Appendix E - Data Transformation

HUD provided Deloitte & Touche with data from the Single Family Data Warehouse for fiscal endorsement years 1975 through 2001 as of March 31, 2001. The following summarizes the process of summarizing the data and preparing the data sets for analysis.

Initial Record Drop Criteria

Our first step in sorting through the data was to take out any files that did not have an original loan amount (orig_mrtg_amt = 0) or a contract rate (int_rt = 0). The following table summarizes the results of this process.

Table E.1

Fiscal Origination Year	Original Number of Loans in Database	Total Initial Drop	Number Remaining Loans After Initial Drop	Percent of Total Original Loans
1975	185,974	35	185,939	0.019%
1976	222,104	50	222,054	0.023%
1977	256,162	79	256,083	0.031%
1978	294,604	147	294,457	0.050%
1979	389,811	852	388,959	0.219%
1980	337,148	539	336,609	0.160%
1981	216,297	245	216,052	0.113%
1982	149,178	7,616	141,562	5.105%
1983	506,098	114	505,984	0.023%
1984	287,189	19	287,170	0.007%
1985	400,632	11	400,621	0.003%
1986	929,000	29	928,971	0.003%
1987	1,126,841	40	1,126,801	0.004%
1988	615,822	148	615,674	0.024%
1989	634,640	107	634,533	0.017%
1990	715,752	50	715,702	0.007%
1991	643,554	57	643,497	0.009%
1992	637,131	1	637,130	0.000%
1993	990,899	0	990,899	0.000%
1994	1,059,778	12	1,059,766	0.001%
1995	521,238	145	521,093	0.028%
1996	728,051	50	728,001	0.007%
1997	740,188	59	740,129	0.008%
1998	955,317	44	955,273	0.005%
1999	1,121,041	0	1,121,041	0.000%
2000	826,491	0	826,491	0.000%
2001	411,910	0	411,910	0.000%
Total	**15,902,850**	**10,449**	**15,892,401**	**0.066%**

Identifying Loan Types

We split the database into six different loan types:

1. Fixed rate 30-year (FX30)
2. Fixed rate 15-year (FX15)
3. Adjustable rate (ARM)
4. Streamline refinance 30-year (SRFX30)
5. Streamline refinance 15-year (SRFX15)
6. Adjustable rate streamline refinance (SRARM)

We identified Streamline Refinanced (SR) loans in fiscal origination years 1988 through 2000 according to three criteria:

1. A refinance code (rfnc_cd) of “H”, “R”, or “S”
2. A streamline flag (pd_strmln_flg) of “R”, or
3. A loan-to-value ratio (ratio_loan_to_vl) coded as 30 or 999 (as opposed to our calculated value of LTV).

We used the adjustable rate indicator and the 15-year term indicator in the Data Warehouse to further classify the loans.

Geography

There are some geographic areas covered by the MMIF but for which some of the external economic information was unavailable. These are, specifically: Puerto Rico, the Virgin Islands, and Guam. Since we did not have complete information about these areas, we had to make simplifying assumptions. Given the small size of this subset of the database (see table below), we believe the assumptions to have an immaterial effect on our results.

We used economic information about Florida as a proxy for information about Puerto Rico. We excluded Virgin Island and Guam records from the regression analysis.

Table E.2

Fiscal Origination Year	Number of Records in Analysis	Virgin Islands, Guam	Virgin Islands, Guam Percentage	Number of Records Remaining in Analysis
1975	185,939	436	0.234%	185,503
1976	222,054	174	0.078%	221,880
1977	256,083	214	0.084%	255,869
1978	294,457	168	0.057%	294,289
1979	388,959	55	0.014%	388,904
1980	336,609	26	0.008%	336,583
1981	216,052	2	0.001%	216,050
1982	141,562	69	0.049%	141,493
1982	505,984	114	0.023%	505,870
1983	287,170	111	0.039%	287,059
1984	400,621	42	0.010%	400,579
1985	928,971	31	0.003%	928,940
1986	1,126,801	43	0.004%	1,126,758
1987	615,674	27	0.004%	615,647
1988	634,533	27	0.004%	634,506
1989	715,702	50	0.007%	715,652
1990	643,497	28	0.004%	643,469
1991	637,130	64	0.010%	637,066
1992	990,899	82	0.008%	990,817
1993	1,059,766	62	0.006%	1,059,704
1994	521,093	25	0.005%	521,068
1995	728,001	34	0.005%	727,967
1996	740,129	65	0.009%	740,064
1997	955,273	50	0.005%	955,223
1998	1,121,041	41	0.004%	1,121,000
1999	826,491	23	0.003%	826,468
2000	411,910	7	0.002%	411,903
Total	**15,892,401**	**2,070**	**0.013%**	**15,890,331**

