Technical Paper Series
Congressional Budget Office

The FHA Budget Subsidy Simulation System: A Dynamic Simulation Model of Budget Outcomes for FHA Single-Family Mortgage Insurance

Charles A. Capone, Jr.
Microeconomic and Financial Studies Division

July 2003
2003-7

For additional information about this paper, contact Charles Capone at (202) 226-2649
or by email at ccapone@cbo.gov.

Abstract

This paper introduces a dynamic simulation model of value to the government from FHA single-family mortgage insurance. The model, known as the FHA Budget Subsidy Simulation System (FHA-BSSS), was built at CBO as a tool for analyzing the sensitivity to economic fluctuations of net receipts from or outlays to FHA, and how that sensitivity could be used to inform initial budget estimates and ongoing re-estimates.

The FHA-BSSS is a software system that utilizes loan-record databases and econometric models to forecast FHA cash flows and resulting budget subsidy rates in a stochastic environment. Separate econometric models are used to create the economic forecasts, relate mortgage termination rates to economic conditions, and to determine post-default outcomes and costs. FHA cash flows are computed from predicted mortgage events, and then are converted into budget subsidy rates and dollars.

This technical paper both describes the research entailed in building the FHA-BSSS, and provides implementation details to allow replication of the model. The paper ends with some basic results from using the FHA-BSSS to make current estimates of subsidy rates on 1992-2007 budget-year cohorts of FHA insured loans. The model version used was current as of March 2003, economic data was current as of December 2002, and FHA loan information was current as of September 30, 2002.

CONTENTS

FIGURES, BOXES, AND TABLES

I. INTRODUCTION[1]

FHA AND ITS BUDGET ACCOUNTING

After nearly 70 years of existence, the Federal Housing Administration (FHA) remains an important federal credit program. It insures nearly 10 percent of all mortgage loans originated in the U.S. each year, and it accounts for half of all federal credit activity–direct loans and guarantees of all types. FHA has a statutory mandate for its major single-family mortgage insurance programs to remain actuarially sound.[2] That mandate requires that net budgetary receipts for the government remain positive under a wide range of economic conditions.

This analysis concludes that the balance of underwriting standards and premium fees now used by FHA do not ensure actuarial soundness into the future. FHA's principal insurance fund is, therefore, not as sound as indicated by annual independent actuarial reviews and financial audit reports. Those reviews and reports include as assets the expected (future) net income from outstanding insurance and, according to this study, that value is overstated.

[1]Several individuals read this manuscript at various stages: Kim Cawley, Ufuk Demiroglu, Robert Dennis, Peter Fontaine, Robert McClelland, John McMurray, Susanne Mehlman, Albert Metz, and Marvin Phaup. The assistance of these persons is greatly appreciated, yet the author maintains responsibility for the paper's final content.

[2]This and all other general comments in this paper refer to single-family mortgage insurance provided in programs that are part of the Mutual Mortgage Insurance Fund. That Fund is required to be actuarially sound. It accounts for around 90 percent of all FHA single-family insurance, and over 80 percent of all insurance (dollar volume) provided by FHA.

Expected net income is overstated because it is based on estimates of insured-loan defaults and prepayments under unrealistic economic conditions, where house prices and interest rates are very stable. In contrast, it is well known that normal economic fluctuations themselves increase loan terminations, which then lowers net receipts from loan guarantees. In addition, FHA faces an asymmetric profit/loss function so that budget receipts estimated using average economic conditions will be larger than the average of receipt levels across all possible economic conditions.

The existing bias in initial budget estimates is seen when one decomposes annual budget revisions required by the Office of Management and Budget (OMB). Those revisions, called re-estimates, are based on actual, to-date program performance and new forecasts of future default and prepayment rates of the insured mortgages. Budget receipt re-estimates based on loan performance have been overwhelmingly downward for all annual budget-year cohorts with measurable experience.[3] The re-estimate numbers published by OMB, however, do not show this result because they also include offsetting amounts from one-time program changes and asset sales. When these offsets are removed, there emerges a clear pattern of over-predicting expected budget receipts from each year's book-of-business and by a large percentage.

To understand the extent of the bias, and to correct for it, requires a simulation model

[3]This type of accounting has been in place for federal credit programs since 1992. For summaries of the FHA re-estimate revisions see, Office of Management and Budget, *FY2004 Federal Credit Supplement*, Table 8.

capable of valuing an insurance portfolio under a wide range of plausible economic conditions. This technical paper describes such a model for FHA single-family mortgage insurance programs under the umbrella of the Mutual Mortgage Insurance Fund.

USE OF SIMULATION MODELS IN MORTGAGE VALUATION

Simulation models are in regular use by private investors in mortgages and mortgage-backed securities to develop risk-adjusted valuations. The most prominent of these are option-adjusted spread (OAS) models that attempt to make default-free but prepayment-prone mortgage yields comparable to (non-callable) bond yields. Simulated yield-adjustment factors account for pool-level prepayment potential and are subtracted from mortgage coupon rates to arrive at bond-like yields. Models that analyze credit-risk yield spreads are less common, principally because mortgage credit risk in the U.S. is highly concentrated in a small number of government agencies, government sponsored enterprises, and private mortgage insurers. However, large commercial banks and thrifts are now looking at developing such models to potentially decrease capital requirements on whole-loan investments under pending revisions to the Basle capital accords.[4]

Simulation models face two primary challenges: developing reliable behavioral equations

[4]A recent issue of the *Journal of Banking and Finance* (vol. 24, 2000) has numerous articles that detail models currently in use at major banks. See also, Kaskowicz, et al, *Best Practices in Mortgage Default-Risk Management and Economic Capital*, unpublished manuscript, Loan Performance Inc. (February 2002), available at http://www.loanperformance.com/library/articles/best_practices_2002.pdf. In April, 2003, Moody's Investor Services introduced a proprietary simulation model for measuring mortgage credit risk and pool-level capital requirements which is similar in nature to the FHA-BSSS.

of how mortgage borrowers respond to economic events; and simulating a reasonable range of future economic outcomes with appropriate correlations between economic measures/variables. When these two elements are pieced together, they must produce reasonable results across a wide spectrum of economic conditions so as to provide a probability distribution of ultimate outcomes that can accurately guide investor decisions.

This study of FHA single-family mortgage insurance addresses these challenges directly. Economic conditions are modeled in a vector autoregressive system of equations, where movements in any one variable have effects on movements of all other variables in the system, in future periods. Behavioral models are built for borrower termination decisions, and for the events and time sequences triggered by borrower default. Adding the latter allows for a full development of the probability distribution of default-related claim costs.

THE FHA BUDGET SUBSIDY SIMULATION SYSTEM (FHA-BSSS)[5]

Performance of government loan guarantee programs like FHA are measured by guarantee-contract values, as defined by current budget accounting rules. The question of interest is, for each dollar of loans insured by FHA, what is the expected net present value to the federal government? This value is called a subsidy rate. The FHA Budget Subsidy

[5]Research behind the FHA-BSSS was accomplished with the help of many individuals. In particular, the author thanks: Robert Dennis, Robert Arnold, John Peterson, and Ufuk Demiroglu of the CBO Macroeconomic Analysis Division for building the stochastic economic-environment models; Malgorzata Klosek, visiting scholar, for the model of regional house-price cycles; Jenny Au, Sean Corcoran, Lori Ellebracht, Erin Hirsch, Joseph Nichols, DaRon Ross, Errick Simmons, and Aurora Swanson for research assistance; Carol Frost for SAS consulting; Georgia Brown, Eric Guille, Guanli Lu, and Rick Williams for IT support; Judy May, Dominique Stasulli, and Edward Szymanoski of FHA for access to data; and Debbie Lucas for research oversight.

Simulation System (FHA-BSSS) described in this study attempts to calculate an unbiased estimate of the true, but unknown subsidy rate that will reveal itself as the future economy unfolds and actual program outcomes are known.

The FHA-BSSS also measures the probability that expected net receipts from FHA could rather turn out to be (unexpected) net outlays for the federal budget. Analysis of such outcomes can lead to measures of economic capital: the level of capital reserves that would be required to assure solvency of the FHA program against a loss-event defined in the tail of the outcome probability distribution. Not that a federal government program can or needs to hold capital reserves, but it is a risk-metric familiar to private investors. Since 1990, the Congress has indicated an interest in FHA producing net receipts for the federal budget, as defined by statutory "capital" requirements.[6]

In process, the FHA-BSSS takes pools of FHA-insured loans and projects their experience, and the resulting cash flows of FHA, through 1000 different future economic paths. The use of 1000 simulation trajectories is arbitrary. It was chosen to balance the need for generating a sufficient number of outlier events with the need to control total computer processing time.

Simulation results shown in this paper (section VI) are for entire budget-year cohorts of loan guarantees–FHA endorsements by fiscal year book-of-business. The structure of the

---

[6] 12 USC 1711(f).

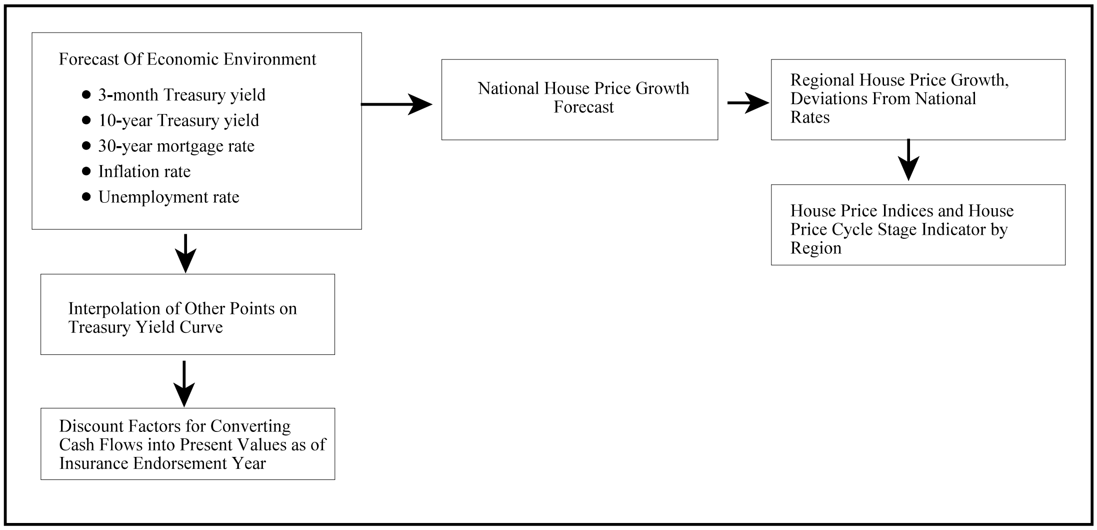
FHA-BSSS allows for further examination of subsidy rate outcomes for sub-groups based upon values of nine different categorical variables, including product type and loan-to-value class (see section III). The simulation system also was built to address a wide range of policy issues involving how changes in administrative policies and program usage affect the Federal budget. These ancillary matters are, however, beyond the scope of this paper and are not addressed here.

