Appendix A – Conditional Claim and Prepayment Rate Models

A class of generalized linear models (GLMs) known as Poisson regression models were developed to model the claim and prepayment patterns of the MMIF's liabilities. A brief review of the theory behind GLMs, Poisson regression models, and the algorithm used to fit this class of models to empirical data is provided in a separate technical appendix – see Appendix J.

In the sections that follow we provide general comments regarding our model-building process for both the claim and prepayment models. In addition, we outline the specific structure of these models along with a description of the various predictor variables.

General Comments that Apply to both the Claim and Prepayment Models

Policy Year 1 Claim and Prepayment Rates

Our models for both the conditional claim and prepayment rates do not attempt to forecast loan terminations during the first policy year. The rates for this early stage in the life of the loan pool are too low and distort the fit of any model for the other policy years. Instead, we reviewed historical claim and prepayment rates during the first policy year for origination years 1975 through 2000[1] and selected conditional claim and prepayment rates for this initial period. Separate policy year 1 rate selections were made for each loan-type/LTV category for which we built a regression model. Table A.1 displays the historical, average, and selected conditional claim rates by loan-type/LTV category and Table A.2 provides the corresponding figures for the conditional prepayment rates.

By selecting the policy year 1 rates, and excluding policy year 1 statistics from the data set used for performing our regressions, we improved the accuracy of the loan termination models in all subsequent policy years and developed more realistic estimates of policy year 1 rates.

[1] Although origination year 2000 is as of 6.30.2000 and is therefore not shown on a full-year basis, we reviewed monthly statistics and did not find any significant evidence of seasonality. Therefore we believe it is appropriate to view the origination year 2000 rates as full-year numbers.

Table A.1

Historical and Selected Policy Year 1 Conditional Claim Rate by Loan-Type/LTV Category

Origination Year	FR30, High LTV	FR30, Mid LTV	FR30, Investor LTV	FR30, Low LTV	FR15, High LTV	FR15, Mid + Investor LTV	FR15, Low LTV	ARM	SRF30	SRF15
1975	0.0616%	0.0274%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%			
1976	0.0562%	0.0195%	0.0160%	0.0000%	0.0000%	0.0000%	0.0000%			
1977	0.0567%	0.0061%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%			
1978	0.0272%	0.0127%	0.0151%	0.0108%	0.5282%	0.0000%	0.0000%			
1979	0.0153%	0.0013%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%			
1980	0.0405%	0.0037%	0.0060%	0.0036%	0.0000%	0.0000%	0.0000%			
1981	0.1189%	0.0275%	0.0098%	0.0126%	0.0000%	0.0000%	0.0000%			
1982	0.2231%	0.0372%	0.0275%	0.0445%	0.0000%	0.0000%	0.0000%			
1983	0.0751%	0.0069%	0.0048%	0.0043%	0.0613%	0.0248%	0.0000%			
1984	0.0483%	0.0305%	0.0082%	0.0042%	0.0788%	0.0168%	0.0000%	0.0000%		
1985	0.0246%	0.0191%	0.0334%	0.0105%	0.0202%	0.0000%	0.0000%	0.0000%		
1986	0.0218%	0.0094%	0.0100%	0.0024%	0.0074%	0.0000%	0.0000%	0.0129%		
1987	0.0153%	0.0115%	0.0061%	0.0173%	0.0000%	0.0019%	0.0000%	0.0079%		
1988	0.0119%	0.0114%	0.0022%	0.0000%	0.0000%	0.0185%	0.0000%	0.0071%	0.0000%	0.0000%
1989	0.0155%	0.0070%	0.0120%	0.0159%	0.0278%	0.0000%	0.0000%	0.0466%	0.0000%	0.0000%
1990	0.0074%	0.0007%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0322%	0.0000%
1991	0.0099%	0.0048%	0.0027%	0.0060%	0.0000%	0.0098%	0.0000%	0.0164%	0.0130%	0.0000%
1992	0.0087%	0.0053%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0024%	0.0040%	0.0000%
1993	0.0040%	0.0000%	0.0000%	0.0000%	0.0078%	0.0000%	0.0000%	0.0053%	0.0047%	0.0000%
1994	0.0023%	0.0161%	0.0000%	0.0111%	0.0000%	0.0000%	0.0000%	0.0008%	0.0074%	0.0011%
1995	0.0050%	0.0000%	0.0000%	0.0000%	0.0117%	0.0000%	0.0000%	0.0121%	0.0371%	0.0000%
1996	0.0027%	0.0041%	0.0073%	0.0000%	0.0000%	0.0000%	0.0000%	0.0020%	0.0144%	0.0000%
1997	0.0099%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0057%	0.0234%	0.0000%
1998	0.0067%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0132%	0.0027%	0.0000%
1999	0.0060%	0.0000%	0.0123%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0081%	0.0000%
2000	0.0038%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
All Yr Avg:	0.0338%	0.0101%	0.0067%	0.0055%	0.0286%	0.0028%	0.0000%	0.0078%	0.0113%	0.0001%
2 Yr Avg:	0.0049%	0.0000%	0.0062%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0040%	0.0000%
Selected:	0.0045%	0.0010%	0.0060%	0.0015%	0.0010%	0.0010%	0.0002%	0.0040%	0.0040%	0.0001%