Loan-to-Value Ratio Calculation

In general, the initial loan-to-value ratio, LTV_0, is calculated using the following formula:

$$\frac{orig\_mrtg\_amt - ufmip\_pd\_amt}{\min(prprty\_aprsl\_vl, prc\_excl\_clsng\_amt)}$$

1. **If both prprty_aprsl_vl and prc_excl_clsng_amt are available**, the LTV_0 ratio is estimated based on the above formula.
2. **If one of prprty_aprsl_vl or prc_excl_clsng_amt is not available**, the LTV_0 ratio's denominator takes the value of the available variable.
3. **If both "previous" prprty_aprsl_vl and "previous" prc_excl_clsng_amt are unavailable**, then we use the ratio_loan_to_vl field in the database.
4. **If ratio_loan_to_vl is unavailable**, then the loan record is excluded from the regression analysis for lack of sufficient information.

Streamline Refinanced Loans

Matching to Original Loan

Because Streamline Refinancing doesn't require an appraisal, we needed to estimate LTV_0 for those loans. We did this by attempting to match each SR loan to the refinanced or "previous" loan. We searched all loans prior to each SR loan for a loan where the refinance case number field (rfnc_cs_nbr) matched the case number of the SR loan.

We were able to match roughly 85% of the SR loans to their "previous" loans. The success rate varied by fiscal origination year as shown in the table below.

Table E.3

Origination Year	Streamline Refinancings	Unmatched	Total Streamline Refinancings Remaining	Percent Unmatched
1988	21,547	20,186	1,361	94%
1989	13,499	8,741	4,758	65%
1990	25,255	10,675	14,580	42%
1991	29,084	10,604	18,480	36%
1992	97,391	21,640	75,751	22%
1993	421,382	42,655	378,727	10%
1994	458,554	76,995	381,559	17%
1995	28,132	14,843	13,289	53%
1996	102,864	22,546	80,318	22%
1997	56,181	11,314	44,867	20%
1998	212,737	28,877	183,860	14%
1999	259,872	33,180	226,692	13%
2000	31,671	7,875	23,796	25%
2001	66,635	4,712	61,923	7%
Total	**1,824,804**	**314,843**	**1,509,961**	**17%**

If we could not match an SR loan to an earlier loan record, we dropped the SR from our regression analysis. Note that, if the "previous" loan had already been dropped from the analysis for lack of sufficient information to calculate LTV_0, then the corresponding SR loan is included in the count of "unmatched" loans.

Estimation of Property Value
Depending on the data available from the "previous" loan, we can estimate the property value of the SR loan based on one of the following scenarios:

1. **If both "previous" prprty_aprsl_vl and "previous" prc_excl_clsng_amt are available**, the SR loan's property value is estimated as the minimum of these two values adjusted by the ratio_loan_to_vl. (Note: If the previous ufmip_pd_amt is unavailable in this scenario, we adjust the estimated property value by an upfront premium factor, based on the upfront premium table shown in Appendix D - The Cash Flow Model).
2. **If exactly one of "previous" prprty_aprsl_vl or "previous" prc_excl_clsng_amt is not available**, the other is assigned as the SR loan's estimated property value.
3. **If both "previous" prprty_aprsl_vl and "previous" prc_excl_clsng_amt are unavailable,** we use the ratio_loan_to_vl field in the database.
4. **If ratio_loan_to_vl is unavailable,** then the SR loan is discarded for lack of sufficient information.

