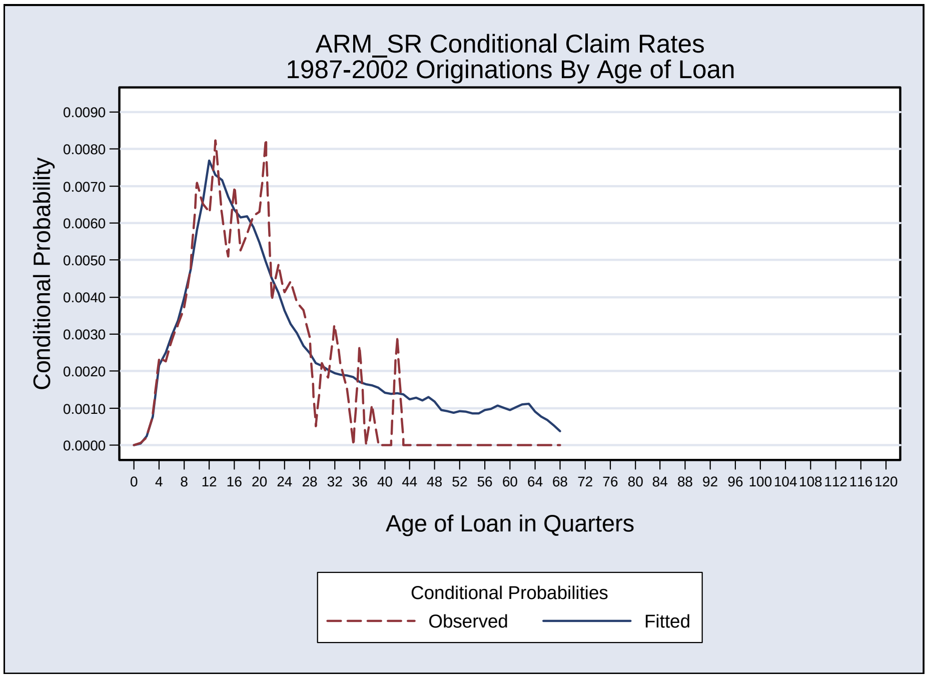

Exhibit A-15

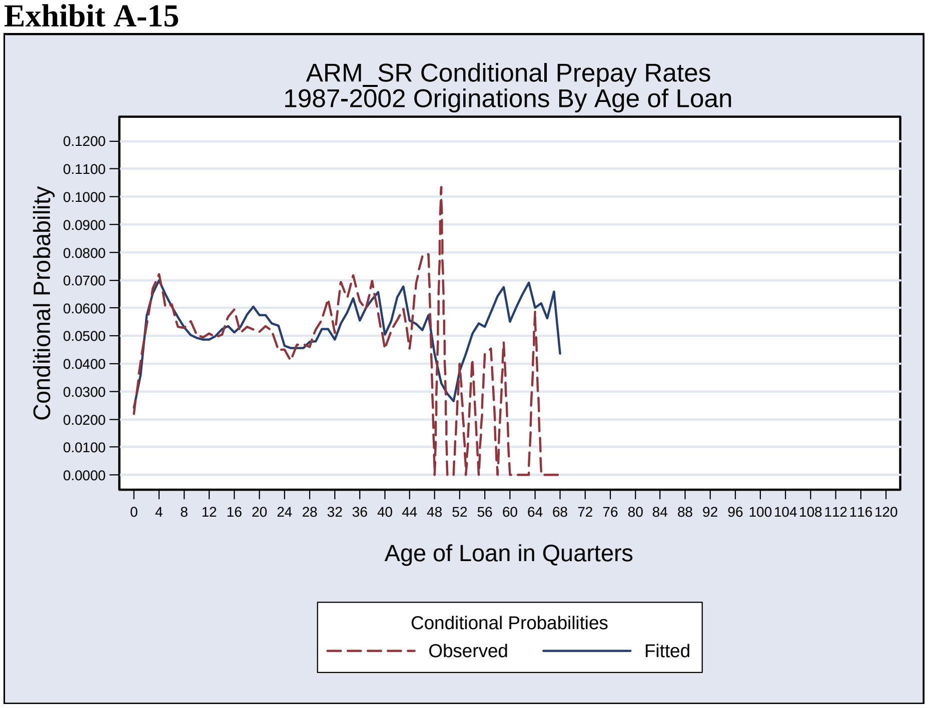

Exhibit A-16