Table A.2

Historical and Selected Policy Year 1 Conditional Prepayment Rate by Loan-Type/LTV Category										
Origination Year	FR30, High LTV	FR30, Mid LTV	FR30, Investor LTV	FR30, Low LTV	FR15, High LTV	FR15, Mid + Investor LTV	FR15, Low LTV	ARM	SRF30	SRF15
1975	0.1349%	0.2434%	0.2618%	0.6244%	0.0000%	0.2419%	1.6719%			
1976	0.1786%	0.2492%	0.2424%	0.9430%	0.2919%	1.0434%	0.0000%			
1977	0.3003%	0.3974%	0.3169%	0.9659%	0.3224%	0.3893%	2.2693%			
1978	0.2116%	0.3348%	0.3769%	1.1203%	0.0000%	0.0000%	0.0000%			
1979	0.2050%	0.1108%	0.1936%	0.4612%	0.0000%	0.4380%	0.0000%			
1980	0.1378%	0.0876%	0.1677%	1.0245%	0.0000%	0.0000%	1.2074%			
1981	0.1354%	0.1594%	0.1383%	0.3615%	0.0000%	0.3205%	0.8341%			
1982	0.1847%	0.1178%	0.2699%	0.5904%	0.0000%	0.0000%	0.6801%			
1983	0.1895%	0.1339%	0.2366%	0.5325%	0.0607%	0.1239%	0.3073%			
1984	0.0834%	0.1166%	0.2501%	0.4631%	0.0000%	0.1917%	0.4139%	0.0000%		
1985	0.1406%	0.2134%	0.3611%	0.7223%	0.1843%	0.2178%	0.3969%	0.3443%		
1986	0.4757%	0.3568%	0.6677%	0.8494%	0.6375%	0.4095%	0.6550%	0.4378%		
1987	0.1445%	0.1412%	0.3150%	0.5305%	0.2130%	0.3162%	0.5899%	0.0421%		
1988	0.1780%	0.2366%	0.6046%	1.0183%	0.3124%	0.4414%	0.7847%	0.3284%	2.5197%	1.0108%
1989	0.2115%	0.3285%	0.9941%	0.6783%	0.3167%	0.3632%	0.6551%	0.4114%	4.1151%	0.0000%
1990	0.1917%	0.2485%	0.4210%	0.7445%	0.4373%	0.3944%	0.8146%	0.0642%	4.6008%	0.8996%
1991	0.2077%	0.2625%	0.4300%	0.7683%	0.4533%	0.4474%	0.9731%	0.2017%	2.7533%	0.4726%
1992	0.2633%	0.5154%	0.5336%	1.1035%	0.5856%	0.5654%	1.2276%	0.2225%	2.4084%	0.5145%
1993	0.5344%	0.8585%	0.7062%	1.4761%	0.6733%	0.6896%	1.3363%	0.4622%	2.7636%	1.0383%
1994	0.2563%	0.3333%	0.1656%	0.9133%	0.6937%	0.3431%	0.6881%	0.2950%	1.9764%	1.5545%
1995	1.7686%	2.0613%	1.9378%	2.5772%	1.3236%	1.2299%	1.1988%	1.6596%	2.0997%	1.4487%
1996	0.3308%	0.4535%	0.3712%	1.2920%	0.6865%	0.7587%	0.5743%	0.4761%	2.0771%	0.7942%
1997	0.6081%	0.8773%	0.9337%	1.6349%	1.0014%	1.0408%	1.3291%	0.8656%	3.0356%	1.5867%
1998	0.8103%	1.1148%	1.7508%	1.9814%	1.0437%	0.4964%	1.3488%	2.6171%	4.0471%	1.1741%
1999	0.4343%	0.6353%	0.8483%	1.4290%	0.6450%	0.5591%	1.1530%	0.3799%	2.6539%	1.4028%
2000	0.3870%	0.4602%	0.6497%	1.0794%	0.3557%	0.1488%	0.7081%	0.4080%	1.6438%	0.8598%
All Yr Avg:	0.3348%	0.4249%	0.5440%	0.9956%	0.3938%	0.4296%	0.8391%	0.5421%	2.8226%	0.9813%
2 Yr Avg:	0.4106%	0.5477%	0.7490%	1.2542%	0.5004%	0.3540%	0.9306%	0.3940%	2.1488%	1.1313%
Selected:	0.4106%	0.5477%	0.7490%	1.2542%	0.5004%	0.3540%	0.9306%	0.3940%	2.1488%	1.1313%

Minimum Number of Loans/Credibility Criteria

The *Data Transformation* Appendix (see Appendix E) provides a detailed description of how we arranged our data for building our regression models. The data sets we developed are "cell-based" in the sense that individual loans are grouped into cohorts and a time series of statistics are developed for each cohort. Our cohorts are designed, and our regression data sets are developed, at a much finer level of detail than the level at which our final regression models are developed – we did this allow for a detailed investigation of specific model behavior and to develop more homogenous cohorts.

By segmenting the available data too finely we run the risk of fitting our models to spurious results. For example, if at any point in time a given cohort has only a few loans surviving, a single claim or prepayment will produce an observation with a falsely high claim or prepayment rate. To address this issue we dropped from our regression data sets any observation that did not have at least 100 loans surviving in the cohort. In some cases, we increased the minimum credibility criteria to 200. In future versions of our model we would like to implement a formal minimum credibility criterion that varies by loan type/LTV category. For a discussion of credibility theory and its applications, we suggest the text *Introduction to Credibility Theory*, by Thomas N. Herzog, Ph.D., ASA.