Note that the three scenarios parallel the LTV_0 calculation described in the description of the LTV_0 calculation in the previous section.

Once we have estimated the property value based on the available information from the "previous" loan, it is then adjusted by a house price appreciation factor. These factors were derived from the house price index (HPI) published by OFHEO by MSA, by state and by census division.

Payment to Income Fix Subroutine

Analyzing the payment to income ratio in the database (ratio_tmp_tei), we have found that a number of records contain a value of zero in this field. We also found other instances in which values in this field were greater than 75%. Therefore, we replaced these values with a reasonable estimate for the ratio, loan by loan. **For each loan type and each fiscal year**, we followed three simple steps to fix the records containing zero values or values greater than 75% in this field:

1. Find all the loans where the ratio_tmp_tei field contains a non-zero value or value less than 75% (judgmentally selected).
2. Calculate a weighted average of ratio_tmp_tei using the non-zero ratios determined in item1 with weights based on the corresponding orig_mrtg_amt.
3. Replace the zero values for ratio_tmp_tei with this weighted average ratio.

The table below shows the calculated average payment-to-income ratio by year and by loan type.

Table E.4

	Average Payment-to-Income Ratio (%)					
Fiscal Origination Year	Fixed Rate, 30-year Loans	Fixed Rate, 15-year Loans	Adjustable Rate Loans	Streamline Fixed Rate, 30-year Loans	Streamline Fixed Rate, 15-year Loans	Streamline Adjustable Rate Loans
1975	20.1548	17.2334	N/A	N/A	N/A	N/A
1976	20.3953	17.2851	N/A	N/A	N/A	N/A
1977	20.2352	16.8393	N/A	N/A	N/A	N/A
1978	21.6138	17.0934	N/A	N/A	N/A	N/A
1979	22.2554	17.1006	N/A	N/A	N/A	N/A
1980	23.3971	18.4563	N/A	N/A	N/A	N/A
1981	24.5346	19.3445	N/A	N/A	N/A	N/A
1982	24.7654	20.6457	N/A	N/A	N/A	N/A
1983	23.4542	22.9785	N/A	N/A	N/A	N/A
1984	24.2275	22.8989	N/A	N/A	N/A	N/A
1985	23.3636	22.8864	22.8387	N/A	N/A	N/A
1986	21.4749	20.4657	21.9264	N/A	N/A	N/A
1987	21.3479	19.8098	21.5248	N/A	N/A	N/A
1988	23.3566	22.4263	23.0647	19.5380	N/A	N/A
1989	25.3315	23.4140	25.4962	25.6982	24.6312	N/A
1990	23.7724	21.7151	23.2271	28.5424	21.0997	N/A
1991	22.9523	20.9496	23.8780	26.3305	23.6200	20.1664
1992	22.7247	20.1324	23.4370	24.0239	22.5823	22.3332
1993	22.4546	19.5859	23.6798	24.0792	21.7586	23.7337
1994	22.8249	19.3739	24.1877	21.5892	20.9847	21.4944
1995	23.9900	20.1626	24.8931	24.5135	21.8920	24.4035
1996	24.0293	20.5510	24.9601	25.1386	21.8197	24.4836
1997	24.3602	21.0897	24.9672	26.6138	22.8645	25.6561
1998	24.2735	21.2361	25.0560	29.9808	22.6867	27.9839
1999	25.0323	21.9598	26.1946	25.6196	21.9082	27.3653
2000	26.9263	23.7122	27.3840	28.5846	25.2308	27.4837
2001	26.8985	24.1555	28.2529	28.5694	25.5207	28.5915

Reasonable Range of LTV_0

We further attempted to remove erroneous records from the data set for regression analysis by checking the calculated LTV_0. We excluded any loan where LTV_0 was less than or equal to 10%, and any loan where LTV_0 was greater than or equal to 140%. The results of this step are summarized for fixed rate, 30-year loans in the table below.