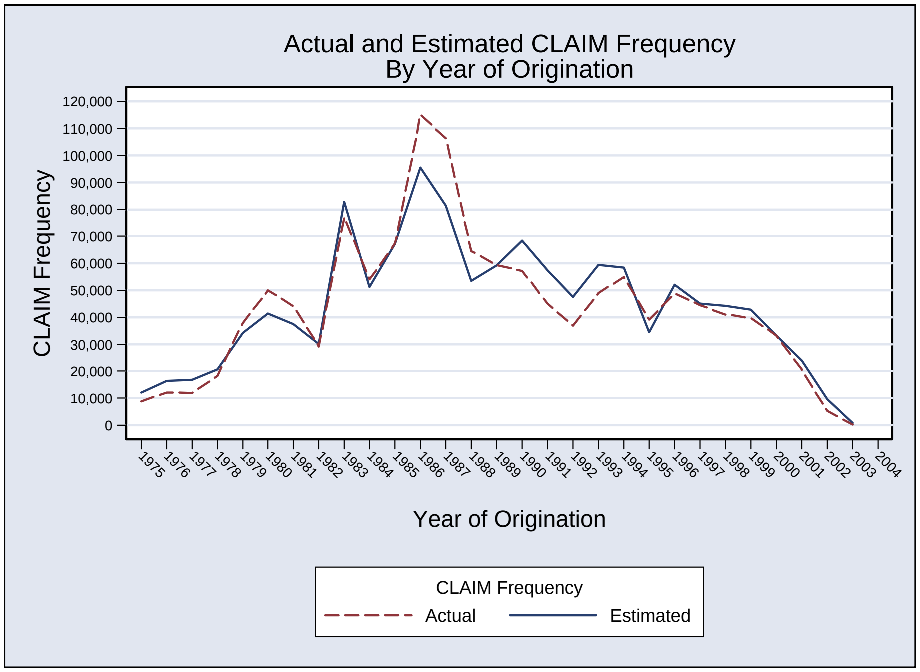

Exhibit A-17

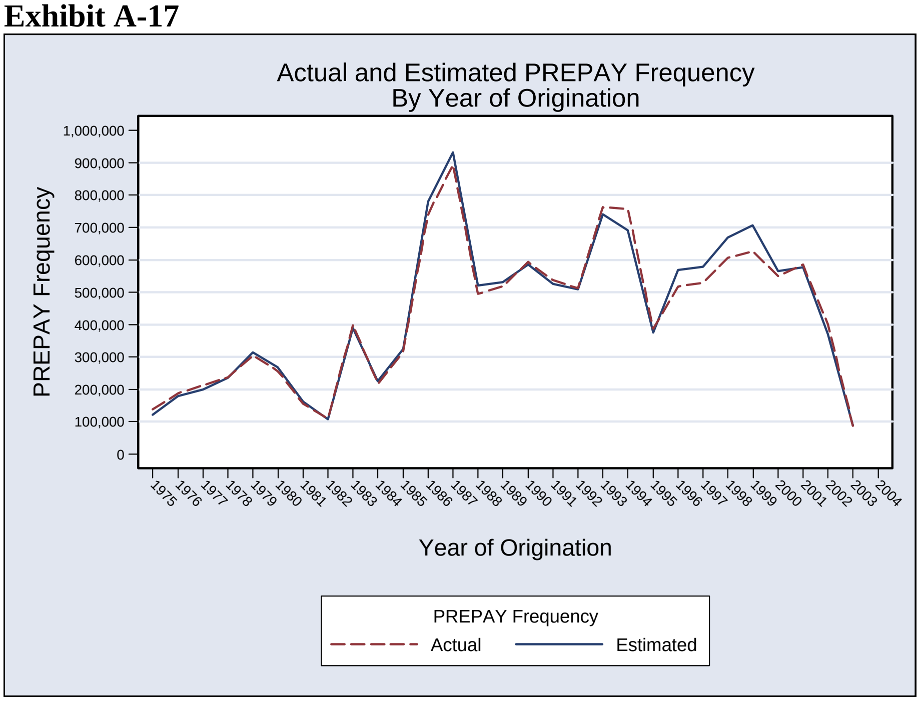

Exhibit A-18

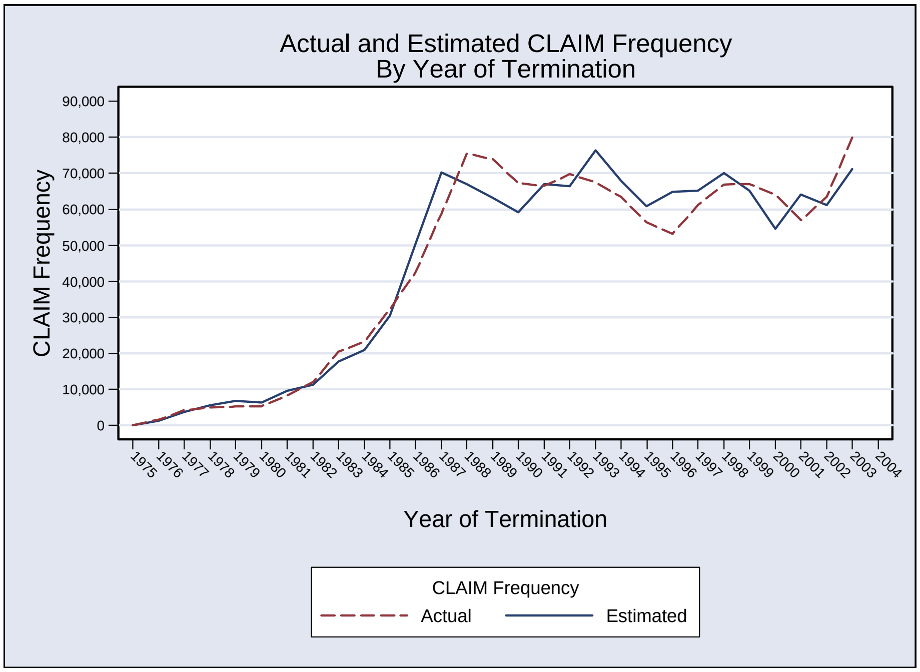