Period of Historical Data and Number of Observations

Table A.3 provides a listing of the historical period and the number of observations used in performing the regression for each loan-type/LTV category.

Table A.3

Loan-Type/LTV Category	Historical Period	Number of Observations
FR30, High LTV	1975 through 1999	14,831
FR30, Mid LTV	1975 through 1999	14,498
FR30, Investor LTV	1975 through 1999	7,232
FR30, Low LTV	1975 through 1999	6,958
FR15, High LTV	1975 through 1999	770
FR15, Mid & Investor LTV	1975 through 1999	1,398
FR15, Low LTV	1975 through 1999	723
ARM	1986 through 1999	1,674
SRFR30 (see note)	1988 through 1999	15,600
SRFR15 (see note)	1991 through 1999	1,101

Note: The historical period listed in Table A.3 for the streamline-refinanced loans is the period for which actual streamline loan experience was available and met our minimum credibility criteria. To improve the fit of our models, we supplemented this data with experience from the FR30, High LTV and the FR15, High LTV data sets – the available history for fixed rate loans runs from 1975 through 1999. The number of observations listed in Table A.3 reflects the combined FR and SRFR experience.

Computing Environment Uused for Data Manipulation and Regression Analysis

The vast majority of the data manipulation, scrubbing, and transformation was performed on a UNIX server using SAS. *Appendix E, Data Transformation* includes a printout of the SAS programs used in this phase of our analysis. We performed a small amount of additional data manipulation in Excel spreadsheets. All of our regression analyses were performed using S-Plus 2000, Professional edition. The regression results for each loan-type/LTV category are included as part of this appendix.

Weighted Averages

Since each observation in the data set we used for performing our regression analysis is comprised of the characteristics for a number of loans, we calculate a weighted average across a given cell for each of the various response and predictor variables. Each of our weighted averages is based on the amortized loan balance of the surviving loan pool during each experience period or policy year. This weighting scheme makes each of our predictor variables, with the exception of the binary variables, an implicitly time-dependent variable, if not an explicitly time-dependent variable.

Conditional Claim Rate Model

We developed ten separate conditional claim rate models. A separate model was developed for each of the four LTV categories[2] in the fixed rate, 30-year loan category. For the fixed rate, 15-year loan category, we developed three separate regression models: one for the High LTV category, one for the Low LTV category, and one for the combined Mid and Investor LTV category. The final three regression models are comprised of one model for adjustable rate loans, all LTV categories combined; one for streamline refinanced, fixed rate, 30-year loans (SRFR30), all LTV categories combined; and one for the streamline refinanced, fixed rate, 15-year loans (SRFR15), all LTV categories combined. (There are so few streamline refinanced, adjustable rate loans that it is not feasible to fit a model for that loan category. To estimate the number of claims and prepayments for this group of loans, we used the rates estimated by the model for non-streamline, adjustable rate loans.)

The basic structure of the conditional claim rate model is set forth in Equation A.1.

$$\begin{aligned}\ln(\lambda_t) = {} & \alpha + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot INT.RT_t + \beta_4 \cdot R.GT1_t + \beta_5 \cdot R.LT1_t + \beta_6 \cdot CUMDIFF_t \\ & + \beta_7 \cdot LTV_0 + \beta_8 \cdot LTV.AGE3_t + \beta_9 \cdot ANN.HPA_t + \beta_{10} \cdot HPA_t \\ & + \beta_{11} \cdot NEGEQ.RGT1_t + \beta_{12} \cdot NEGEQ.RLT1_t + \beta_{13} \cdot RHP + \beta_{14} \cdot UNEMP.L0_t \\ & + \beta_{15} \cdot PAY.INC.AGE4_t + \beta_{16} \cdot SR \quad \textbf{(A.1)}\end{aligned}$$

Presented below is a brief definition of each term in the conditional claim rate model. *Appendix E, Data Transformation* includes a more detailed description of the exact method used to calculate each of the predictor variables. Exhibits A.1 through A.10 and Graphs A.1 through A.30 included with this Appendix A provide results of each regression analysis, including the coefficient values, standard errors, t-statistics, null and residual deviance values, analysis of deviance tables, plots of actual versus fitted conditional claim rates by policy year, scatter plots of actual versus fitted claims, and scatter plots of the square root of the absolute value of the deviance residuals versus fitted values.

Definitions

1. $\ln(\lambda_t)$ = the natural log of the estimated Poisson parameter during policy year t for a given loan-type/LTV category. Since, for a Poisson distribution, the Poisson parameter is also the mean of the distribution, λ_t is the expected number of claims (per 10,000 contracts) in policy year t within a given loan type category. In effect, we are saying that, at each stage of a loan pool's life (where we define a stage to be a policy year), there exists a Poisson distribution that defines the conditional probability distribution of insurance claims. The Poisson

[2] The four LTV categories are High, Mid, Investor, and Low. A detailed description of these categories is provided in the *Data Transformation Appendix*, which we summarize here for the readers convenience. Low LTV is defined by LTV values less than 83%, Mid LTV is defined by LTV values between 87% and 96%, and the High LTV is defined by LTV values greater than 96%. The Investor LTV category overlaps slightly the Low and Mid LTV categories with respect to the actual LTV but is further defined by the number of living units (>1), the borrower type (landlord, builder, operative builder, escrow commitment or corporation), and LTV values less than 87%. The definition of LTV categories is constant across all loan types.

parameter provides the expected number of claims, but we could just as easily calculate the number of claims we would expect at the 75th or 25th percentiles for a given set of values for the predictor variables. This additional information is available since a Poisson distribution is completely specified by its mean.