Table E.5

Origination Year	Number of Loans, All Loan Types	LTV 10% or Less	LTV 140% or Greater	Remaining Loans	Percent Excluded
1975	185,503	26,500	432	158,571	15%
1976	221,880	28,291	663	192,926	13%
1977	255,869	24,179	950	230,740	10%
1978	294,289	41,279	1625	251,385	15%
1979	388,904	67,199	1855	319,850	18%
1980	336,583	36,790	2135	297,658	12%
1981	216,050	47,198	1605	167,247	23%
1982	141,493	20,766	724	120,003	15%
1983	505,870	88,609	1044	416,217	18%
1984	287,059	8,015	599	278,445	3%
1985	400,579	4,433	7360	388,786	3%
1986	928,940	5,090	4102	919,748	1%
1987	1,126,758	2,420	2814	1,121,524	0%
1988	595,461	303	2573	592,585	0%
1989	625,765	1,557	1540	622,668	0%
1990	704,977	198	2318	702,461	0%
1991	632,865	5,838	1396	625,631	1%
1992	615,426	3,768	4893	606,765	1%
1993	948,162	47	5842	942,273	1%
1994	982,709	31	5850	976,828	1%
1995	506,225	19	3949	502,257	1%
1996	705,421	12	4991	700,418	1%
1997	728,750	7	5619	723,124	1%
1998	926,346	25	6708	919,613	1%
1999	1,087,820	8	7039	1,080,773	1%
2000	818,593	2	3359	815,232	0%
2001	407,191	4	1126	406,061	0%
Total	**15,575,488**	**412,588**	**83,111**	**15,079,789**	**3%**

Relative House Price

HUD provided us with median house prices (MHP) through 1997 for some MSAs, and for all states. We estimated MHPs for 1998-2001 based on changes in HPI.

We calculated the relative house price (*RHP*) for a given loan to be consistent with our calculation of LTV_0. For each loan,

$$RHP = \frac{orig\_mrtg\_amt - ufmip\_pd\_amt}{LTV_0} \cdot \frac{1}{MHP}.$$

This guarantees that the "price" used in the RHP calculation for each loan was the same as the property value used to calculate the loan-to-value ratio. We used the MHP by MSA where it was available; otherwise we used MHP by state.

RHP and LTV Categories

Table E.6

LTV Range		Percentage of Loans in Range	Cumulative Percentage
10%	15%	0.0078%	0.0078%
15%	20%	0.0061%	0.0139%
20%	25%	0.0128%	0.0267%
25%	30%	0.0247%	0.0513%
30%	35%	0.0925%	0.1438%
35%	40%	0.0844%	0.2283%
40%	45%	0.1250%	0.3533%
45%	50%	0.1903%	0.5436%
50%	55%	0.2830%	0.8266%
55%	60%	0.3936%	1.2203%
60%	65%	0.5797%	1.8000%
65%	70%	0.8879%	2.6879%
70%	75%	1.5416%	4.2295%
75%	80%	2.6213%	6.8508%
80%	85%	5.2054%	12.0563%
85%	90%	8.8885%	20.9448%
90%	91%	2.0227%	22.9675%
91%	92%	2.6398%	25.6073%
92%	93%	3.2579%	28.8653%
93%	94%	4.1592%	33.0245%
94%	95%	6.3659%	39.3905%
95%	96%	11.6736%	51.0641%
96%	97%	19.7194%	70.7835%
97%	98%	18.0089%	88.7923%
98%	99%	5.9161%	94.7084%
99%	100%	3.0115%	97.7199%
100%	101%	0.6773%	98.3972%
101%	102%	0.3533%	98.7505%
102%	103%	0.2192%	98.9697%
103%	104%	0.1303%	99.1000%
104%	105%	0.0491%	99.1491%
105%	110%	0.2903%	99.4394%
110%	115%	0.1584%	99.5978%
115%	120%	0.1207%	99.7185%
120%	125%	0.0947%	99.8131%
125%	130%	0.1869%	100.0000%

