Appendix E - Data Transformation

HUD provided Deloitte & Touche with data from the Single Family Data Warehouse for fiscal endorsement years 1975 through 2002 as of March 31, 2003. 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	185,967	25	185,942	0.0134%
1976	222,097	33	222,064	0.0149%
1977	256,153	65	256,088	0.0254%
1978	294,582	104	294,478	0.0353%
1979	389,783	128	389,655	0.0328%
1980	337,117	520	336,597	0.1542%
1981	216,269	0	216,269	0.0000%
1982	149,114	0	149,114	0.0000%
1983	505,956	2	505,954	0.0004%
1984	287,136	0	287,136	0.0000%
1985	400,623	0	400,623	0.0000%
1986	929,125	1	929,124	0.0001%
1987	1,126,925	0	1,126,925	0.0000%
1988	615,830	0	615,830	0.0000%
1989	634,642	0	634,642	0.0000%
1990	715,756	0	715,756	0.0000%
1991	643,553	0	643,553	0.0000%
1992	637,162	0	637,162	0.0000%
1993	991,030	0	991,030	0.0000%
1994	1,059,866	0	1,059,866	0.0000%
1995	521,263	0	521,263	0.0000%
1996	728,073	0	728,073	0.0000%
1997	740,215	0	740,215	0.0000%
1998	955,486	0	955,486	0.0000%
1999	1,121,712	0	1,121,712	0.0000%
2000	832,977	0	832,977	0.0000%
2001	1,058,860	0	1,058,860	0.0000%
2002	1,088,895	0	1,088,895	0.0000%
2003	450,756	3	450,753	0.0007%
Total	**18,096,923**	**881**	**18,096,042**	**0.0049%**

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	185,942	438	0.2356%	185,504
1976	222,064	172	0.0775%	221,892
1977	256,088	213	0.0832%	255,875
1978	294,478	169	0.0574%	294,309
1979	389,655	55	0.0141%	389,600
1980	336,597	26	0.0077%	336,571
1981	216,269	2	0.0009%	216,267
1982	149,114	71	0.0476%	149,043
1983	505,954	114	0.0225%	505,840
1984	287,136	111	0.0387%	287,025
1985	400,623	39	0.0097%	400,584
1986	929,124	29	0.0031%	929,095
1987	1,126,925	43	0.0038%	1,126,882
1988	615,830	28	0.0045%	615,802
1989	634,642	25	0.0039%	634,617
1990	715,756	50	0.0070%	715,706
1991	643,553	28	0.0044%	643,525
1992	637,162	64	0.0100%	637,098
1993	991,030	82	0.0083%	990,948
1994	1,059,866	64	0.0060%	1,059,802
1995	521,263	25	0.0048%	521,238
1996	728,073	34	0.0047%	728,039
1997	740,215	65	0.0088%	740,150
1998	955,486	50	0.0052%	955,436
1999	1,121,712	41	0.0037%	1,121,671
2000	832,977	25	0.0030%	832,952
2001	1,058,860	22	0.0021%	1,058,838
2002	1,088,895	11	0.0010%	1,088,884
2003	450,753	0	0.0000%	450,753
Total	**18,096,042**	**2,096**	**0.0116%**	**18,093,946**

Loan-to-Value Ratio

In the 2003 study, as in 2002, HUD provided us with a variable referred to as "ratio_loan_to_vl_new". This variable has had a significant impact on the LTV_0 distribution, especially for the streamline refinanced loan types.