2. t = age of the loan in years, or the policy year of a given loan cohort.
3. t^2 = age of the loan in years, squared.
4. $INT.RT_t$ = the weighted average loan contract rate.
5. $R.GT1_t$ = the weighted average refinance incentive ratio at a given age if the value of the refinance incentive ratio is greater than one; otherwise this variable takes a value of zero. The refinance incentive ratio is defined as the ratio of the loan contract rate to the available refinance rate at a given age.
6. $R.LT1_t$ = the weighted average refinance incentive ratio at a given age if the value of the refinance incentive ratio is less than or equal to one; otherwise this variable takes a value of zero.
7. $CUMDIFF_t$ = the weighted average cumulative positive difference between the loan contract rate and the historically available refinance rate through each age of a loan pool's life. See *Appendix E, Data Transformation* for a more detailed description of this variable and a graphical display of its value over time.
8. $LTV.0$ = the weighted average loan-to-value ratio that exists at the time of loan origination.
9. $LTV.AGE3_t$ = the weighted average, time-dependent loan-to-value ratio, beginning at age 3. The time-dependent loan-to-value ratios are updated at each policy year for scheduled amortization and for house price appreciation.
10. $ANN.HPA_t$ = the weighted average house price appreciation that has taken place during a given fiscal year, weighted for the geographical distribution of the particular loan pool.
11. HPA_t = the weighted average cumulative house price appreciation since the origination of the loan pool, weighted for the geographical distribution of the particular loan pool.
12. $NEGEQ.RGT1_t$ = the weighted average probability of negative equity at a given age if the value of the refinance incentive ratio is greater than or equal to 1.02; if the value of the refinance incentive ratio is less than 1.02, this variable takes on a value of zero.
13. $NEGEQ.RLT1_t$ = the weighted average probability of negative equity at a given age if the value of the refinance incentive ratio is less than 1.02; if the value of the refinance incentive ratio is greater than or equal to 1.02, this variable takes on a value of zero.
14. RHP = the weighted average relative house price at loan origination. Relative house price is calculated as the ratio of the property value associated with each loan relative to the corresponding MSA median house price. The median house price used in this instance is not an FHA/MMIF specific median; it is the median for the entire MSA.
15. $UNEMP.L0_t$ = the weighted average of state unemployment rates in the current fiscal period.
16. $PAY.INC.AGE4_t$ = the weighted average, time-dependent payment-to-income ratio as of a given age. The time-dependent payment-to-income ratios are updated at each policy period for changes in personal income levels; in addition for adjustable rate loans the ratio is updated for changes in loan payment levels that result from changes in the loan interest rate.
17. SR = a binary variable that indicates whether the observation is pre- or post-introduction of the Streamline Refinance program. The variable takes on a value of 1 if the fiscal/calendar period is 1989 or later, and zero otherwise.
18. $\boldsymbol{a}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, ..., \boldsymbol{b}_{16}$ = estimated regression coefficients for the Poisson regression model that result from applying an iteratively re-weighted least squares (IRWLS) methodology.

The regression models applied to the streamline refinanced loans have an additional explanatory variable, *REFI*. We introduced this binary variable when we augmented the streamline data with data from the corresponding fixed rate, high LTV category. *REFI* takes on a value of 1 if the observation is an actual streamline refinanced loan observation, and a value of zero otherwise. The negative sign on the regression coefficient for this variable indicates that frequency of claims for streamline refinanced loans is less than that of purchase origination loans. Based on our analysis, there is a relative difference of approximately 25% lower claim rates for streamline loans.

For adjustable rate loans, we modified equation A.1 slightly. We dropped the age-squared term and added two other age variable, one that takes on a value up to age 2, and zero thereafter, and a second that takes on a value of 1 up to age 4, and a value of zero thereafter. We found that this shift in the use of age variables improved the fit of our models to historical experience. In addition, we dropped the *CUMDIFF*, *HPA*, *NEGEQ.RLT1*, and *NEGEQ.RGT1* variables and introduced a *NEGEQ.AGE6* variable. The *NEGEQ.AGE6* variable takes on the value of the probability of negative equity for policy year 6 and later, and has a zero value otherwise. We examined a series of scatter plots showing conditional claim rates versus the probability of negative equity and observed that a clear relationship between the two variables did not emerge until policy year 6 and subsequent.

Conditional Prepayment Rate Model

We developed conditional prepayment rate models for the same loan-type/LTV categories for which conditional claim rate models were developed. That is, we developed ten prepayment models: four for the fixed rate, 30-year loan type, three for the fixed rate, 15-year loan type, and one for each of ARM, streamline refinanced, fixed rate, 30-year loans, and streamline refinanced, fixed rate, 15-year loans.