Table E.7

RHP Range		Percentage of Loans in Range	Cumulative Percentage	RHP Range		Percentage of Loans in Range	Cumulative Percentage
0%	10%	0.0015%	0.0015%	96%	97%	1.1239%	63.1626%
10%	20%	0.0791%	0.0805%	97%	98%	1.1022%	64.2648%
20%	30%	0.8346%	0.9151%	98%	99%	1.0933%	65.3581%
30%	40%	2.8479%	3.7630%	99%	100%	1.0884%	66.4465%
40%	50%	5.6966%	9.4597%	100%	101%	1.0506%	67.4971%
50%	60%	9.2682%	18.7279%	101%	102%	1.0196%	68.5167%
60%	61%	1.0334%	19.7612%	102%	103%	1.0224%	69.5391%
61%	62%	1.0726%	20.8339%	103%	104%	0.9817%	70.5208%
62%	63%	1.0784%	21.9123%	104%	105%	0.9957%	71.5165%
63%	64%	1.1180%	23.0303%	105%	106%	0.9678%	72.4843%
64%	65%	1.1234%	24.1537%	106%	107%	0.9240%	73.4083%
65%	66%	1.1331%	25.2868%	107%	108%	0.9148%	74.3231%
66%	67%	1.1687%	26.4555%	108%	109%	0.8887%	75.2118%
67%	68%	1.1490%	27.6045%	109%	110%	0.8616%	76.0735%
68%	69%	1.2114%	28.8159%	110%	111%	0.8620%	76.9355%
69%	70%	1.2010%	30.0169%	111%	112%	0.8241%	77.7596%
70%	71%	1.1989%	31.2158%	112%	113%	0.8080%	78.5676%
71%	72%	1.2207%	32.4364%	113%	114%	0.7686%	79.3362%
72%	73%	1.2185%	33.6549%	114%	115%	0.7549%	80.0910%
73%	74%	1.2611%	34.9160%	115%	116%	0.7397%	80.8307%
74%	75%	1.2522%	36.1682%	116%	117%	0.6989%	81.5296%
75%	76%	1.2603%	37.4286%	117%	118%	0.7089%	82.2385%
76%	77%	1.2539%	38.6825%	118%	119%	0.6649%	82.9034%
77%	78%	1.2656%	39.9481%	119%	120%	0.3250%	83.2285%
78%	79%	1.2889%	41.2370%	120%	130%	5.4808%	88.7093%
79%	80%	1.2903%	42.5273%	130%	140%	3.7456%	92.4549%
80%	81%	1.2479%	43.7752%	140%	150%	2.5212%	94.9761%
81%	82%	1.2572%	45.0324%	150%	160%	1.6883%	96.6644%
82%	83%	1.2820%	46.3144%	160%	170%	1.0966%	97.7609%
83%	84%	1.2648%	47.5792%	170%	180%	0.7106%	98.4715%
84%	85%	1.2476%	48.8268%	180%	190%	0.4702%	98.9417%
85%	86%	1.2555%	50.0824%	190%	200%	0.3073%	99.2490%
86%	87%	1.2626%	51.3449%	200%	210%	0.2078%	99.4568%
87%	88%	1.2202%	52.5651%	210%	220%	0.1434%	99.6002%
88%	89%	1.2370%	53.8021%	220%	230%	0.1021%	99.7023%
89%	90%	1.2153%	55.0174%	230%	240%	0.0749%	99.7772%
90%	91%	1.1956%	56.2129%	240%	250%	0.0547%	99.8319%
91%	92%	1.1958%	57.4087%	250%	260%	0.0392%	99.8711%
92%	93%	1.1872%	58.5960%	260%	270%	0.0272%	99.8983%
93%	94%	1.1649%	59.7608%	270%	280%	0.0195%	99.9178%
94%	95%	1.1373%	60.8982%	280%	290%	0.0160%	99.9337%
95%	96%	1.1405%	62.0387%	290%	300%	0.0663%	100.0000%

The two previous tables illustrate the distribution of loans (across fixed year 30 loans) by LTV ratio and by RHP ratio, respectively. (The calculation of each of these ratios for individual loans was described above.) Our definition of the LTV and RHP ranges was based on examination of these tables.

We further subdivided the LTV categories into increments for purposes of accuracy. In particular, the calculation of the probability of negative equity for a "cell" of loans requires a finer definition of the LTV range. The table below shows the definitions of the LTV increments, as well as the value for each increment that we used as a proxy for each value within the range in calculating the probability of negative equity.