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-3

	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	20.1547	17.2260	N/A	N/A	N/A	N/A
1976	20.3955	17.2851	N/A	N/A	N/A	N/A
1977	20.2348	16.8675	N/A	N/A	N/A	N/A
1978	21.6115	17.0934	N/A	N/A	N/A	N/A
1979	22.2515	17.0985	N/A	N/A	N/A	N/A
1980	23.3949	18.4603	N/A	N/A	N/A	N/A
1981	24.5209	19.4220	N/A	N/A	N/A	N/A
1982	24.7577	20.5804	N/A	N/A	N/A	N/A
1983	23.4481	22.9855	N/A	N/A	N/A	N/A
1984	24.2124	22.8784	N/A	N/A	N/A	N/A
1985	23.3324	22.8712	22.8563	N/A	N/A	N/A
1986	21.4750	20.4722	21.9185	N/A	N/A	N/A
1987	21.3457	19.8183	21.5216	N/A	N/A	N/A
1988	23.3609	22.4398	23.0581	22.0005	21.7702	22.5783
1989	25.3360	23.4132	25.4995	22.9545	21.1502	23.6427
1990	23.7811	21.7217	23.2271	22.6490	20.9291	N/A
1991	22.9644	20.9718	23.8764	24.0409	22.1722	21.6555
1992	22.7286	20.1361	23.4386	23.5244	22.1034	22.3442
1993	22.4560	19.5675	23.6837	23.8364	21.6317	23.4983
1994	22.8231	19.3591	24.1932	21.5813	20.6633	21.6082
1995	23.9895	20.1669	24.8931	23.9421	21.9757	23.8516
1996	24.0296	20.5343	24.9627	24.3811	21.4591	24.2194
1997	24.3603	21.0721	24.9650	25.5726	22.3733	25.4589
1998	24.2761	21.1932	25.0568	28.7751	22.4749	27.2623
1999	25.0286	21.9187	26.1875	24.9982	21.3667	26.7515
2000	26.9208	23.6614	27.3889	27.6116	24.1360	27.2267
2001	26.4429	23.9416	27.3987	26.9950	24.5625	26.6314
2002	26.3779	24.2515	27.2767	26.1146	23.8840	25.4031
2003	26.2602	24.2351	27.7072	25.8167	23.5384	26.1610

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-4

Origination Year	Number of Loans, All Loan Types	LTV 10% or Less	LTV 140% or Greater	Remaining Loans	Percent Excluded
1975	185,504	26,503	407	158,594	15%
1976	221,892	28,293	604	192,995	13%
1977	255,875	24,188	853	230,834	10%
1978	294,309	41,327	1,359	251,623	15%
1979	389,600	67,734	1,465	320,401	18%
1980	336,571	36,926	1,527	298,118	11%
1981	216,267	47,328	1,023	167,916	22%
1982	149,043	22,620	511	125,912	16%
1983	505,840	88,820	718	416,302	18%
1984	287,025	8,039	419	278,567	3%
1985	400,584	4,559	4,079	391,946	2%
1986	929,095	5,822	2,696	920,577	1%
1987	1,126,882	2,817	2,112	1,121,953	0%
1988	615,802	415	1,680	613,707	0%
1989	634,617	3,239	670	630,708	1%
1990	715,706	10,047	906	704,753	2%
1991	643,525	22,775	580	620,170	4%
1992	637,098	17,862	2,805	616,431	3%
1993	990,948	7,050	3,800	980,098	1%
1994	1,059,802	2,303	4,691	1,052,808	1%
1995	521,238	243	2,558	518,437	1%
1996	728,039	430	4,112	723,497	1%
1997	740,150	336	4,555	735,259	1%
1998	955,436	8,016	4,530	942,890	1%
1999	1,121,671	42,722	24	1,078,925	4%
2000	832,952	5,082	0	827,870	1%
2001	1,058,838	5,622	0	1,053,216	1%
2002	1,088,884	19	0	1,088,865	0%
2003	450,753	10	0	450,743	0%
Total	**18,093,946**	**531,147**	**48,684**	**17,514,115**	**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-5