The basic structure of the conditional prepayment rate model is set forth in Equation A.2.

$$
\begin{aligned}
\ln(\,\boldsymbol{l}_t) = \; & \boldsymbol{a} + \boldsymbol{b}_1 \cdot INT.RT_t + \boldsymbol{b}_2 \cdot R.PRIME_t + \boldsymbol{b}_3 \cdot YIELD.DIFF_t + \boldsymbol{b}_4 \cdot FR30.PDIFF_t \\
& + \boldsymbol{b}_5 \cdot FR30.NDIFF_t + \boldsymbol{b}_6 \cdot TBOND.VOL_t + \boldsymbol{b}_7 \cdot NEGEQ.RGT1_t \\
& + \boldsymbol{b}_8 \cdot NEGEQ.RLT1_t + \boldsymbol{b}_9 \cdot CUMDIFF_t + \boldsymbol{b}_{10} \cdot RHP + \boldsymbol{b}_{11} \cdot ANN.HPA_t \\
& + \boldsymbol{b}_{12} \cdot LTV_0 + \boldsymbol{b}_{13} \cdot LTV.AGE3_t + \boldsymbol{b}_{14} \cdot PAY.INC.AGE4_t \\
& + \boldsymbol{b}_{15} \cdot UNEMP.L0_t + \boldsymbol{b}_{16} \cdot t + \boldsymbol{b}_{17} \cdot t.2 + \boldsymbol{b}_{18} \cdot SR \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textbf{(A.2)}
\end{aligned}
$$

Presented below is a brief definition of each term in the conditional prepayment rate model. *Appendix E, Data Transformation* includes a more detailed description of the exact method used to calculate each of the predictor variables. Exhibits A.11 through A.20 and Graphs A.31 through A.60 included with this Appendix A provide results of each regression analysis, including the coefficient values, standard errors, t-statistics, null and residual deviance values, analysis of deviance tables, plots of actual versus fitted conditional claim rates by policy year, scatter plots of actual versus fitted claims, and scatter plots of the square root of the absolute value of the deviance residuals versus fitted values.

Definitions

1. $\ln(\,\boldsymbol{l}_t) =$ the natural log of the estimated Poisson parameter at during policy year t for a given loan-type/LTV category. Since, for a Poisson distribution, the Poisson parameter is also the mean of the distribution, $\boldsymbol{l}_t$ is the expected number of prepayments (per 10,000 contracts) in policy year t within a given loan-type/LTV category. In effect, we are saying that, at each stage of a loan pool's life (where we define a stage to be a policy year), there exists a Poisson distribution that defines the conditional probability distribution of prepayments. The Poisson parameter provides the expected number of prepayments, but we could just as easily calculate the number of prepayments we would expect at the 75^{th} or 25^{th} percentiles for a given set of values for the predictor variables. This additional information is available since a Poisson distribution is completely specified by its mean.
2. $INT.RT_t$ = the weighted average loan contract rate.
3. $R.PRIME_t$ = the weighted average, exponentially weighted moving average[3] refinance incentive ratio at age t.

---

[3] $R'_t = z \cdot \overline{R}_t + (1 - z) \cdot R'_{t-1}$, where $\overline{R}_t =$ the arithmetic mean of prior refinance ratios up to time t, and z = the weight assigned to the mean of prior ratios. For this Review, we selected $z = 0.75$.

4. $YIELD.DIFF_t$ = is the difference between the yield on 30-year U.S. Treasury bonds and the yield on 52-week U.S. Treasury bills.
5. $FR30.PDIFF_t$ = the difference between the weighted average loan contract rate and the available contract rate on a fixed rate, 30-year mortgage. If the loan contract rate is higher than the available refinance rate this variable reflects that difference; otherwise it has a value of zero.
6. $FR30.NDIFF_t$ = the difference between the weighted average loan contract rate and the available contract rate on a fixed rate, 30-year mortgage. If the loan contract rate is lower than the available refinance rate this variable reflects that signed difference; otherwise it has a value of zero.
7. $TBOND.VOL_t$ = the annual volatility of the 30-year U.S. Treasury bond.
8. $NEGEQ.RGT1_t$ = the weighted average probability of negative equity at a given age if the value of the refinance incentive ratio is greater than or equal to 1.02; if the value of the refinance incentive ratio is less than 1.02, this variable takes on a value of zero.
9. $NEGEQ.RLT1_t$ = the weighted average probability of negative equity at a given age if the value of the refinance incentive ratio is less than 1.02; if the value of the refinance incentive ratio is greater than or equal to 1.02, this variable takes on a value of zero.
10. $CUMDIFF_t$ = the weighted average cumulative positive difference between the loan contract rate and the historically available refinance rate through each age of a loan pool's life. See the *Appendix E, Data Transformation* for a more detailed description of this variable and a graphical display of its value over time.
11. *RHP* = the weighted average relative house price at loan origination. Relative house price is calculated as the ratio of the property value associated with each loan relative to the corresponding MSA median house price. The median house price used in this instance is not an FHA/MMIF specific median; it is the median for the entire MSA.
12. $ANN.HPA_t$ = the weighted average house price appreciation that has taken place during a given fiscal year, weighted for the geographical distribution of the particular loan pool.
13. *LTV.0* = the weighted average loan-to-value ratio that exists at the time of loan origination.
14. $LTV.AGE3_t$ = the weighted average, time-dependent loan-to-value ratio, beginning at age 3. The time-dependent loan-to-value ratios are updated at each policy year for scheduled amortization and for house price appreciation.
15. $PAY.INC.AGE4_t$ = the weighted average, time-dependent payment-to-income ratio as of a given age. The time-dependent payment-to-income ratios are updated at each policy period for changes in personal income levels; in addition for adjustable rate loans the ratio is updated for changes in loan payment levels that result from changes in the loan interest rate.
16. $UNEMP.L0_t$ = the weighted average unemployment rate in the current fiscal period.
17. *t* = age of the loan in years, or the policy year of a given loan cohort.
18. *t.2* = a binary variable that takes on the value 1 if the policy year is less than or equal to 2, and zero otherwise.
19. *SR* = a binary variable that indicates whether the observation is pre- or post-introduction of the Streamline Refinance program. The variable takes on a value of 1 if the fiscal/calendar period is 1989 or later, and zero otherwise.
20. $\boldsymbol{a}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, ..., \boldsymbol{b}_{18}$ = estimated regression coefficients for the Poisson regression model that result from applying an iteratively re-weighted least squares (IRWLS) methodology.