OUTLINE OF REMAINING SECTIONS

Figure 1 provides a schematic outline of how the various parts of the FHA-BSSS work together to create subsidy rate forecasts. The (downward) flow of that illustration is matched by the order of material presented in the following sections of this study. Relevant parts of Figure 1 are repeated at the beginning of each section, as a guide to readers. Each segment of Figure 1 is mapped to sections of the study as follows:

Section of Figure 1	Part of the FHA-BSSS	Sections of this Study
Top	Forecast economic inputs for mortgage performance models	II. Economic Forecasting Models
Center	Forecast mortgage default and prepayment rates	III. Regression Models of Mortgage Default and Prepayment
Center	Forecast default resolution types and timing and losses for foreclosures	IV. Post-Default Outcome Regression Models
Bottom	Calculate cash-flow time series and discount to net present values	V. Final Forecast Assumptions and Subsidy Rate Calculations
(not in figure)	Discussion of simulation results	VI. Simulation Results for the 1992-2007 Cohorts

II. ECONOMIC FORECASTING MODELS



The first thesis of simulation analysis is that normality of probability density functions cannot always be assumed. When the outcome of interest is the product of numerous stochastic processes, each with its own unique distributional form, then the distribution of final outcomes can be best understood through repeated-experiment simulations. The mean outcome across a large number of simulated economic environments can provide an unbiased estimate of the actual, but unknown outcome of the process in question. In this case the outcome is an (average) predicted subsidy rate, which is the net present value of cash flows from loan guarantees, per dollar of guarantees, for each budget year cohort or book-of-business.

For mean simulation outcomes to be unbiased estimators, the simulations of economic environments must be realistic. Such realism includes the size and distribution of potential values of economic variables, the correlations between their values, and their movements over time. To create realistic economic environments, the FHA-BSSS uses both regression analysis and stochastic (random) shocks to the values produced using the regression equations. The various models and steps involved are:

- A vector autoregressive (VAR) model of macro economic changes;
- Regression of national house price movements on the economic variables of the VAR model;
- Addition of random shocks or disturbances to values predicted with the regression models, to create the full range of possible variable values;
- Addition of the (random) possibility of Great-Depression-like events;
- Generation of random local house price paths and cycles around national price movements; and
- Interpolation of points along the Treasury yield curve to create interest rates for discounting cash flows to produce subsidy estimates.

VECTOR AUTOREGRESSIVE MODEL OF ECONOMIC CONDITIONS

The FHA-BSSS projects economic conditions using a five-equation Vector Autoregression (VAR) model. VAR techniques are especially suited to time-series data where there are strong correlations among variables but the direction of causation is difficult to determine. VAR is also known as impulse-response modeling because when one variable changes, its effects ripple over time through all other variables. The magnitudes of these responses are modeled in a system of equations where contemporaneous values of each variable are regressed on lagged values of all variables in the system.

The VAR used to project quarterly economic conditions for the FHA-BSSS uses a 3-quarter lag structure in equations for the following variables:[7]

- 30-year Mortgage Interest Rate, in log form;
- National Civilian Unemployment Rate, in log form;
- Inflation, measured as the annual rate of change in the core component of the urban consumer CPI each quarter, transformed as described below;[8]
- Ratio of the 30-year mortgage rate to the 10-year constant maturity Treasury rate, in log form; and
- Ratio of the 10-year Treasury yield to 3-month Treasury bill rate , in log form.

Logarithmic (log) transformations yield regression equations based on percentage changes in variables rather than on levels. This change is desirable for stochastic simulations of how economic conditions change over time, and is also valuable for creating regression variables with constant variances. Using the untransformed variables (actual interest rates, unemployment rates, and inflation rates) directly in the regression would produce sub-optimal results because error variances would be proportional to the

[7]Longer lag structures were tested but provided no net benefits.

[8]Core inflation excludes food and energy price movements.

values of the underlying variables, creating a heteroskedasticity problem. Log transformations generally alleviate this problem; however, log transformations are also restricted to positive variable values. While this constraint is not a problem for estimating a regression model using recent history, it then restricts the values that can be forecast for the future.

The possibility of negative forecast values is desirable for one variable, inflation. To allow for negative forecast values while also managing the inherent heteroskasticity problem, a transformation other than the logarithmic is required. The transformation used here relies upon the volatility of the inflation time series to structure an appropriate function. Its form is derived and shown in Box 1.

The VAR is estimated using quarterly data from 1968 to 2001, and results are in Table 1. All interest rates come from historical series published by the Federal Reserve Board, while unemployment and core inflation rate data are from the Bureau of Labor Statistics of the U.S. Department of Commerce.[9] The unemployment rate is the adult, non-institutionalized population unemployment rate.

[9]The mortgage rate series is the conventional 30-year mortgage rate reported by Freddie Mac to the Federal Reserve Board.

FORECASTING THE VAR VARIABLES

The VAR equations are the basis for forecast simulations but, by themselves, do not produce the values necessary for predicting mortgage performance and subsidy rates. To generate forecast predictions, the VAR equations (Table 1) are seeded from recent historical experience. Doing so provides lagged values for computing predictions in the first quarter of the forecast period. Random shocks (or deviations) are next added to those first quarter fitted values using re-sampling techniques. These shocks are random draws from the regression residuals for each VAR equation. They provide deviations of actual historical data from regression equation predictions.[10]

Once first quarter values are determined for VAR variables, they become one-quarter lagged values when forecasting the second quarter and the re-sampling procedure is again used to generate random shocks. This two-step process continues for the entire 30-year forecast period. Once finished, the VAR variables are un-transformed to get values of the economic variables of interest. For interest rates and unemployment, the inverse transformation is exponentiation. For inflation, the inverse transformation function is described in Box 1.

Quarterly forecasts for each variable are normalized so that mean values across all

[10] Readers wanting to replicate this forecasting system, who do not have the residuals from the VAR estimation, can generate random shocks by assuming they are normally distributed random variables with zero means and standard deviations equal to the standard errors of the regression equations reported in Tables 1 and 2.

simulations match CBO baseline economic forecasts, for each quarter. This baseline forecast is reported by CBO in its most recent *Budget and Economic Outlook*.[11] The normalization is the additive difference between the simulation mean in any given quarter, and the CBO baseline value. The normalization constant for each variable, in each quarter, is added to predicted values in each simulation run.

HOUSE–PRICE GROWTH-RATE REGRESSION MODEL

A national house-price growth equation is estimated using functions of economic variables found in the VAR system. This creates one-way causation where the macro economy affects housing prices but house-price growth does not affect the macro economy. In forecast simulations, this approach allows house prices to be affected by shocks in the macro economy, without temporal shocks in the housing market having feedback effects on the economy.[12]

The estimated house-price-growth (least squares) regression equation is reported in Table 2. It models annual-equivalent rates of house price growth each quarter as a function of:

- The change in expected inflation, from previous quarter;

---

[11] Results shown in this technical paper use the December 2002 *Budget and Economic Outlook.*

[12] While there is evidence that house price growth affects consumer expenditures through the cashing out of home equity via second mortgages and cash-out refinancing, CBO does not predict house prices as part of its larger macroeconomic model. Thus, making house price growth a recursive function to the VAR provides needed correlations of house prices with economic conditions without violating the premises of the larger macroeconomic model. There is no consensus in the broader research community as to whether housing has measurable feedback effects into the general economy or not.

- The change in real mortgage interest rates, from previous quarter;
- One quarter lag in the change in real mortgage rates; and
- The change in the unemployment rate, from four quarters earlier.

Expected inflation is a weighted average of inflation in the previous 12 quarters. The weights are given with Table 2 and come from the Phillips Curve model used in CBO's macroeconomic forecasts.

FORECASTING HOUSE PRICE GROWTH

Forecast values of the VAR economic variables, with the random shocks, provide inputs into the house price growth equation. To increase the variability of house-price growth forecasts, regression coefficient values are assumed normal (Gaussian) random variables with means equal to the regression estimates, and standard deviations from those same estimates (see Table 2).[13] Random shocks are added to house-price growth predictions using the same re-sampling techniques employed for the VAR equations.

A supplement to the house-price growth equation is added to account for the possibility of a cyclical peak in national house price growth at the start of the forecast period (2002, fourth quarter). At that time, real national house price growth had been above average for

[13]The effect of this tactic is minor compared to the size of random shocks that are added to the predicted growth-rate series. Yet allowing the regression coefficients to be random variables when forecasting the future results in a more continuous distribution of outcomes and more accurate confidence intervals, when compared to historical variations in house price growth rates.

over three years and was just beginning to show signs of slowing down. To look for the evidence of a cyclical peak, we examined the time trend of the ratio of a national house price index (HPI) to the Bureau of Labor Statistics residential rent price index (rent of primary residence series).[14] This ratio shows cycles of roughly ten years duration, with current market conditions at a peak with house prices 12.8 percent above what would be supported by their average relationship to rental prices (over 1975-2002). This 12.8 percent represents cumulative growth of house prices in excess of what would maintain long-run parity between the purchase and rental markets.[15]

This excess growth may represent a housing market disequilibrium, rather than a permanent shift in the demand for owner-occupied housing. To deal with this uncertainty, a probabilistic approach is employed for when and by how much to dissipate the 12.8 percent excess growth . The dissipation in any simulated economic scenario is based a first-order autocorrelation coefficient for the HPI-to-rental-price ratio series. Using regression analysis, the quarterly autocorrelation coefficient of the series is estimated as 0.994, with a standard deviation of 0.018. Values used for the autocorrelation coefficient in the economic simulations are randomly drawn from a normal distribution with mean 0.994 and standard deviation 0.018, but truncated at 1.00. When the randomly selected coefficient is 1.00 (37 percent of the time) there is no dissipation of the 12.8 percent

---

[14] The house price series is the national index produced by the Office of Federal Housing Enterprise Oversight (OFHEO). It is a repeat-transactions price index based on loan purchase information provided to OFHEO from Fannie Mae and Freddie Mac.

[15] Previous peaks occurred in 1979 and in 1990.

"excess growth," so that such growth is assumed to be due to a permanent shift in demand for owner-occupied housing. Thus, the adjustment factor imbeds a probabilistic estimate of whether or not there truly has been abnormally high real house price growth and a disequilibrium in housing markets.