LTV Range		Percentage of Loans in Range	Cumulative Percentage
10%	15%	0.0059%	0.0059%
15%	20%	0.0055%	0.0114%
20%	25%	0.0119%	0.0233%
25%	30%	0.0224%	0.0457%
30%	35%	0.0831%	0.1288%
35%	40%	0.0775%	0.2062%
40%	45%	0.1170%	0.3232%
45%	50%	0.1795%	0.5027%
50%	55%	0.2738%	0.7765%
55%	60%	0.3789%	1.1554%
60%	65%	0.5635%	1.7189%
65%	70%	0.8528%	2.5717%
70%	75%	1.4908%	4.0625%
75%	80%	2.5204%	6.5829%
80%	85%	5.1445%	11.7273%
85%	90%	8.6922%	20.4196%
90%	91%	1.9348%	22.3544%
91%	92%	2.5034%	24.8578%
92%	93%	3.1031%	27.9609%
93%	94%	3.9696%	31.9305%
94%	95%	6.1824%	38.1129%
95%	96%	10.7787%	48.8916%
96%	97%	22.1423%	71.0339%
97%	98%	19.2702%	90.3041%
98%	99%	5.3076%	95.6117%
99%	100%	2.4347%	98.0464%
100%	101%	0.5550%	98.6014%
101%	102%	0.2788%	98.8802%
102%	103%	0.1693%	99.0496%
103%	104%	0.0974%	99.1469%
104%	105%	0.0363%	99.1833%
105%	110%	0.2039%	99.3871%
110%	115%	0.1019%	99.4891%
115%	120%	0.0773%	99.5664%
120%	125%	0.0625%	99.6289%
125%	130%	0.3711%	100.0000%

Table E-6

RHP Range		Percentage of Loans in Range	Cumulative Percentage	RHP Range		Percentage of Loans in Range	Cumulative Percentage
0%	10%	0.0016%	0.0016%	96%	97%	1.1524%	63.69%
10%	20%	0.0882%	0.0899%	97%	98%	1.1193%	64.81%
20%	30%	0.8686%	0.9585%	98%	99%	1.1079%	65.92%
30%	40%	2.8280%	3.7865%	99%	100%	1.0976%	67.01%
40%	50%	5.6738%	9.4603%	100%	101%	1.0681%	68.08%
50%	60%	9.2480%	18.7083%	101%	102%	1.0395%	69.12%
60%	61%	1.0384%	19.7467%	102%	103%	1.0317%	70.15%
61%	62%	1.0577%	20.8044%	103%	104%	0.9968%	71.15%
62%	63%	1.0858%	21.8902%	104%	105%	0.9930%	72.14%
63%	64%	1.1120%	23.0021%	105%	106%	0.9780%	73.12%
64%	65%	1.1357%	24.1378%	106%	107%	0.9239%	74.04%
65%	66%	1.1398%	25.2776%	107%	108%	0.9278%	74.97%
66%	67%	1.1660%	26.4436%	108%	109%	0.8881%	75.86%
67%	68%	1.1582%	27.6018%	109%	110%	0.8639%	76.72%
68%	69%	1.2124%	28.8142%	110%	111%	0.8623%	77.59%
69%	70%	1.2187%	30.0328%	111%	112%	0.8257%	78.41%
70%	71%	1.1977%	31.2305%	112%	113%	0.8032%	79.22%
71%	72%	1.2318%	32.4624%	113%	114%	0.7658%	79.98%
72%	73%	1.2379%	33.7003%	114%	115%	0.7467%	80.73%
73%	74%	1.2659%	34.9661%	115%	116%	0.7474%	81.48%
74%	75%	1.2556%	36.2217%	116%	117%	0.7080%	82.18%
75%	76%	1.2726%	37.4944%	117%	118%	0.6891%	82.87%
76%	77%	1.2876%	38.7819%	118%	119%	0.6734%	83.55%
77%	78%	1.2734%	40.0553%	119%	120%	0.3196%	83.87%
78%	79%	1.2926%	41.3479%	120%	130%	5.4263%	89.29%
79%	80%	1.3002%	42.6481%	130%	140%	3.6831%	92.97%
80%	81%	1.2846%	43.9327%	140%	150%	2.4515%	95.43%
81%	82%	1.2663%	45.1990%	150%	160%	1.6085%	97.03%
82%	83%	1.3104%	46.5094%	160%	170%	1.0237%	98.06%
83%	84%	1.2885%	47.7980%	170%	180%	0.6511%	98.71%
84%	85%	1.2857%	49.0837%	180%	190%	0.4153%	99.13%
85%	86%	1.2725%	50.3562%	190%	200%	0.2683%	99.39%
86%	87%	1.2789%	51.6351%	200%	210%	0.1701%	99.56%
87%	88%	1.2636%	52.8986%	210%	220%	0.1143%	99.68%
88%	89%	1.2540%	54.1526%	220%	230%	0.0834%	99.76%
89%	90%	1.2241%	55.3768%	230%	240%	0.0601%	99.82%
90%	91%	1.2373%	56.6141%	240%	250%	0.0429%	99.86%
91%	92%	1.2232%	57.8372%	250%	260%	0.0301%	99.89%
92%	93%	1.2023%	59.0395%	260%	270%	0.0221%	99.92%
93%	94%	1.1832%	60.2227%	270%	280%	0.0161%	99.93%
94%	95%	1.1630%	61.3858%	280%	290%	0.0131%	99.95%
95%	96%	1.1503%	62.5361%	290%	300%	0.0545%	100.00%