The regression models applied to the streamline refinanced loans have an additional explanatory variable, *REFI*. We introduced this binary variable when we augmented the streamline data with data from the corresponding fixed rate, high LTV category. *REFI* takes on a value of 1 if the observation is an actual streamline refinanced loan observation, and a value of zero otherwise.

For adjustable rate loans, we modified equation A.2 slightly. We dropped the *CUMDIFF*, and *ANN.HPA* variables. Originally we included these variables in our ARM prepayment model but found that they added little in terms of explanatory power based on a review of the analysis of deviance table.

Explanatory Variables

The explanatory variables used in the claim and prepayment models can be categorized into several groups: interest rate variables, income variables, equity variables, and baseline hazard. In the sections that follow we provide an explanation for these variables and provide some economic insight behind the signs on the corresponding regression coefficients in the claim and prepayment equations. Following this section is a bibliography of reference sources for both generalized linear models and the loan termination models.

Interest Rate Variables

The contract rate on a loan is a primary determinant of the borrower's payment burden. In general, lower contract rates increase the demand for "new loans" in two ways. First, there is the refinancing effect where current mortgage holders choose to take a new loan with a lower contract rate and repay their current loans. The second effect is the origination of new loans by borrowers who are attracted by the lower monthly mortgage payments that result from the lower loan contract rate.

We use several variables to capture the effect of interest rates on prepayment options and payment burden levels. First is the refinance incentive ratio that is defined as the ratio of the loan contract rate to the currently available refinance rate. When the ratio takes on a value greater than 1.0, there is an economic incentive for the borrower to refinance since the borrower can reduce their monthly payment by refinancing at the lower rate. The refinance incentive ratio is an approximation of the ratio of the present value of an annuity at the available refinance rate to the present value of an annuity at the loan contract rate. At most ages, the refinance incentive ratio provides a reasonable approximation to this ratio of present values; only at very late stages of a loan's life does this approximation begin to break down; see Richard and Roll (1995). Our refinance incentive ratio is similar to the prepayment option, or POPTION, covariate used in other loan termination models.

The refinance incentive ratio provides information regarding when there is an economic advantage to refinancing a loan, but it does not provide information about the recent history of refinance options. For this purpose, we use an exponentially weighted moving average refinance incentive ratio. This variable is a weighted average of the refinance incentive ratios that have existed since loan origination where the most recent ratio receives the highest weight and prior ratios receive a weight that decays at an exponential rate.

Another variable that keeps track of the refinance options over the life of a loan is the CUMDIFF variable. This is the sum of the cumulative positive differences between the loan contract rate and the available refinance rate. In the prepayment rate model, CUMDIFF provides information regarding the level of burnout for a pool of loans – that is, the change in sensitivity to economically advantageous refinance options. In the claim rate model, CUMDIFF is intended to provide insight regarding adverse selection of remaining loans. The theory is that, given the fact that an increase in the value of the CUMDIFF variable represents a history of favorable refinance opportunities, the loans remaining in a pool characterized by a high CUMDIFF variable will be those that were unable to refinance and are more likely to result in a claim. Given this, the CUNMDIFF regression coefficient should have a positive sign in the claim model. For some of our models the results meet this expectation. However, for the fixed rate, 30-year loan category this is not the case; the sign on the CUMDIFF coefficient is negative. For the most part, we believe this is due to the characteristics of the surviving loans and the relationship of the interest rate to the loans that have already refinanced. We are investigating this further and will address it in more detail in subsequent reports.

FR30.PDIFF and FR30.NDIFF are time-dependent variables defined as the difference between the loan contract rate and the available refinance rate. The PDIFF or NDIFF suffixes discriminate between cases where the variables take on a positive or negative value. We expect the regression coefficients to have a positive sign in the prepayment model. The covariates are not used in the claim model.

YIELD.DIFF is a time dependent variable that is the difference between the short- and long-end of the U.S. Treasury yield curve. Specifically, it is the difference between the rate on the 30-year Treasury bond and the 52-week Treasury bill. The variable attempts to predict the direction of future interest rates. It is similar in its basic structure to the YLDCURVE used in the Abt Associates Microsimulation model[4] and the YIELDCUR variable used in the GAO model[5]. There are, however, differences in its specification that result in the expectation that our YIELD.DIFF has a positive sign in the prepayment model whereas YLDCURVE and YIELDCUR are expected to have a negative sign. The difference is that both Abt and GAO specify their variables as the difference between the yield on 10-year Treasury bond and a 52-week treasury bill less 250 basis points. Both Abt and GAO further constrain their variable to only positive values. YIELD.DIFF is not used in the claim rate model.