If house price growth predicted by the regression equation is $\Delta HPI(t,i)$, for forecast quarter $t = \{1,..., 120\}$ in simulation $i = \{1,...,1000\}$, the final, adjusted growth rate with the excess-growth adjustment factor, $\Delta H\tilde{P}I(t,i)$, is:

$$\Delta H\tilde{P}I(t,i) = \Delta HPI(t,i) + \varepsilon(i,t) - 4 \cdot \left(12.8 \cdot \left(1 - \rho(i)\right) \rho(i)^{t-1} \right)$$

where $\varepsilon(i,t)$ is the random shock adjustment, $\rho(i) \sim N(.994, (.018)^2)$ is the random correlation coefficient for simulation i (censored at 1.0), 12.8 (percent) is the excess growth to be removed over time, and multiplying the right-hand-side by 4 converts the quarterly dissipation effect to an annual rate. For all values of $\rho_i < 1.0$, $\Delta H\tilde{P}I(t,i)$ declines with t. Smaller values of ρ_i result in faster dissipation, while larger values result in slower dissipation.[16]

[16]The dissipation process is not a correction for any implied housing price "bubble." Indeed, even with low values of ρ, dissipation of the 12.8 percent "excess" growth occurs very slowly. With a ρ_i value two standard deviations below its mean, one-half of the excess growth is dissipated in four years. At one standard deviation below the mean, it takes seven years to remove one half of the excess growth, and at the mean value itself it takes 29 years to reach the 50 percent dissipation mark. While this correction does not address the question of housing price bubbles, the potential for local price bubbles is addressed in the next section.

Because house-price growth-rate predictions are annual rates, they must be divided by four to obtain quarterly rates needed for mortgage performance forecasts. Dividing by four is appropriate here because, growth rates are transformed into house price indices by taking the exponential of cumulative growth rates.

ACCOUNTING FOR THE POSSIBILITY OF DEPRESSIONS

Events like the Great Depression do not appear in the data available for regression analysis of the economic environment. To add risk of depression-like events, a stochastic depression generator is added to the VAR and house-price-growth equations. The probability of a depression-like event occurring in any given time series simulation is arbitrarily set to 2.5 percent. That is, one-in-forty simulations are likely to experience a depression event, meaning 25 events in 1000 simulations. When a depression is triggered, changes in unemployment, inflation, and house price growth are patterned after the 1930-1942 experience in the U.S. Unemployment rates increase by 20 percentage points over VAR simulation values in the first four years, and then decline back to their VAR simulation values over the ensuing 8 years. Inflation drops by 7.7 percentage points per year below VAR simulated values for each of 3 years, and annual HPI growth rates drop by 11.1 percentage points below simulated values for each of 3 years. After year three, both inflation and HPI growth return to their simulated values.

Because depression events create deviations from projections of economic conditions, they do not necessarily result in economic environments with the conditions of the Great Depression of the 1930s. Also, their impact on forecasts of mortgage performance outcomes and FHA cash flows depends on when these events actually occur during the 30-year forecast horizon, vis-a-vis mortgage origination. Depression events generated in the last ten years will have little effect on predictions of mortgage performance and subsidy rates.

The occurrence of depression-like conditions is not wholly dependent on these add-on events. In the VAR-and-house-price-growth system of equations, it is possible for a congruence of interest rates, unemployment, and inflation to create recessionary and depressionary housing market conditions apart from this depression-event add-on. However, given the range of economic conditions existing over the historical period available for regression analysis, the rate of naturally occurring housing recessions and depressions coming out of the simulations is too small, without the add-on event generator.

RANGE OF SIMULATED ECONOMIC VALUES

The range of final output values from the VAR and house-price growth equations, for use in predicting mortgage performance and subsidy rates, is illustrated in Figures 2 and 3. These Figures show historical, to-date movements in mortgage interest rates (Figure 2) and national house price growth rates (Figure 3), along with one and 99 percent confidence bounds for the forecast simulation series for each variable.

ADDING LOCAL HOUSING CYCLES

As Federal Reserve Board Chairman, Alan Greenspan, recently noted, housing markets are, by nature, local: "A home in Portland, Oregon is not a close substitute for a home in Portland, Maine, and the 'national' housing market is better understood as a collection of small, local housing markets."[17] Study of Census Division and Metropolitan level (MSA) house price growth rate series published by the Office of Federal Housing Enterprise Oversight (OFHEO) reveals clear cyclical patterns around national trends.[18] The amplitude of those cycles was larger in the 1980s, yet cyclical patterns continued through the 1990s.[19]

The statistical characteristics of these patterns are replicated in two parts. First is a process that generates serially correlated (random) growth paths, and second is a cycle-generating process. Serial correlations are created using the Ornstein-Uhlenbeck (OU)

---

[17] Statement of Alan Greenspan, Chairman, Federal Reserve Board, before the Joint Economic Committee of Congress, April 17, 2002. (http://www.federalreserve.gov/boarddocs/testimony/2002/20020417/default.htm)

[18] These data are available at www.ofheo.gov. The data were studied by Malgorzata Klosek, while a visiting scholar at CBO. She also developed the Ornstein-Uhlenbeck process usage described in this section.

[19] Local cyclical patterns around long-term national trends, was also found by Joseph Gyouko and Richard Voith, "Local Market and National Components in House Price Appreciation," *Journal of Urban Economics* vol. 32 no. 1 (July), 1992, pp. 52-69.

process, which generates exponentially correlated growth rates.[20] The OU process is the continuous version of a distributed lag function, and is modeled here in its discrete form. The discrete formulation mimics daily price movements, from which end-of-quarter values are drawn. The OU process allows for the type of variation seen in the historical data, namely, that some regions and localities may cycle near the national growth rate over time, while others will diverge. Adding a cycle-generating process to this forces each house price series to have more pronounced cyclical movements, creating a self-correcting process for housing booms and busts. The phase of each cycle at the start of the forecasting period is determined by recent deviations in each price series' movements from national house price movements.

Deviations of cumulative local house-price growth from the national level are computed using the zero-mean OU process, $\eta(t)$, which satisfies the differential equation:

$$\eta'(t) = -\alpha \cdot \eta(t) + \sqrt{2}\omega'(t) .$$

Here, α is a time-series correlation parameter (affects half-life of correlations), $\omega'(t)$ is standard Brownian motion, and t represents quarterly time realizations of the Brownian-motion process. A discrete form of the OU process is used to simulate the cumulation of daily fluctuations in house prices:

$$\eta\left(t,(k+1)\right) = \eta(t,k)(1-\alpha/92)+\sqrt{2/92}\cdot N(0,1)$$

[20] G.E. Ornstein and L.S. Uhlenbeck, "On the Theory of the Brownian Motion," in *Selected Papers on Noise and Stochastic Processes*, N. Wax ed. New York: Dover Press, 1954, pp. 93-112. Reprinted from *Physical Review* vol. 36 (September 1, 1930), pp. 823-841.

where k represents each (daily) realization of the process, there are 92 days in each quarter, $N(0,1)$ is a standard normal random variate, and $\eta(t,0) = \eta(t\text{-}1, 92)$. The OU process is seeded by $\eta(0,0) \sim N(0,1/\alpha)$, which assures stationarity in the resulting $\eta(t,k)$ time series. Adding the volatility or amplitude parameter, ε, leads to quarterly realized values of cumulative house price deviations, $\gamma(t)$:

$$\gamma(t) = \eta(t, 92)*\varepsilon$$

The parameters for these processes (α and ε) are chosen separately for Census Divisions and for MSAs to account for higher volatility in the historical MSA house price series (see Table 3). The OU process uses random number generation, so that each specific locality–MSA and Census Division–will undergo the full range of possible deviations from national house price growth, across the 1000 simulated economic environments.

The cyclical effect added to the OU price-growth factor each quarter is:

$$\sigma \cos(\Omega \cdot t + \rho)$$

where σ is the amplitude of the cycle, Ω determines length of cycle (in quarters, t), and $\rho = (0, 2\pi)$ indicates the phase of the cycle at time zero. Cycle phase, ρ, is a uniform random variable whose range is zero to 2π, the length of the cosine wave. The starting phase for each region in the forecast period is set using a uniform random variable within a segment of the full $(0, 2\pi)$ interval. The particular segment is chosen according to the relationship between regional and national house price growth over the most recent five-year period. The difference between average annual growth at the regional and national levels determines whether the individual region will begin the forecast period in an expansion or contraction, and in what part of that cycle phase (See Table 3).

Final Census Division and MSA house price indices used to predict mortgage performance are calculated as the exponential of the sum of national (cumulative) house-price growth rates and local deviations. Use of the exponential function follows from assuming that house prices in any given locality are lognormally distributed, at any given point in time.[21] This assumption underlies the OFHEO house price indices used here for statistical analysis of house price growth rates.

THE TREASURY YIELD CURVE AND DISCOUNT FACTORS

Following budget preparation rules established by OMB, a "basket-of-zeros" approach is used to discount FHA cash flows and calculate net present value subsidy rate estimates. This approach uses implied spot interest rates across the full length of the constant maturity Treasury yield curve. Cash flows in year one are discounted with the one-year spot rate, those in year two with the two-year spot rate, and so on. To be consistent with budget practice in the Executive Branch, the FHA-BSSS obtains discount rates for new budget cohorts from the forecast Treasury yield curve in the first quarter of each calendar

[21] Lognormal price indices follow from normally distributed cumulative house price growth rates, within any given geographic area, over any specific time period.

year. This matches the January timeframe for budget preparation.[22] For existing cohorts, OMB determines the yield curves to be used by federal agencies for discounting cash flows for credit programs. These yield curves are averages across each fiscal year. Yields used for 1992-2002 are replicated in Table 4.

The VAR system of equations predicts only the 3-month and 10-year Treasury yields. All other rates are computed using interpolation techniques. First, the basic shape of the yield curve is determined through relationships of other major points to the 3-month and 10-year yields. These relationships are average, proportional shape factors derived with regression analysis on monthly data, 1980-2000.[23] Equations used to compute them are reported in Table 5.

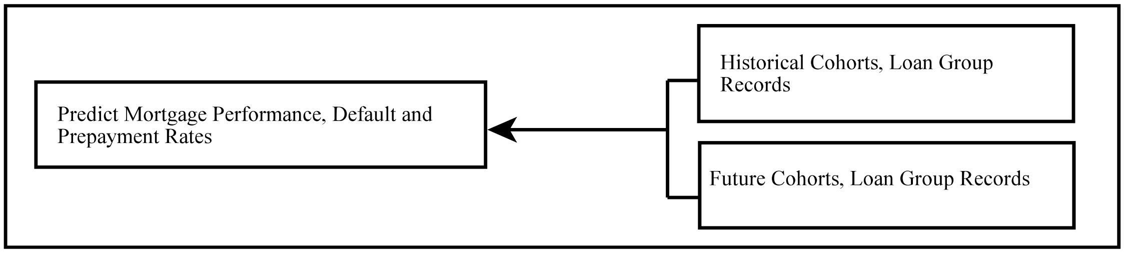
Further interpolations required to complete the entire yield curve are made assuming logarithmic shapes for movements of implied interest rates on sub-intervals of the yield curve, as described in Box 2. These sub-interval rates are called forward rates, and they are computed over successive 6-month intervals between the yield curve points calculated from the formulas in Table 5. Forward rates are first converted into spot-rate equivalents, which are term-equivalent yields that match a series of forward rates over the entire term

[22] Often OMB picks a yield curve from a week in December, rather than January, to give the Executive Branch agencies more time to prepare their final budget numbers. The FHA-BSSS, however, only forecasts data on a quarterly basis and, therefore, cannot pick individual weeks. Instead, first-quarter (January-March) predictions are used to match the timing of the President's Budget. Using the previous fourth quarter (October-December) would have little impact on simulation results.