Table E.8

LTV Category	Proxy Value	Incremental Range	
Low	77.5%	0%	80%
	81.5%	80%	83%
Investor	84%	0%	85%
	86%	85%	87%
	90%	87%	140%
Mid	88.5%	87%	90%
	91%	90%	92%
	93%	92%	94%
	95%	94%	96%
High	97%	96%	98%
	99%	98%	100%
	105%	100%	140%

Age

Throughout this document, we will refer to the age of a pool of loans in terms of time t or policy year. In each case, we are defining the age of the pool of loans in terms of the number of years since the inception of the fiscal origination year (or endorsement year, if applicable). Therefore, policy year 1 for fiscal origination year 1985 is the time period between the inception of the period, October 1, 1984, and the date one year later, October 1, 1985. Fiscal origination year 1999 will reach age 4 ($t = 4$) on October 1, 2002.

Unemployment Rates

Unemployment rates are based on information extracted from the U.S. Department of Labor - Bureau of Labor Statistics. Downloaded from their website (Local Area Unemployment Statistics - http://stats.bls.gov/lauhome.htm) in August 2001, the available monthly civilian unemployment rates spanned from January 1978 through and including June 2001. The website

provided unemployment rates by state. Jim Campbell, Bureau of Labor Statistics, also provided unemployment rates as far back as 1970 for many of the states.

Based on the above information, we constructed one table of annual unemployment rates by calendar year. However, the HUD database was organized by fiscal orgination year. One fiscal origination year runs from October 1st through September 30th of each year. As a result, we converted the calendar year rates to fiscal origination year rates by taking 25% of the previous calendar year plus 75% of the current calendar year. For example, fiscal origination year 1975 is equal to 25% of 1974 and 75% of 1975.

We lagged the unemployment rate by two years due to the fact that when an individual becomes unemployed, the effects are not immediate mainly due the existence of unemployment benefits and personal savings. When an individual becomes unemployed he/she can first claim unemployment benefits and when that has run out his/her personal savings can be utilized. Any means of staying out of the red is explored before an individual would default on a loan. Consequently, it may take up to a year or two before unemployment actually affects an individual's mortgage payments. Based on this logic, we model expected loan termination behavior using lagged unemployment rates.

Time-adjusted Loan-to-Value Ratio (LTV_t)

We calculated LTV_t by individual loan. The time variable, t, represents the age of the fiscal origination year, where $t = 1$ represents the end of the fiscal year itself, $t = 2$ is the date one year later, and so on. Therefore, LTV_t is evaluated for a given loan as of October 1 of the fiscal year, plus t years, minus 1 (or as of 10/31/[FY + t – 1]).

$$LTV_t = LTV_0 \cdot \frac{SAF_t}{HPAF_t}, \text{ where}$$

$HPAF_t = \frac{HPI_t}{HPI_0}$, an adjustment for change in house prices between the time of the origination of the loan and the age t, and SAF_t is the scheduled amortization factor, or the percentage of the original loan amount estimated as still outstanding at age t.

Time-adjusted Payment-to-Income Ratio ($PAY.INC_t$)

$$PAY.INC_t = PAY.INC_0 \cdot \frac{contractrate_t}{contractrate_0} \cdot \frac{personalincome_0}{personalincome_t}$$

We obtained personal income per capita by MSA through 1999, and by state through the first quarter of 2001, from the Bureau of Economic Analysis (BEA) website. The BEA data was supplemented with house price index data from the OFHEO website in order to estimate per

capita personal income by MSA for the most recent years, and per capita personal income by state for remainder of fiscal origination year 2001.

The adjustment for change in personal income levels were made loan by loan. We made the adjustment for changes in the contract rate for groups of loans. The contract rate changes between time *t* and time 0 only on adjustable rate loans. The adjusted rate is estimated for a group of loans based on the historical changes in the index for adjustable rate loans, the 1-year, constant maturity T-bill rate. We also assumed that, on average, MMIF loans originated on April 15, which accounts for the seasonality in MMIF originations.