INT.RT is the loan contract rate. This time-dependent covariate is used in both the claim and prepayment rate models. In the claim rate model, we expect the regression coefficient to have a positive sign under the belief that, all else being equal, a higher monthly loan payment driven by a higher contract rate will increase the chance of a claim. In a prepayment model that used the loan contract rate as a predictor variable but did not make use of the refinance incentive ratio or FR30.PDIFF and FR30.NDIFF, we would expect the regression coefficient of INT.RT to have a positive sign. Since our model makes use of each of these variables, we believe that INT.RT is picking up a variety of effects including policy level changes.

---

[4] Abt Associates Inc., 1998, Report of the Loan Termination Models for the Microsimulation Model of FHA's Mutual Mortgage Insurance Fund.

[5] General Accounting Office, 1996, Mortgage Financing: FHA Has Achieved Its Home Mortgage Capital Reserve Target, GAO/RCED-96-50.

Income Variables
PAY.INC.AGE4 is the ratio of the monthly loan payment to borrower income – see Campbell and Dietrich, 1983. It is a time-dependent variable that is adjusted for changes in income levels over the life of a loan. It is used in both the claim and prepayment rate models. We expect the sign of the coefficient to be positive in both the claim and prepayment rate models. The expectation of the sign for the claim rate model is obvious; a higher ratio of mortgage payment to income increases the chance of a claim. For the prepayment rate model, the expectation of a positive sign is driven by a borrower's desire to lower their monthly payment burden with a lower contract rate loan.

UNEMP.L0 is a contemporaneous countrywide unemployment statistic. Since higher levels of unemployment decrease household income levels, we expect the sign on this coefficient to be positive in the claim rate models and negative in the prepayment rate models. Our results for the prepayment models are consistent with expectations; the claim rate model results are, however, counter-intuitive. While we believe there is good economic rationale for including unemployment as a covariate in the claim rate model, we considered removing it due to the resulting sign. In the end, we decided to retain this variable and to look for a better specification (based on geographic distribution) in future models and further investigate the current results. In addition, we constrained the unemployment variable in our alternative economic scenarios by holding them at baseline levels.

Equity Variables
Based on option pricing theory, a mortgage can be viewed as a set of options held by the borrower. First, it is a call option in the sense that the borrower can call in their loan at any time by prepaying the outstanding balance; and second, it is a put option in the sense that the borrower can force their creditor to purchase their property at current market prices by defaulting. The call, or prepayment, option is reflected in our refinance incentive ratio covariate; the put, or default, option is reflected in our NEGEQ.RGT1 and NEGEQ.RLT1 covariates.

The rationale behind the default options is that each month, or on some periodic basis, the homeowner compares the market value of their property with the outstanding on their loan. The difference between the two is the amount of equity in the property. In general, as a loan ages we expect the level of equity in a property to increase for two reasons: first the loan balance is paid down over time resulting in higher equity, and second the market value of the property can increase over time. It is, however, also possible for the market value of the property to decline, which under certain circumstances can result in negative equity. The likelihood of a claim is increased with the probability of negative equity for a given property. We calculate the probability of negative equity based on the following formula:

$$\Pr(NEG.EQUITY) = \Phi\left(\frac{\ln(pv.bal) - \ln(mkt.val)}{s}\right)$$

where $pv.bal$ is the present value of the remaining mortgage payments, discounted at the loan contract rate, and $mkt.val$ is the current market value of the house estimated using the OFHEO repeat sales index. The value in the denominator, s, is the volatility of the house price index that is estimated since loan origination. The function, Φ, is the standard normal distributiuon. For further details see Ambrose and Capone (1997), Cooperstein, Redburn, and Meyers (1991), or Deng (1995), Deng, Quigley, and Van Order (1994). For a specification that uses the negative equity variable split based upon the value of the refinance incentive ratio, or a similar variable, see Mattey and Wallace (1999).

We expect the sign on NEQEQ.RGT1 and NEG.RLT1 to be positive in the claim rate model and negative in the prepayment rate model.

We use two separate loan-to-value ratios in our models, the original LTV, LTV.0, and a time-dependent LTV that phases in at policy year 3, LTV.AGE3$_t$. LTV.0 specifies the initial equity position of the borrower; LTV.AGE3 tracks increases in house prices and scheduled amortization. For the claim rate model, we expect the coefficient on both variables to be positive; for the prepayment rate model, we expect that higher LTV loans are less likely to prepay which would result in a negative coefficient.

HPA and ANN.HPA reflect cumulative and contemporaneous changes in property values, respectively. We expect the signs on the coefficients for these covariates to be positive in the prepayment rate models - higher levels of appreciation result in greater overall wealth and loan prepayment due to an upgrade in housing – and negative for the claim rate models – an increase in the market value of a house results in a lower probability of a loan resulting in a claim.