[23] The relationship of the 30-year (constant maturity) Treasury yield to the 10-year yield is based on analysis of 1985-2000 data.

of interest. The spot rates are then used for discounting cash flows from each year, starting with the year the loan guarantees are made. All cash flows are treated as if they occurred at the midpoint of each fiscal year, so the required spot rates and discount multipliers are for securities with maturities of : 6, 18, 30,...,372 months.

III. MORTGAGE DEFAULT AND PREPAYMENT REGRESSION MODELS



Simulated economic conditions are used to predict future performance of guaranteed loans, throughout the 30-year life of each budget year cohort. The first outcomes to be measured are mortgage defaults and prepayments each quarter. Defaults lead to another sequence of events (discussed in Section IV) to determine net costs. Prepayments can both curtail premium income and trigger premium rebates to insured borrowers.

Established economic theory on mortgage terminations uses a framework of financial options, whereby borrowers are assumed to hold two valuable options imbedded in their home mortgages.[24] The first is a *call* option, which permits borrowers to buy mortgages from lenders for par, at any time, and without penalty.[25] This option identifies the prepayment feature of nearly all home mortgages issued in the United States.

---

[24]Prices of these options are effectively imbedded in mortgage interest rates. Those rates are the sum of the risk-free rate, a credit-risk (default) premium, and an early-termination premium. For a recent exposition of the options approach to mortgage performance analysis see, Deng, Quigley, and Van Order, "Mortgage Terminations, Heterogeneity, and the Exercise of Mortgage Options," *Econometrica*, vol. 14, no. 3 (2000), pp. 275-307.

[25]In options markets, the agreed upon purchase price is the *strike* price. The ability of holders to exercise their options at any time, up to maturity, is a feature associated with what are generally called American options.

The call option is *in-the-money* whenever current mortgage rates are lower than the coupon rate on the outstanding mortgage by an amount exceeding transaction costs of obtaining new financing. Being in-the-money means the mortgage liability–the present value of future payments due on the mortgage–is greater than the par value (outstanding loan balance). When market interest rates fall below the mortgage coupon rate, the present value of future payments on the existing mortgage increase to an amount greater than the outstanding balance. When this situation occurs, it is in the borrower's best interest to obtain new financing at the lower interest rate, using those funds to purchase the original mortgage back from the lender. Exercise of the call option in this fashion is a rational response to changes in market prices, as the homeowner attempts to maximize lifetime consumption flows from expected lifetime income. Optimal exercise timing is a function of expectations of future movements in interest rates–whether they might drift downward or upward–and how long the homeowner expects to stay in the property. There is no limit to how often homeowners can refinance their properties with new mortgages, but there are transaction costs that influence optimal call option exercise.

The second option imbedded in home mortgages is a default option. In most cases the mortgaged property is the sole collateral for the loan, thus it can be in the borrower's best financial interest to *put*, or sell, the mortgage back to the lender when the value of that collateral is less than the value of the mortgage liability. This transaction effectively takes place in property foreclosure, where the borrower exchanges the property for release from the debt obligation. This is not a costless transaction for the borrower because of impairments to future credit and costs associated with moving the household. Yet when the value of the option is large enough to outweigh these costs, homeowners maximize their lifetime consumption flows by exercising this put option.[26]

The value of the put option increases when the call option is in the money. A valuable call option means the effective cost of the mortgage liability is greater than the outstanding loan balance, which only enhances the value of the put option. Thus, as the value of the excess mortgage cost increases (market mortgage rates decline), both the call and the put options become more valuable. Which one, if either, is exercised depends on the difference between property value and outstanding loan balance, and on expectations of future movements in interest rates and house values.[27]

Predicting mortgage terminations is not as easy as just measuring interest rate and house price movements. The actual cost of option exercise varies by loan, property, and household characteristics. In addition, while market interest rates are readily available current values of individual properties are not. Because of these additional considerations, empirical studies of mortgage termination rates use regression techniques that measure

---

[26]Many states permit lenders to pursue deficiency judgments against borrowers to recover their loss in foreclosure. However, FHA has a policy of not pursuing this option except on repeated foreclosures.

[27]In the pure options theory of mortgage borrower behavior, having in-the-money options does not mean immediate exercise. Wealth-maximizing individuals will wait until the expected value of an option is maximized in the present. This means looking for cyclical lows in house prices and interest rates. Not that borrowers can predict these things with great accuracy but, rather, one expects to see greater rates of option exercise when their values are greater. Empirical analysis focuses on probabilities of events, and increasing values of option exercise increase the probability that any given borrower will chose to an immediate exercise.

average response rates of borrowers to changes in option values. Researchers also add to their analyses various loan, borrower, and property characteristics that help identify factors making option values different across borrowers.[28]

GROUPING LOAN RECORDS FOR USE IN REGRESSION ANALYSIS AND FORECASTING

Mortgage default and prepayment forecasting equations used in the FHA-BSSS were estimated with records on over 16 million FHA loan guarantee originations,1975-1999.[29] To expedite data processing and statistical estimation, loans were first aggregated into groups or "cells," based on common values of nine essential characteristics.[30] These characteristics, and the value ranges (classes) used to determine which loans are aggregated together are:

- Mortgage product type (fixed-rate 30-year, fixed-rate 15 year, adjustable interest rate, graduated payment, graduated equity)

---

[28]The regression analysis detailed in this study follows most closely from the following two works: Ambrose, Capone, and Deng, "Optimal Put Exercise: An Empirical Examination of Conditions for Mortgage Foreclosure," *The Journal of Real Estate Finance and Economics*, vol. 23, no. 2 (2001), pp. 213-234; and U.S. Department of Housing and Urban Development, Office of Federal Housing Enterprise Oversight, "Risk Based Capital," *Federal Register*, vol. 64, no. 70 (April 13, 1999), pp. 18172-18192.

[29]These data were provided by FHA from its Single Family Data Warehouse. The request was for loans originating on or after January 1, 1975. The data were re-classified to create a new data warehouse with an organization that facilitates the statistical analysis and forecast simulations described here.

[30]Prior to this aggregation, there are a large number of screening and cleaning routines employed to eliminate bad data and to replace bad or missing data elements using other available information, when possible. Some classifications use characteristics that show up through values of other data fields. For example, FHA contractors traditionally have used groupings of non-standard LTV ratios to label both streamline refinance loans and investor loans. Those same criteria are employed here.

- Mortgage purpose (purchase, refinance)
- Loan-to-Value ratio (80% or less, 81-90%, 91-95%, 96-100%, over 100%)
- Owner type (occupant, non-occupant (investor))
- Interest rate (under 4% to over 18%, grouped in 50 basis point increments)
- Property location (9 Census Divisions and the 25 metropolitan areas where FHA does its greatest volume of business)
- Property price (up to 50% of area median price, 51-100%, over 100%)
- Property age (new, 1-15 years, 16-30 years, over 30 years)
- Loan origination quarter (starting with 1975Q1=1)

Grouping loans according to common values of these nine characteristics maintains essential information for measuring differences in loan performance while, at the same time, the number of records to be processed is greatly reduced.[31] The characteristics are measured as discrete classification variables, producing a finite number of possible combinations of values. Each unique combination of values of the nine variables defines a record. Each record includes the total number and dollar volume of loans matching a given classification, and the numbers of defaults and prepayments occurring each quarter. Average values of some variables, across all loans in the group, are also maintained in

[31]Starting with roughly 16 million loans, the aggregations yield less than 1 million loan groups. For the forecast simulations, property age is dropped as a classification variable, and fiscal year of loan endorsement replaces calendar quarter of origination, yielding just over 300,000 records. Because many loan groups have no loans surviving at the start of the forecast period, only 102,000 loan groups on historical cohorts actually enter the forecast simulations. Approximately 3,000 additional loan groups are added for each future cohort year added to the simulations. Tables A.9-10 show that the number of observations in the regression analyses can still be in the millions. This is because each loan group provides up to 80 quarterly observations to the regression analysis.

each record (e.g, loan-to-value ratio and mortgage coupon rate).

The property location variable identifies whether a loan is in one of 25 Metropolitan Statistical Areas (MSA) or not.[32] The 25 are chosen because they had the highest volumes of FHA insurance business in the 1990s. Those not in one of these MSAs are assigned to their respective Census Divisions. This geographic mapping is used for assigning house price growth rates.

DEFINING MORTGAGE TERMINATION EVENTS

The particular events studied in this analysis are prepayment terminations and default events. Borrower default is defined here as a non-cured 90-day delinquency rather than a loan termination. Defining default as non-cured 90-day delinquencies permits a secondary analysis of whether borrowers lose their homes in foreclosure or not. This second step is important now that FHA has a fully functioning loss mitigation program designed to create workout options. Such options enable many defaulted borrowers to avoid foreclosure and keep their homes and mortgages. Regression modeling of rates of workout offers versus property foreclosure is discussed in Section IV.

---

[32]The MSAs (with their identification numbers) used for location grouping in this study are, in order from largest to smallest FHA volume: Chicago (1600), Riverside, CA (6780), Los Angeles (4480), Washington, DC (8840), Atlanta (520), Phoenix-Mesa (6200), Denver (2080), Minneapolis (5120), Baltimore (720), Dallas (1920), Philadelphia (6160), Detroit (2160), Salt Lake City (7160), Sacramento (6920), Las Vegas (4120), St. Louis (7140), Seattle-Bellvue (7600), New York (5600), Houston (3360), Fort Worth (2800), Indianapolis (3480), Nashville (5360), Norfolk-New Port News (5720), Orlando (5960), and Kansas City (3760).

Borrowers that self-cure 90-day delinquencies by making back payments are treated as having chosen to continue servicing the mortgage rather than defaulting. If they instead sell the property to cure a default, their choice is tagged as a payoff. If neither of these occurs, then the loan is tagged as a (non-cured) default for this analysis.