Exhibit A-19

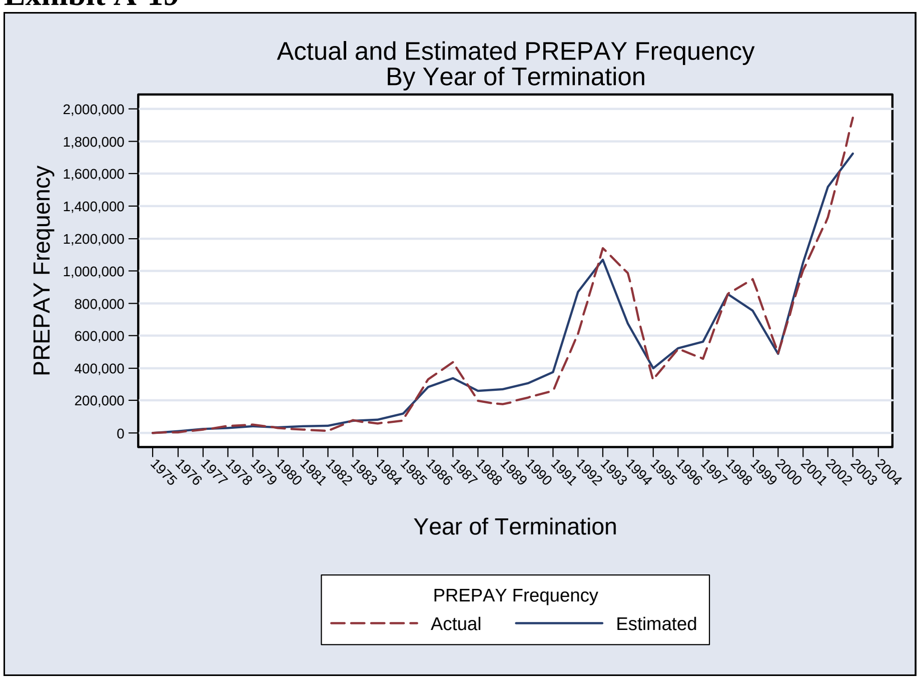

Exhibit A-20

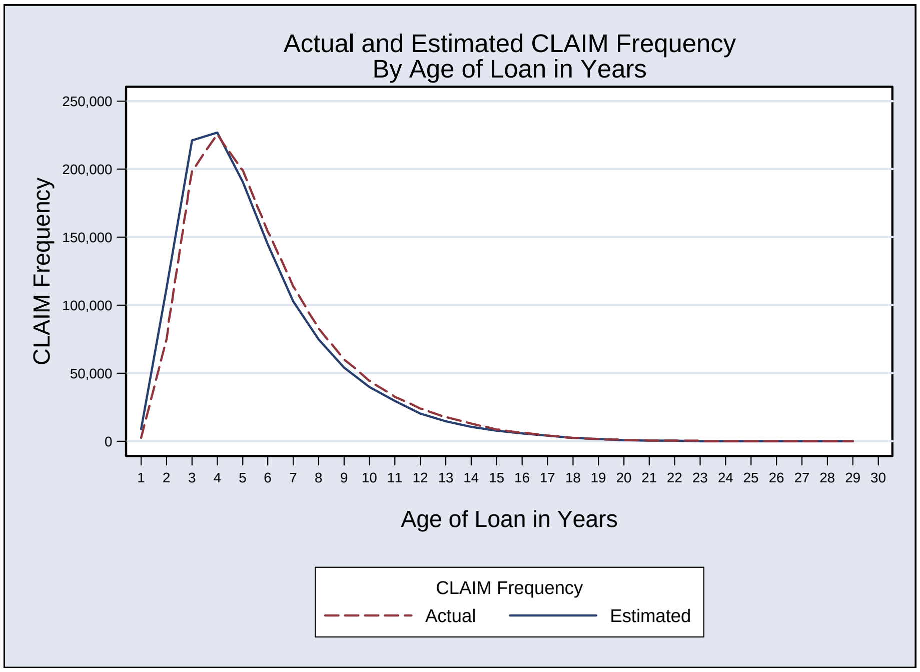

Exhibit A-21