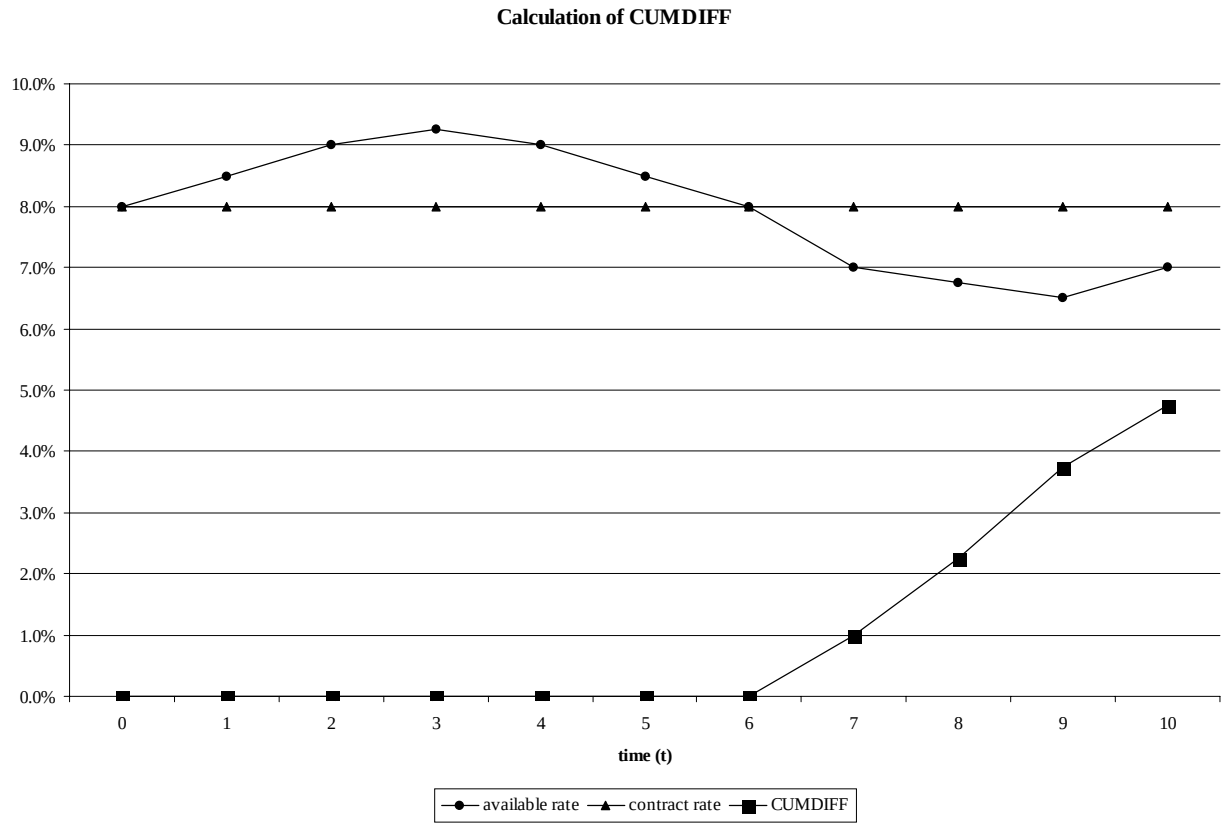
Refinance Incentive Ratio and Related Values

The refinance incentive ratio at a given time *t*, R_t, is defined as the ratio of the contract rate on a given loan to the available refinance rate at time *t*. If R_t is greater than one, the contract rate is higher than currently available rates at time *t*, and refinancing is an attractive prospect. A refinance incentive ratio less than one would imply little or no incentive to refinance at time *t*.

The variable used to indicate the level of the propensity to refinance is the exponentially weighted, moving average refinance incentive ratio at age *t*, or R'_t. $R'_t = z \cdot \overline{R}_t + (1-z) \cdot R'_{t-1}$, where $\overline{R}_t$ = the arithmetic mean of prior refinance incentive ratios up to time *t*, and *z* = the weight assigned to prior refinance incentive ratios. For this Review, we selected *z* = 0.75.