Loan prepayments arise from both property sales and refinancing. While it would be beneficial to treat these as distinct events/choices, the required data are only available beginning in 1991. Starting the analysis in 1991 would miss the richness of the 1980s economic experience, which included wide swings in interest rates and deep regional housing recessions. Therefore, both types of loan payoffs are combined together and statistical analysis of loan performance begins in 1975.[33]

Loan prepayments due to property sale do not represent call option exercise. They still, however, represent utility-maximizing choices of households, given expected lifetime income, the price of housing relative to other goods, and the relative cost of owning versus renting a home. The incentives for such mobility-induced loan terminations–both defaults and payoffs–are not as easily measured with empirical data, as are the incentives for option exercise, because they rely more upon unknown individual borrower circumstances. Still, it is important to develop regression and forecasting models that use

[33] Refinancings are actually of two types, but data are not available to distinguish them. These types are the pure interest-rate refinancings done to lower monthly payments, and cash-out refinancings, which are often used to consolidate debt. Cash-out refinancings are somewhat insensitive to mortgage interest rate changes because they are driven by needs for large amounts of cash, either to payoff higher-cost consumer debt, or else to pay for large non-housing expenditures.

available data to capture systematic patterns in homeowner mobility.

THE LOGISTIC REGRESSION APPROACH TO ESTIMATING DEFAULT AND PREPAYMENT PROBABILITY EQUATIONS

Forecast equations for mortgage performance in the FHA-BSSS predict probabilities of borrower choices each quarter, given economic conditions and loan characteristics. The statistical analysis used to generate these equations was performed on FHA-insured loans originating between 1975 and 1999, with termination activity followed (quarterly) through 2000 for defaults and 2001 for prepayments. The difference in final observation dates for defaults and prepayments is necessary to give time to determine which 90-day delinquencies were not cured by borrowers.[34] In addition, loan termination activity is only analyzed through the twentieth year of each insurance cohort. After that, there is little available data on loan performance, and such latter-year events will have little effect on subsidy rate estimates. Likewise, very few loans terminate in the first two quarters of loan life, and so these observations also are excluded from the regression analysis. Section V describes how default and prepayment rates are determined in the forecast simulations for very young and very old loans.

Mortgage performance between the third-quarter and twentieth-year of loan life is

[34] In the simulations, all open 90-day delinquencies without a recorded loan workout agreement at the end of FY2002 are considered non-cured defaults that will result in insurance claims during the first year of the forecast period (FY2003).

modeled with pairs of binomial-choice equations. In each equation, the choice is either to make monthly mortgage payments due that quarter, or else enter a termination path. The termination paths are payoff (prepayment) or default. The choice to enter either of these paths is modeled using logistic regressions. Logistic or, simply, *logit* models are consistent with consumer utility/wealth maximization, can be used with longitudinal data (time series on individual observations), easily handle variables whose values change over time, and have a peculiar property making them especially suited for this analysis. That property, *independence of irrelevant alternatives,* means default and prepayment decisions can be modeled separately and then recombined to form an equation used for calculating probabilities of default and prepayment in each period (calendar quarter).[35] Separate default and prepayment estimations are useful for testing different variable specifications for default and prepayment equations. When the probabilities estimated from these equations are applied to grouped data, they generate forecasted rates (and numbers) of terminations each quarter.

To measure major differences in borrower responses to economic environments, separate pairs of regression equations are estimated for: purchase-money fixed-rate mortgages, refinance fixed-rate mortgages, purchase-money adjustable-rate mortgages, and refinance

[35]See C.B. Begg and R. Gray, "Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions," *Biometrika* vol. 71 (1984), pp. 11-18.

adjustable-rate mortgages.[36, 37]

These equations resemble:

$$\log\left(\frac{P_d}{1 - P_d - P_p}\right) = X_d\beta_d + \varepsilon_d \quad \text{for default, and}$$

$$\log\left(\frac{P_p}{1 - P_d - P_p}\right) = X_p\beta_p + \varepsilon_p \quad \text{for prepayment.}$$

Where P_d is the probability of default, P_p the probability of prepayment, and ε_d and ε_p are i.i.d. error terms with logistic distributions. In the FHA regression analysis, the two probabilities are computed as relative frequencies within each loan group, in each quarter.[38] The left-hand-side construction in each equation is the log of the odds ratio for

---

[36]The different risks and risk-preferences of borrowers who choose adjustable-rate mortgages has been discussed and analyzed in: Jan Breukner and James Follain, "The Rise and Fall of the ARM: An Econometric Analysis of Mortgage Choice," *Review of Economics and Statistics*, vol. 70 (February, 1988), pp. 93-102; Charles Capone and Donald Cunningham, "Estimating the Marginal Contribution of Adjustable-Rate Mortgage Selection to Termination Probabilities in a Nested Model," *Journal of Real Estate Finance and Economics,* vol. 5 (1992), pp. 333-357.

[37]Mortgages with shorter-terms (15 years) and those on investment properties exhibit unique termination patterns, but there are too few of them to warrant separate statistical analyses. Mortgage performance forecasts on 15-year mortgages use the estimated equations for 30-year mortgages. The 15-year loans will have smaller predicted default rates and larger predicted prepayment rates than comparable 30-year loans because they accumulate equity more quickly. Investor-property loans have higher default and prepayment rates, which are measured in regression models through dummy-variable scaling factors. Thus, the investor loans are included in the regression estimations along with non-investor loans.

[38]Logit estimation is designed to model "conditional" rates of event choices. That means the probabilities in question are in reference to the number of loans in a group that actually survive and are current at the start of each observation quarter. So, P_d is computed as the number of loans defaulting in a given quarter, divided by the number of loans active at the start of the quarter less the number of loans prepaying during the quarter. Prepayments are removed because the logit model estimates probabilities of events (default or prepay) versus the non-event (continuing mortgage).

each termination choice, which is assumed a linear function of explanatory variables denoted by the *X* vectors. The estimated coefficient vectors, $\hat{\beta}$, provide weights associated with each variable's impact on incentives to enter a termination path. These weights are estimated on the historical data using maximum likelihood techniques.[39]

When estimating probabilities for simulation forecasting, $X_d\hat{\beta}_d$ represent the weighted sum of default incentives, and $X_p\hat{\beta}_p$ the weighted sum of prepayment incentives. *X* values for each loan group include both fixed loan and property characteristics, and time-varying loan and property characteristics that are functions of economic conditions in the forecast period. The weighted sums of incentives to default and prepay are re-combined to calculate the two probabilities of interest:[40]

$$P_d = \frac{e^{X_d\hat{\beta}_d}}{1 + e^{X_d\hat{\beta}_d} + e^{X_p\hat{\beta}_p}} \qquad \text{and} \qquad P_p = \frac{e^{X_p\hat{\beta}_p}}{1 + e^{X_d\hat{\beta}_d} + e^{X_p\hat{\beta}_p}} .$$

Estimated probabilities for each loan group are multiplied by numbers of surviving loans in each quarter to arrive at numbers of predicted defaults and prepayments. These

[39]Such techniques choose the values of $\hat{\beta}$ that maximize the joint probability that the observed choices are the "correct" choices, given the values of the explanatory variables, *X*.

[40]These probability formulas result from rearranging the exponentials of the logit equations shown above.

terminations are then subtracted from beginning-of-quarter survivors to calculate the number of surviving loans used to predict defaults and prepayments in the next quarter.[41]

The statistical regression estimation procedure weights each loan group observation according to the number of surviving loans in the group, at the start of each quarter. Thus, final coefficient estimates ($\hat{\beta}_d$, $\hat{\beta}_p$) match what would be obtained by using individual loan data.[42]

EXPLANATORY VARIABLES USED IN REGRESSION ANALYSIS AND PROBABILITY FORECASTS

Explanatory variables used to estimate the logit regressions (X_d and X_p) and, subsequently, to predict probabilities of events in the forecast simulations, include loan-group characteristics, economic condition indicators, and more direct measures of the financial incentives to exercise imbedded call and put options (Table 6). All variables were chosen

[41] In the forecast simulations, the number of predicted defaulted-loan workouts in each quarter is first netted from the number of defaults so that the new number of surviving loans includes the worked-out defaults. Analysis of workouts is covered in Section IV. In the historical data, loan workouts were not a measurable phenomenon until 1997. Before that time, loan default, as measured here, included property foreclosure and the repurchase of defaulted loans by FHA. That repurchase program (Single Family Mortgage Assignment Program) accounted for up to 25 percent of non-cured defaults before it was eliminated by the Congress in favor of direct loan workout assistance. Other minor sources of loan defaults in the historical data include pre-foreclosure sales of properties, where borrowers receive financial assistance to sell their homes, and voluntary deed transfers. These are both minor programs. In the forecast simulations, deed transfers are ignored and pre-foreclosure sales are treated by adjusting the cost of paying insurance claims on defaults that go to foreclosure. The exact treatment is discussed in Section V.

[42] Loan groups only enter the regression analysis if they start with at least 10 loans in the quarter of mortgage origination. For fixed-rate mortgages, where the longitudinal databases are quite large, a one-in-three sample of loan groups is used in the regression analyses.

for their expect ability to measure differences in incentives to terminate mortgages. Motivations to prepay (call) or default (put) are determined primarily by movements in property values and interest rates, along with the more general economic conditions that affect household mobility. These two choices are substitutes for one another; once one is chosen the other option is no longer available. The result of substitutability and common determining factors is that variables used in the regression equations are mostly the same. There are slight differences in the variables used for default and prepayment, and two additional variables are used only for adjustable rate mortgages.

Non-financial variables, like property price (*PRICE_CLASS*) and property age (*AGE_CLASS*) are important proxies for differences in financial incentives by classes of mortgages when those differences cannot be measured directly. For example, borrowers who purchase older homes are more likely to encounter significant maintenance and repairs costs that could precipitate default (as put option), and those purchasing older homes are also more likely to see these as starter homes, building equity that will permit trading-up at a future date. Thus, property age, though not a direct financial variable itself, measures general class-level incentives tied to unmeasurable financial details and borrower preferences that affect default and prepayment rates.

Most of the variables used in this analysis are in categorical rather than continuous form (see Table 6). This approach has numerous advantages. First, the regression model is less restricted because no functional form is imposed on the explanatory variables. With continuous variables, one must assume something about whether the X's enter in linear or some other form. Without some prior expectation of what that form should be, one is left assuming a linear form and thus assuming that the incentive-weight $\hat{\beta}$'s are constant across all values of each X. Categorical variables, on the other hand, allow the data to speak for themselves with regard to how the response function behaves as X values change.[43] Each category will have its own value of $\hat{\beta}$.

A second advantage is that use of categorical variables eliminates the chance of predicted probabilities reaching unreasonably large levels when *X* values take on extreme values. When continuous variables enter the regression in linear form, the $\hat{\beta}$'s are most accurate when *X* values are near their mean-values across the regression sample, and can produce unrealistic results when *X* variables take on outlying values.[44] With categorical variables, however, the $\hat{\beta}$'s produce fixed response effects for X-values in each defined range.

[43]This approach requires that the underlying continuous variables be placed into value-classes in a systematic and meaningful fashion. In the extreme, categorizing variables leads to kernel estimation where the category sizes are as small as possible so that a non-parametric regression results.