Baseline Hazards

The amount of time that has elapsed since loan origination is a determinant of claim and prepayment behavior that underlies most other variables. At loan origination, borrowers begin with reasonably stable income, equity, and wealth levels as a result of the underwriting process. As a loan pool ages, various events take place that can change a borrower's ability to make timely loan payments, for example: job loss, change in family status, and house price depreciation. These events can make default more likely. After several years, a given loan will have built up significant equity as a result of house price appreciation and scheduled amortization – characteristics that decrease the likelihood of default. Given this, we expect baseline claim activity to peak in the first three to six years of the life of a loan pool and to steadily decrease beyond that point.

The baseline pattern for prepayments is similar but may be delayed for several years due to transaction costs associated with loan origination.

Bibliography

Generalized Linear Model and Statistics References

Christensen, R. 1997. *Log-Linear Models and Logistic Regression,* 2nd Edition. Springer-Verlag. Chapter 9, Generalized Linear Models.

Huber, P.J. 1973. Robust Regression: Asymptotics, Conjectures, and Monte Carlo. *Annals of Statistics* 1: 799-821.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*. Second Edition. London: Chapman and Hall.

Kao, E. 1997. *An Introduction to Stochastic Processes*, Duxbury Press. Chapter 2.

Ross, S. 1997. *Introduction to Probability Models*, 6th Edition, Academic Press. Section 5.3 and 5.4.

Stokes and Koch, 1983; *A Macro for Maximum Likelihood Fitting of Log-Linear Models to Poisson and Multinomial Counts*; Proceedings of the Eighth Annual SAS Users Group International; Cary, North Carolina: SAS Institute, pp. 795-800.

Real Estate and Real Estate Economics References

Ambrose, B., C. Capone, *Modeling the Conditional Probability of Foreclosure in the Context of Single-Family Mortgage Default Resolutions*, Real Estate Economics, Vol. 26, No. 3, 391-429, Fall 1998.

Calhoun, C., *OFHEO House Price Indexes: HPI Technical Description,* Office of Federal Housing Enterprise Oversight, March 1996.

Campbell, T.S. and K.J. Dietrich. 1983. *The Determinants of Default on Insured Conventional Residential Mortgage Loans.* Journal of Finance 38(5): 1569 – 1581.

Cooperstein, R.L., F.S. Redburn, and H.G. Meyers, *Modeling Mortgage Terminations in Turbulent Times,* AREUEA Journal, 19(4): 473-494, Winter 1991.

Deng, Y., J. Quigley, and R. Van Order, *Mortgage Default and Low Downpayment Loans: The Costs of Public Subsidy,* Regional Science and Urban Economics, 26 (1996) 263-285.

Deng, Y., and C. Calhoun, *A Dynamic Analysis of Adjustable- and Fixed-Rate Mortgage Termination,* Office of Federal Housing Enterprise Oversight, December 1996.

Deng, Y., *Mortgage Termination: An Empirical Hazard Model with a Stochastic Term Structure,* Journal of Real Estate Finance and Economics, 14:3, 309-331 (1997).

Deng Y., J. Quigley, and R. Van Order, *Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options,* Econometrica, 1998.

Fabozzi, Ramsey, and Ramirez. 1994. *Collateralized Mortgage Obligations: Structures and Analysis*, 2nd Edition. Frank J. Fabozzi and Associates. Chapter 3, pp. 48-55.

Fabozzi, F. 1998. *The Handbook of Structured Financial Products*, Frank J. Fabozzi and Associates. Chapter 22, Mortgage Backed Securities, pp. 337 – 340.

Foster, C. and R. Van Order, *An Option-Based Model of Mortgage Default,* Housing Finance Review, 3(4):351-77, October 1984.

Kau, J., T. Springer, *The Prepayment of Fixed Rate Mortgages Underlying Mortgage Backed Securities: A Switching Regimes Analysis,* Journal of Housing Economics, 2, 257 - 275, 1992.

Mattey, J., N. Wallace, *Housing Prices and the Instability of Mortgage Prepayment Models: Evidence from California,* Federal Reserve Bank of San Francisco, Working Paper 98-05, July 1998.

Mattey, J., N. Wallace, *Housing Price Cycles and the Prepayment Rates of U.S. Mortgage Pools,* Federal Reserve Bank of San Francisco, August 1999.

Peristiani, S., P. Bennett, G. Monsen, R. Peach, J. Raiff, *Effects of Household Creditworthiness on Mortgage Refinancings,* Federal Reserve Bank of New York, Research Paper No. 9622, August 1996.

Peristiani, S., P. Bennett, G. Monsen, R. Peach, J. Raiff, *Credit, Equity, and Mortgage Refinancings,* Federal Reserve Bank of New York Economic Policy Review, 83-99, July 1997.

Quigley, J. and R. Van Order, *More on the Efficiency of the Market for Single Family Homes: Default,* University of California at Berkley.

Richard, Scott and Roll, Richard. Goldman Sachs Mortgage Securities Research report titled *Modeling Prepayments on Fixed Rate Mortgage-Backed Securities*. September 1988. Fabozzi, F.J. (editor). 1995.

Schwartz, E.S., W.N. Torous, *Prepayment and the Valuation of Mortgage Backed Securities,* Journal of Finance, 44 (2), 375-92.

Stanton, R., *Rational Prepayment and the Valuation of Mortgage Backed Securities*, The Review of Financial Studies, Volume 8, No. 3, 677 - 708, Fall 1995.

The Handbook of Mortgage Backed Securities. McGraw-Hill. Section II - Prepayment Behavior and Forecasting, Chapters 8, 9, 11, 12, and 13.