The variable $CUMDIFF_t$ and the age of the loan pool determine the degree to which the pool has burned out. $CUMDIFF_t$ is defined as the cumulative positive difference between the loan interest rate and the historically available refinance interest rate. The graph below illustrates this definition for the case of a loan with a fixed rate of 8 percent.

Chart E.9



As long as the available (refinance) rates are higher that the contract rate, there is no incentive to refinance and $CUMDIFF_t$ is zero. As the rates drop below the contract rate, however, there is incentive to refinance. As the positive differences accumulate, there will be very few borrowers left who will prepay and the pool "burns out".

In this Review, we calculated R_t, R_t', and $CUMDIFF_t$, at the "cell" level of detail. That is, we calculated R_t as the ratio of the average contract rate for a group of loans at a given age to the market rate available at the same point in time. R_t' was calculated based on the cell-level R_t. Similarly, we calculated $CUMDIFF_t$ based on the average contract rate for the group relative to the available market rate. It is our belief that there is very little difference between the values calculated at the cell-level and those calculated at the loan level of detail and weighted by amortized loan values.

House Price Appreciation

There are two house price appreciation variables used in the claims and prepayment rate models, an annual rate and a cumulative rate. Both are based on the historical house price index published by OFHEO.

We calculate the cumulative rate of house price appreciation by individual loan, and weight it based on the amortized values of loans surviving to age *t*. The cumulative rate for an individual loan is the ratio of the index value for the MSA (or state or census division) where the property is located at time *t* (plus three months) to the index value at the time the loan began amortizing (plus three months). We built a lag of three months into the index.

The annual rate of house price appreciation was based on the ratio of the average cumulative rate at time *t* to the cumulative rate at the previous age. This estimate of annual house price appreciation is slightly less clean than the calculation of the cumulative rate in that the mix of surviving loans by MSA may be slightly different between the two points in time. We do not consider that this "impurity" had a material effect on the results of our analysis.

The Probability of Negative Equity

In general, a normal and lognormal distribution is defined as follows:

$$Normal.Dist\left(\frac{x-\mu}{\sigma}\right) = f(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

$$Lognormal.Dist(x,\mu,\sigma) = Normal.Dist \cdot \left(\frac{\ln x - \mu}{\sigma}\right)$$

The probability of negative equity is defined within the parameters of the lognormal distribution. We have defined the lognormal parameters x, μ, and σ as follows:

$$x = 0$$

$$\mu = -\left[\left(\frac{LTV_0 \cdot SAF}{HPAF}\right) \cdot \frac{1}{\theta} + \frac{\theta}{2}\right], \text{ where } \theta = \sqrt{4A(t-0.5) + 16B(t-0.5)^2}$$

$$\sigma = 1$$

Please note that θ is defined as the volatility parameter by OFHEO. Other acronyms are defined as follows:

- LTV_0 is the loan-to-value ratio at time zero.
- *SAF* is the systematic amortization factor at time t.
- *HPAF* is the house price appreciation factor at time t.

We calculated probabilities of negative equity based on historical house price volatilities by MSA, by state, and by rural census division, published by OFHEO. The threshold for negative equity is an LTV ratio of 100%. Therefore, the calculated probabilities represent the probability that a loan with a given initial LTV will achieve a time-adjusted LTV of 100% or greater by time *t*.

The calculation of the probability of negative equity is by far the most labor-intensive calculation in terms of the required computer processing time. In order to save processing time, at what we felt was little or no cost in accuracy, we summarized the loans in our regression data sets by MSA. (Loans belonging to no MSA [i.e., rural properties] were grouped by census division, while non-rural properties that could not be assigned to an MSA were grouped by state.) We calculated a probability of negative equity for each MSA (or state or census division) at each point in time *t*, for each LTV increment proxy value. We could then weight the calculated probabilities for each "cell" based on the amortized value of surviving loans by MSA (or state or census division).

The historical probability of negative equity was estimated as described above. When we applied the results of our regression analysis to the forecast period, we did so on a country-wide basis. After discussion with OMB as to the proper means of accounting for regional covariance, we employed an adjustment suggested by OMB for purposes of estimating the probability of negative equity in the forecast period.