[44]For categorical variable coefficients to give reasonable results, one then must have a wide variability in *X* values in the regression data. This is why the 1980s and 1990s experience are both valuable in estimating a regression model that can be used with forecast simulations of a wide range of economic conditions. During the 1980s, interest rates reached historical peaks, and house prices experiences strong regional cycles. During the 1990s, interest rates experienced large declines while regional house-price cycles continued to appear. While use of categorical variables caps the influence of outlier values of the X variables, the economic simulations used in the FHA-BSSS are designed to provide X-values in the highest and lowest value classes. Thus, we balance limits on the response of borrowers to extreme economic events, we make sure that such events are adequately represented in the forecast simulations.

A third advantage of categorical variables is that the relative magnitudes of termination incentive effects ($\hat{\beta}$'s) can be more easily compared both within and across the *X* variables. Because the scale of categorical variables is identical for all X variables, the relative size of the $\hat{\beta}$'s can be directly interpreted as relative magnitude of effects, both within and across the X variables.

RESULTS OF LOGISTIC REGRESSION ESTIMATIONS

Results of the mortgage performance regressions are reported in Tables 7-10. First are Tables of analysis-of-variance tests that measure the relative influence of each explanatory variable on the estimated probabilities in the regression sample. These are followed by Tables of the regression estimates of the $\hat{\beta}$'s .

The regression equations use an *effects* parameterization. With this approach, the coefficients for classes within each categorical variable measure effects relative to the average across all value classes. Thus, the coefficients across all classes of any one variable sum to zero.[45] The last value-class for each variable is then dropped from the

[45]This is also known as a full-rank or centered-effects parameterization. The alternative, and more common parameterization is the *reference* form. In that specification, the net effect of the coefficients on the (omitted) highest value classes of all categorical variables are jointly imbedded in the constant term of the regression.. The parameterization type affects how the data are organized for the regression and the interpretation of coefficient estimates, but not the actual predicted probabilities that come from the estimated equation.

regression estimation because their coefficients are merely the negative sum of all other (estimated) coefficients. They are, however, reintroduced into Tables 9-10 and are highlighted in italics.

Several coefficients could not be estimated, or not reasonably estimated with the available data. These omissions mostly affect adjustable-rate refinance loan equations for which data are only available starting in 1992. Reasonable values are imputed from estimated coefficients within and across equations. Imputed coefficients are highlighted in the Tables in boldface. The procedures used for imputation are described in Section V.

Analysis-of-Variance Tests

Analysis-of-variance (ANOVA) tests reported in Tables 7 and 8 measure the influence of adding each variable to the regression equation. Test statistics highlight the relative influence of each explanatory variable in measuring differences in actual historical outcomes. Because these rankings can be influenced by which variables had the most variability in the historical period, one cannot say that the same rank ordering of influence will hold in a forecast environment. However, the historical period in this analysis did have good variability in all variables, and simulations of economic environments used for forecasting do replicate that variability. Thus, this analysis point to which factors should be most influential in predicting default and prepayment rates in the forecast period.

Rankings of Variables by Influence in Default Regressions. ANOVA test results emphasize the importance of option-related financial variables. In the default regressions, the number one influence by far is the percentage of mortgages in a loan group likely to have loan balances below their property values (*NEQ_EQ_CLASS*). This is the most direct measure of put-option value, and it is followed in importance by the call-option value variable (*SPREAD_CLASS*).[46] These two variables complete the put option value, which in a direct measure would be the sum of (negative) property equity and excess mortgage value (present value of the future payment stream less the mortgage balance).[47] From there, variable influence across equations starts to diverge, yet the variables which stand out most are property age (*AGE_CLASS*), origination year (*COHORT_YEAR*), and property price (*PRICE_CLASS*).

As seen in Table 9, the property age variable picks-up higher default incentives with new homes, below-average incentives on homes aged 1-30 years, and average incentives on older homes (over 30-years of age). With new homes, builders often capitalize many

[46]The adjustable-rate, refinance-loan default equation deviates from the others at this point. The call-option variable is fourth most influential, behind both property age and property price. This is not too surprising, given that borrowers who refinance into adjustable-rate products have shown a lack of interest in locking into fixed-rate products when interest rates were relatively low. In addition, the call-option value on adjustable-rate mortgages, when it is in-the-money, is less than it is for fixed-rate mortgages because of the interest-rate adjustment features of these loans. Also, this equation only has data beginning with 1992 loan originations, because there was no measurable FHA refinance activity with adjustable-rate mortgages until that year.

[47]Separating these components in empirical/statistical analysis is reasonable because we do not have direct measures of either one. Indeed, direct estimates of the values of these components can only be made using repeated simulations of future (random) paths of interest rates and house prices from positions at the time of observations, and using backward-solving algorithms from the time of mortgage maturity. The variables used here are common representations of option value used in empirical research of mortgage terminations.

buyer transaction costs into the price of the home. This practice makes it easier for the buyer to purchase new homes, but it also increases the mortgage balance. Because the homeowner cannot recoup these costs upon resale, new home values exhibit lower rates of appreciation than do previously existing homes, for a period of time. This then leads to higher put option values on these mortgages. The $\hat{\beta}$'s also indicate higher put option value on older homes versus younger (but not versus new) homes because of the higher incidence of high-cost repairs and maintenance in the early years of mortgage life. These costs, when they exist, both lower the effective value of the property and require cash outlays that the homeowner may not be able to provide.

The loan origination year/cohort variable is a catch-all variable that measures differences in the quality of loans as underwriting standards tighten and loosen over time. In some years, these effects can be quite dramatic. In the early 1980s, when interest rates were high, lenders initiated many creative financing arrangements to qualify borrowers for mortgages. These all entailed hidden costs which effectively increased loan-to-value ratios of mortgages to property, increasing put option values. The mortgage industry, led by Fannie Mae, Freddie Mac, and FHA, eliminated many of these practices in 1985 and 1986. FHA, in particular, restricted the use of prepaid interest buydowns and tightened standards for other loan quality characteristics.[48] Then, in 1995, FHA took deliberate steps to lower underwriting standards to attract more first-time and minority homebuyers.

[48] Mortgagee Letter 86-15, *Change to Single Family Programs*, August 8, 1986.

These changes included broader acceptance of alternative sources of income and downpayment, and the use of non-standard means of verifying borrower creditworthiness.[49] The result has been a steady increase in the underlying default incentives of successive insurance cohorts, as shown in the cohort year coefficients in Table 9.

Property-price class is another variable which may be an indicator of a number of factors contributing to differences in default incentives. The estimated coefficients in Table 9 show an almost symmetrical relationship between default incentives on low-valued properties (below 50 percent of area median price) and on high-valued properties (above 100 percent of area median price), with those at the low-end being above and those at the high-end below average. The low-value class represents roughly the bottom 25 percent of homes in a given market. Homeowners in this price range will have lower income levels, have fewer financial resources to draw on in emergency situations, and may be more prone to job loss during economic downturns. In addition, the lower-and upper-ends of the housing distribution are sub-markets with less activity and thus less stable house prices than the broad mid-section of the distribution. FHA does not insure homes at the upper-end, as mortgage limits set by FHA are generally somewhere between 95 and 150

[49] Mortgagee Letter 95-7, *Single Family Loan Production - Revised Underwriting Guidelines and Other Policy Issues*, January 27, 1995.

percent of area median price.[50] However, FHA does a significant amount of its business in the low end of local housing markets.

Rankings of Variables by Influence in Prepayment Regressions. The loan payoff/prepayment regressions (Tables 8 and 10) use the same explanatory variables as do the default regressions, with just two exception: first, the probability of having at least 20 percent positive equity (*POS_EQ_CLASS*) replaces probability of negative equity (*NEG_EQ_CLASS*); and, second, the slope of the Treasury yield curve (10-year to 1-year) is added as a new determinant. The probability of positive equity variable is more appropriate here because it measures the ability to pay selling costs and move, to invest in a more expensive trade-up home, refinance into the conventional market, or do a cash-out refinance. The yield curve slope indicates changing incentives to refinance between fixed- and adjustable-rate mortgages as economic conditions change.

ANOVA results show that call-option-value (*SPREAD_CLASS*) is the dominant influence on fixed-rate loan payoff rates. It is roughly equal in predictive power with property equity (*POS_EQ_CLASS*) for adjustable-rate (home) purchase mortgages, and second to origination-year effects (*COHORT_YEAR*) for adjustable-rate refinance loans. The lesser

---

[50] FHA establishes county-by-county mortgage limits within statutory bounds. By law, FHA mortgage amounts must be at least 48 percent of the mortgage limit for Fannie Mae and Freddie Mac (the so-called conforming loan limit). They also can be no more than 87 percent of that same limit. For calendar year 2002 these limits were $144,336 and $261,609. FHA receives petitions from localities to adjust their mortgage limits above the statutory minimum, based on evidence that the area median price is above the lower statutory bound. FHA can adjust county-level loan limits up to the minimum of 95 percent of area median house price or the upper statutory bound.

influence of the call-option value in adjustable-rate mortgage equations results from the smaller incentive these borrowers have to refinance because, in any given quarter, many borrowers are eligible for the annual allowable and automatic interest-rate adjustment of up to one percentage point.

Property equity (*POS_EQ_CLASS*) is an important determinant of prepayment rates in all equations. It signals the abilities of borrowers to sell their homes for mobility or trade-up reasons, and it also indicates potential for cash-out refinancing. Both adjustable-rate mortgage prepayment equations are also strongly influenced by the current market rate for fixed-rate mortgages (*CUR_MKT_RATE*). When this rate is low, borrowers with adjustable-rate mortgages have incentives to eliminate the future risk associated with such instruments and secure fixed payments.

A more detailed discussion of each variable and its estimated regression coefficients follows in the next section.

Discussion of Coefficient Estimates

Nonlinear models, like the logistic, do not lend themselves to direct interpretation of how explanatory variable values influence loan termination rates. As discussed earlier, the estimated Beta coefficients are weighting factors that indicate the relative influence of each explanatory variable on the log-odds of a particular termination choice. Therefore they are discussed as default and prepayment incentive effects. Actual changes in probabilities resulting from changes in variable values require a complex manipulation of the coefficient estimates and variable values.[51]

In this section, the estimated coefficients reported in Tables 9-10 are reviewed for direction of effects (positive or negative), relative magnitudes of effects (for categorical variables), and the reliability of the measured effects (levels of statistical significance). As each variable is introduced, some explanation of why it is included in the regression equations is also provided.

Default Regression Results: Loan Group Characteristics.