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-7

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 5 ($t = 5$) on October 1, 2003.

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) on May 28, 2003, the available monthly civilian unemployment rates spanned from January 1978 through and including March 2003. 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 origination 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 2001, and by state through 2002, 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 fiscal origination years 2002 and 2003, and per capita personal income by state for fiscal origination year 2003.

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.

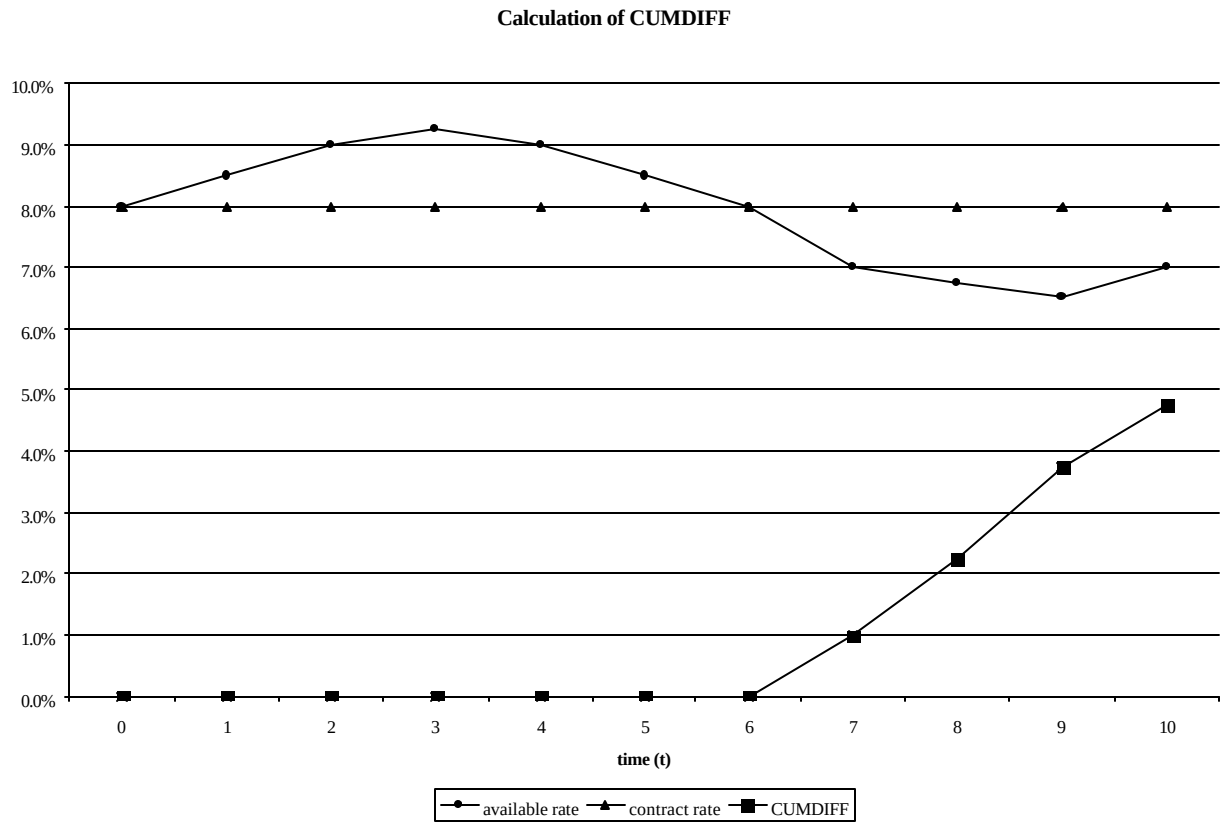
Refinance Incentive Ratio and Related Values