Loan-to-value ratio. Downpayments made at loan origination are an indicator of household abilities to generate liquid assets and avoid default. Researchers have also found that borrowers who make larger downpayments tend to have stronger credit histories.[52] Fixed-rate loans show greatest credit quality in the 91-95 percent *LTV_CLASS*.[53] The best performing adjustable-rate refinance loans are those in the 81-90 percent *LTV_CLASS*, and the best performing adjustable-rate purchase loans are in the

---

[51]The change in event probability associated with changes in explanatory variables are also dependent on the starting values of those variables. Thus, one cannot say, for example, that the increased in the probability of default for borrowers in *LTV_CLASS*=3 (91-94%) over borrowers in *LTV_CLASS*= 2 (81-90%) is a constant value. It depends on the values of all other variables for the borrower (or loan group) in question.

[52]Anthony Pennington-Cross and Joseph Nichols, "Credit History and the FHA-Conventional Choice," *Real Estate Economics*, vol. 28, no. 2, 2000 (summer), pp. 307-336.

[53]Though the distinction appears small, Pennington-Cross and Nichols, *Ibid*., show data that also indicates FHA borrowers in the 91-95 percent LTV class having the highest overall credit quality of all FHA borrowers. Those researchers mix all mortgage types together.

combined 81-95 percent *LTV_CLASS*(es).

The result for fixed-rate loans stands out because it does not conform with options theory: loans with larger downpayments (lower *LTV_CLASS*) should have lower incentives to default, and will find the put option in-the-money less often than borrowers with smaller downpayments.[54] However, among FHA insured loans, borrowers with 10 percent and larger downpayments have other financial weaknesses that prevent them from obtaining less costly conventional financing.[55]

There are relatively few loans with LTV ratios of 80 percent or less, so they are not included in the statistical analysis. In the simulation forecasts, these loans receive an implied *LTV_CLASS* effect value of zero (no positive or negative effect). They have small default rates because of their larger equity positions, both in the historical data and in the forecast simulations.

Price class, Property Age Class, Owner type, and Mortgage Age Class. All of the default regressions show peaks in the underlying time patterns (*AGE_CLASS*) of default rates in

[54]This assertion holds in this context because the regression models control for locality specific house price movements.

[55]This adverse selection problem, whereby FHA insures loans that perform worse than would be expected by observable measures, is mostly an issue with fixed-rate loans. FHA insurance pricing is more attractive for adjustable-rate mortgages, compared with private mortgage insurance, making FHA more competitive for higher credit-quality borrowers who want adjustable-rate products. In addition, there is less-upside default risk from FHA adjustable-rate mortgages because their maximum annual rate changes are only one percent, compared with a standard two percent annual change "cap" on conventional adjustable-rate mortgages.

either the 3-4 or 5-6 year time period, and strong declines after year 10. The peak periods correspond with a confluence of rising rates of household mobility with insufficient time for property appreciation to provide home equity for paying selling and moving costs. By year ten, loan amortization has reached roughly 10 percent, so as long as property value has not declined, there will be enough equity to pay the transaction costs of selling and moving and default rates should drop dramatically.

Investors (*INVESTOR*=YES) have measurably higher default incentives than do owner-occupants, resulting from the risks they take with occupancy and renter-induced property depreciation. Owners of lower-priced properties (*PRICE_CLASS*=LOW) have higher incentives than do owners of higher-priced homes (*PRICE_CLASS*=HIGH); and borrowers purchasing new homes or refinancing them in the first year (*PROP_AGE_CLASS*=NEW) have higher default incentives than do other borrowers. These last two effects were discussed above in the context of the ANOVA Tests.

Cohort-year effects. FHA underwriting policies have undergone many changes over time. Two of the most important in the study period occurred in 1986 and in 1995. These were discussed above, with the ANOVA test results. Compounding the effects of the 1995 changes was a reduction in minimum downpayment requirement, under the rubric of *downpayment simplification*.[56] The principal effect of this "simplification" was to permit a minimum three-percent downpayment for all FHA loans. Previously, there was a tiered structure that required three percent on the first $25,000, five percent on the next $100,000, and 10 percent on any amounts above $125,000.

The trend of declining quality in insurance cohorts since 1995 was recently confirmed in a report by the U.S. General Accounting Office (GAO).[57] That report cites both reduced underwriting stringency and problems in California as reasons for increased default rates on insurance cohorts written between 1995 and 1999. It is possible, however, that this trend may have stopped in 2001, when FHA lowered its insurance premiums to levels that are now highly competitive with private insurance for borrowers with 5-9 percent downpayments, and much less expensive than private insurance for borrowers with 3-4 percent downpayments. FHA premiums are also much lower than private insurance for adjustable-rate and graduated-payment mortgages, and for so-called A-minus borrowers whose credit is just below prime, conventional quality. Evidence that the deterioration in loan quality in the late 1990s may have stopped comes from a recent shift in the loan-to-value distribution of FHA loans. In the first quarter of fiscal year 2002 (the last data used in this analysis), there was a shift in the distribution of loans by loan-to-value class, with

---

[56]Congress made this change in the FY1999 budget for the Department of Housing and Urban Development (HUD), and FHA implemented it in Mortgagee Letter 98-29, *Single Family Loan Production - Mortgage Calculation Simplification*, October 22, 1998.

[57]General Accounting Office, *Mortgage Financing Changes in the Performance of FHA-Insured Loans*, GAO-02-773 (July 10, 2002).

over five percent moving from the above-95-percent loan-to-value class to the 91-95 percent class.

One additional, important change in underwriting occurred for adjustable-rate mortgages (ARMs) in 1998. At that time, FHA eliminated the use of "teasers" on ARMs.[58] Teasers are interest-rate discounts in year one that allow borrowers to qualify for loans with monthly payments lower than they are likely to be in the future. Prior to 1998, average teasers were one percentage point. That one percent could make a big difference in qualifying borrowers for loans, but it also increased the risk of default when borrowers could not make the (expected) higher payments in the future.[59] Eliminating teasers led to a dramatic drop-off in ARM originations in 1998. ARMs, as a percent of all FHA originations, dropped from over 30 percent (of dollar volumes), to just five percent in 1998, and have stayed below that level since then.

Default Regression Results: Economic Conditions.

House-price-cycle stage. House-price-cycle stage is added to the regression equations for two reasons. First, the options theory of default suggests that borrowers looking to

---

[58]Mortgagee Letter 98-1, *Single Family Loan Production - Underwriting Adjustable Rate Mortgages, Interest Buydowns, Homeownership Counseling and Other Credit Policy Issues*, January 2, 1998.

[59]For example, in an environment where initial rates on ARMs are six percent, a borrower that would qualify for a $125,000 mortgage would, with the same monthly payments, qualify for a $139,600 mortgage with the teaser-discounted rate of five percent. If interest rates were stable and the mortgage rate adjusted back to six percent at the end of year one, then the borrower's monthly payment would increase by $86 dollars. While all borrowers who take out ARMs are taking interest rate risk, the ARM borrower that qualifies because of the teaser is taking an added risk that even in an environment of stable interest rates, they will experience increased payments in year two.

exercise put options on their mortgages will want to do so when a cycle is approaching a trough.[60] At that point, the value of the option is at its maximum.[61] Regression results indicate that incentives to default do increase during housing recessions (*CYCLE_STAGE*= { 1,2}) and, for fixed-rate-refinance loans, continue into the early stages of expansions (*CYCLE_STAGE*= 3).[62]

A second reason for including a cycle-stage variable is that, in addition to price declines, housing recessions cause illiquidity. Housing recessions are caused by contractions in area employment, which lead to significant reductions in home buying activity. This decrease in market liquidity means that many homeowners cannot sell their mortgaged properties in a reasonable time frame without experiencing a substantial financial loss. Difficulty in selling properties increases incentives to default, especially for borrowers

---

[60] The application of this theory to empirical investigations of mortgage performance across housing-price cycles is discussed in Brent W. Ambrose, Charles A. Capone, Jr., and Yongheng Deng, "Optimal Put Exercise: An Empirical Examination of Conditions for Mortgage Foreclosure," *Journal of Real Estate Finance and Economics* vol. 23 no. 2 (September, 2001), pp. 213-234.

[61] In pure, theoretical, mortgage-pricing models, optimal default occurs when the value of exercising the option today is greater than the present value of option exercise across all possible future dates. In empirical analysis, however, given that housing cycles average around ten years in length, and given the larger levels of loan amortization that exist after the tenth year of mortgage life, when a housing downturn does exist, the current cycle is the only one of interest for modeling put option exercise. Borrowers will not necessarily anticipate cyclical troughs with great accuracy. Yet, to the extent there are borrowers who desire to exercise put options on their mortgages, there should be greater default activity around the trough of the cycle, as economic indicators start to point to potential near-term stabilization in the housing market and an up-coming turning point. This effect is independent of the dollar size of individual property negative equity positions: all borrowers with in-the-money put options, who desire to exercise those options, will benefit most by doing so near the trough of the housing cycle. Therefore, house-price-cycle stage is an important indicator variable distinct from property equity (*NEQ_EQ_CLASS*).

[62] Stages 1 and 2 represent the first and second halves of house-price downturns, measured by the actual price declines. Stage 3 is that period of time in which prices are rising but remain in the range previously defined as Stage 2.

who may need to move to a new area to obtain work.

National unemployment rate. Coefficient signs for the national unemployment rate variable (*US_UN_RATE*) are consistently negative across all four default equations. This counter-intuitive result–that increases in the national unemployment rate decrease default incentives–also has been found in other mortgage research.[63] Note that this effect is in addition to the local economic conditions captured in both the probability of negative equity (*NEG_EQ_CLASS*) and house-price-cycle-stage (*CYCLE_STAGE*) effects. Thus, the national unemployment rate effect indicates a lack of household mobility when the national economy is in recession and, likewise, an increase in mobility when jobs across the nation are plentiful. When borrowers cannot move to other areas to obtain new work, they will more likely find ways to maintain their current homes and mortgages even if the put option is in-the-money.[64]

Market shift. There were many major changes in the mortgage market during the 1990s. Because it is not possible to account for each one, a shift variable is added for all observations starting in 1993Q1 (*MKT_SHIFT*). This variable shows a measurable

---

[63]Donald Cunningham and Charles Capone, "The Relative Termination Experience of Adjustable to Fixed-Rate Mortgages," *Journal of Finance* vol. XLV, no. 5 (December, 1991), pp. 1687-1703. That study used regional unemployment rates and local house price indices.

[64]This mobility constraint introduces the issue of what researchers call "sub-optimal" option exercise. If an unemployed worker can only find work in another city, then he/she may exercise the put option even if there is positive equity, because the net equity position in the property is negative after selling costs. If, instead, the homeowner were to find work locally, then the put option would not be exercised either because it is not directly in-the-money (positive equity before property selling costs) or else the timing of option exercise is not yet optimal.

reduction in default rates for fixed-rate mortgages as refinancing became an easier option in the 1990s (*MKT_SHIFT*=ON).[65]

One of the most significant changes for borrowers was the movement toward zero-cost refinancing that occurred in the 1993-1994 refinance period. Along with this was a change by the private mortgage insurance industry to monthly premium payments, and the opening of the conventional market to accept refinance originations with less than a 10 percent downpayment.[66] These and other factors ushered in an era of frequent, if not regular, refinancing of homes. This shift in market operations had the effect of lowering the role of call-option value in determining put-option value.