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-8



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 believe 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:

$$Normal.Dist\left(\frac{x-m}{s}\right) = f(x,m,s) = \frac{1}{\sqrt{2p}} \cdot e^{-\left(\frac{(x-m)^2}{2s^2}\right)}$$

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

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

$$x = 0$$

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

$$s = 1$$

Please note that q 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 countrywide 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.

External Data and Sources Used in Building Regression Data Sets

External Data	Source	Website
Unemployment rates - seasonally adjusted monthly civilian unemployment rate - by State	U.S. Department of Labor, Bureau of Labor Statistics	http://stats.bls.gov/
One-Year U.S. Treasury Constant Maturity Rate	Economic Research – Federal Reserve Bank of St. Louis	http://research.stlouisfed.org/fred2/categories/22/downloaddata
10-Year Treasury Constant Maturity Rate	Economic Research – Federal Reserve Bank of St. Louis	http://research.stlouisfed.org/fred2/categories/22/downloaddata
30-Year Conventional Mortgage Rate	Economic Research – Federal Reserve Bank of St. Louis	http://research.stlouisfed.org/fred2/categories/22/downloaddata
Monthly Average Commitment Rates on 30-Year Fixed Rate Mortgages since 1971	Freddie Mac	http://www.freddiemac.com/pmms/pmms30.htm
Monthly Average Commitment Rates on 15-Year Fixed Rate Mortgages since 1991	Freddie Mac	http://www.freddiemac.com/pmms/pmms15.htm
Monthly Average Commitment Rates on 1-Year Adjustable Rate Mortgages since 1984	Freddie Mac	http://www.freddiemac.com/pmms/pmmsarm.htm

External Data and Sources Used in Building Regression Data Sets, continued …

External Data	Source	Website
House Price Indices (as of 2002 4th quarter) - by State, including District of Columbia - by MSA - by Census Division	Office of Federal Housing Enterprise Oversight (OFHEO)	http://www.ofheo.gov/house/download.html
House Price Volatility Parameters - by State, including District of Columbia - by MSA - by Census Division	Office of Federal Housing Enterprise Oversight (OFHEO)	http://www.ofheo.gov/house/download.html [Note: MSA data provided by Shelley Dreiman]
Per Capita Personal Income - by MSA - by State	U.S. Department of Commerce, Bureau of Economic Analysis	http://www.bea.doc.gov/bea/regional/reis/ and http://www.bea.doc.gov/bea/regional/spi/
Median House Price - by MSA - by State	1975 through 1997 from PriceWaterhouseCoopers File, 1998 through 4Q2002 estimated based on OFHEO HPI series	