To see this result, note that FHA allows refinancing of insured mortgages without property appraisals, so that even borrowers with negative equity positions can reduce their mortgage liability and thus lower put-option values. However, if there is no appraisal to prove sufficient property value, borrowers must pay cash for closing costs on the new loan, rather than financing them in the insured mortgage. Changes in the mortgage market that greatly reduced closing costs and eliminated the need for cash at closing allowed more FHA borrowers to exercise refinance call options rather than to see the value of

[65]The opposite effect shown for ARMs implies nothing of importance because ARM volumes did not reach measurable levels until 1992, while the shift variable is turned on in 1993.

[66]Changes in underwriting rules and costs in the conventional market have marked effects on FHA prepayments. Borrowers most often use FHA for qualifying for their first home, and then refinance with FHA only in the early years of their mortgage, while property equity is still relatively low. There are virtually no refinances of FHA loans into new FHA loans after the tenth year of being in a home.

unattainable call options enhancing the value of put options.[67]

Market rates on fixed-rate mortgages: ARM effects. Fluctuations in the long-term mortgage rate (*CUR_MKT_RATE)* reflect market expectations of future movements in short-term rates. Thus, higher long-term rates signal to ARM borrowers that their payments are likely to only continue to increase in the future. For both purchase and refinance equations, upward movements in *CUR_MKT_RATE* create increasing incentives to default as it increases the present value of the mortgage liability. The measured effect in the regression equations is four times as large for purchase mortgages as it is for refinance mortgages. This variable is the predominant ARM-specific effect in the purchase equation, indicating larger anticipated rises in mortgage payments in the future causes borrowers to look for less costly housing alternatives in the present.

Default Regression Results: Financial Incentives

Probability of Negative Equity. The probability of negative equity (*NEG_EQ_CLASS*) variable is the most direct measure of financial incentive to default, and it has the largest influence on default incentives (See Table 7). For loan groups used in this analysis, the probability represents the expected percentage of loans in a group with in-the-money put option values (absent the call option effect) at each point in time. It is calculated with three pieces of information: a to-date price-index level, representing house-price

[67]The primary means for eliminating cash requirements on new loans, without adding the transaction costs to the loan balance, is by adding a premium on the mortgage interest rate of 0.25 to 0.50 percent.

appreciation (or depreciation) since loan origination; amortized loan-to-value (*LTV*) ratio, representing the outstanding loan balance as a percent of the original house price; and a measure of the variance of individual property appreciation rates around the index level. The three are combined to yield a standard-normal statistic that is used to measure the probability that the value of any one property could have fallen enough, compared to the local market, to eliminate all market-generated and owner-invested equity.[68] This is the probability that there could be to-date depreciation of value on the property in question, and that it could be large enough to create a negative equity position for the home owner.

The house-price series and associated volatility measures are produced by OFHEO on a quarterly basis. While not based on FHA loans, they are based on Fannie Mae and Freddie Mac loan purchases, and so represent the broad mid-section of all housing markets. As discussed above, under *Grouping Loan Records for Use in Regression Analysis and Forecasting*, separate house price indices produced by OFHEO are used to calculate *NEQ_EQ_CLASS* for the 25 MSAs where FHA business volumes are largest, and for the nine Census Divisions. Loans not in one of these MSAs are assigned price indices from their respective Census Divisions. The Census Division indices produced by

---

[68] The standardized normal variable defined in the text is: $\dfrac{\ln\left(\dfrac{LTV_t}{HPI_t}\right)}{\sigma_t}$, where HPI is the house price index, *LTV* the loan to value ratio, and σ the square root of the variance of cumulative house price growth rates around the mean index level. All values are as of the time (t) of observation. This statistic is used as the lower bound on an integration of the standard normal density function to compute the probability that an individual property underlying a given mortgage could have depreciated in value enough to eliminate all equity.

OFHEO are, however, first adjusted by population weights to extract out the effects of price changes unique to MSAs within their boundaries.[69]

Interest rate spreads. The interest rate spread (*SPREAD_CLASS*) effects are in line with predictions of option-pricing theory. When spreads are large and negative (market rates below the note rate), higher default incentives result; when they are large and positive, smaller default incentives exist.

When market interest rates are high relative to the existing mortgage contract rate, borrowers are more willing to service their existing mortgages, knowing that their present housing is cheap relative to market-rate alternatives. Also, because FHA loans are fully assumable by new property owners, borrowers can essentially sell the value of a below-market interest rate and thus capture the value of the negative-call option in sales price. In both cases, the value of the put option is reduced because the value of the mortgage liability is less than the outstanding loan balance.

Burnout variables: number of refinance opportunities and new opportunity flag. For many years, mortgage researchers have been concerned with a phenomenon called

---

[69]The adjusted Division price series are computed as:

$\widetilde{HPI}_d = \left(\frac{1}{1-\sum p_m}\right)\left(HPI_d - \sum p_m HPI_m\right)$, where HPI_d is the OFHEO house price index for Census Division, d, p_m is the ratio of MSA population to Division population, and HPI_m is the MSA price index. The summations are over all MSAs within a particular Division. Population weights are from the 1990 Census.

burnout. This is where pools of mortgages get *burned out* by successive refinance waves, so that remaining loans are less sensitive to future refinance options. Borrowers whose loans remain in the pools after these periods may be expecting to move soon anyway,[70] others may be overwhelmed by the mortgage qualification and settlement processes, and still others may have impaired credit and not qualify for a new loan.[71] This last category of borrowers left in burned-out pools is of most concern to mortgage investors. Their presence implies that one can expect higher default rates from survivors in burned-out pools. Two variables are used to test for burnout effects: *BURN_SUM* and *NEW_REFI*.

BURN_SUM is the more important of the two measures (see Table 7). It counts the number of past quarters in which market interest rates were at least 200 basis points below the contract rate on existing mortgages.[72] These were times when refinancing was clearly in borrowers' best interests. The regression results show positive burnout effects (increased default incentives) for adjustable rate mortgages, but negative effects for fixed-rate mortgage borrowers. It has long been argued that borrowers who take out ARMs are less risk-averse than those who use fixed-rate mortgages, meaning that they are more

---

[70]Gordon Crawford and Hsiu-Wen Wu. "On the Rationality of Non-Refinancing Behavior," unpublished manuscript, presented at the 1998 ASSA/AREUEA Annual Conference, Chicago, IL. Washington, DC: Fannie Mae, December 1997.

[71]Stavros Peristiani, Paul Bennett, Gordon Monsen, Richard Peach, and Jonathan Raiff, "Credit, Equity, and Mortgage Refinancings," *Federal Reserve Bank of New York Economic Policy Review* vol. 3 no. 2 (July 1997), pp. 83-99.

[72]A variable that counted the number of times the mortgage spread was 100 basis points or more was also tested in the regressions, but it proved to be a less valuable predictor of default and prepayment incentives than the deep-in-the-money 200 basis point variable.

willing to chance default and foreclosure.[73] The result found here indicates an increased potential for ARM borrowers to impair their credit in ways that make qualifying for a mortgage refinance more difficult. The negative effects found in fixed-rate loan equations suggest that these concerns are not important there. Something else is going on with these borrowers, perhaps as suggested by the authors cited above.

The second burnout variable is *NEW_REFI*, which is a flag (dummy) variable turned on when: 1) the current period interest rate spread is at least -1.50 percentage points, and 2) the spread was smaller (market interest rates higher) than -1.50 percentage points in lagged quarters three through seven. The intent is to capture whether borrowers are in a new refinance period, with enough lapsed time from the previous period to repair any damaged credit and reverse any "burnout" type problems of poor credit quality in a surviving loan pool. Such a reversal would be indicated by negative coefficients for *NEW_REFI*. The regression results confirm that there is some reversal effect for purchase-money ARMs.

The combined negative coefficients on *BURN_SUM* and *NEW_REFI* in the fixed-rate purchase equation suggest that poor quality borrowers leave FHA mortgage pools quickly, leaving higher quality borrowers in "burned-out" pools. The strong influence of the call-option variable (*SPREAD_CLASS*) on default rates, combined with these effects,

---

[73] Breukner and Follain , op. cit.; Capone and Cunningham, op. cit.

suggests that to the extent interest rate effects enhance put option values, they do so independent of borrower credit-quality concerns.

ARM payment adjustment. Because adjustable-rate mortgages have changing payment amounts over time, they face additional sources of incentives to default, beyond those faced by fixed-rate, fixed-payment mortgages. The primary variable used to capture ARM-specific effects is the relative size of the payment adjustment in the current year (*PMT_ADJ_CLASS*). It is the percentage increase or decrease in monthly payments in the present year, as compared with the previous year.[74] Larger positive adjustments should increase default incentives as payment burdens increase, and negative adjustments should decrease default incentives.

Regression results indicate that while this variable does not represent a major influence on ARM default incentives (Table 7), it is relatively more important for refinance than for purchase mortgages.[75] The coefficient signs and sizes for refinance loans are as expected, except for the largest payment-reduction class. The positive default response to large negative adjustments for all ARMs likely indicates that large declines in short-term interest rates are monetary policy responses to recession-level increases in national unemployment rates. This unemployment-enhancement effect overwhelms what

[74] This payment adjustment is based only on the principal and interest components of the mortgage payment.

[75] Following the rank order of ANOVA (Wald) test statistics in Table 7, payment adjustment is the fifth most significant variable influencing ARM refinance-loan default, but the ninth most important variable for determining ARM purchase-loan default incentives.

otherwise should be a decrease in default incentives strictly based on payment burdens. This effect clearly dominates all payment-adjustment size effects for purchase loans, though not for refinance loans.

Prepayment Regression Results: Loan Group Characteristics.

Loan-to-value ratio. Differences in prepayment incentives by original *LTV_CLASS* are limited and mixed. While this variable has some predictive power in both default and prepayment equations, it is never among the top five influencing variables. As an indication of the abilities of borrowers to raise cash to manage financial set-backs, this variable is more important in the default equations. Here, however, it only provides some indication of differences in borrower mobility rates because borrowers can refinance with FHA regardless of their equity position.

Price class, Property Age Class, Owner type, and Mortgage Age Class. Peak age-related payoff incentives (*AGE_CLASS*) are in the 3-4 year age bracket for borrowers with fixed-rate mortgages, and in the 1-2 year age bracket for those with adjustable-rate mortgages. *AGE_CLASS* effects stay larger longer for fixed-rate mortgages, which is consistent with the hypothesis that more mobile borrowers choose ARMs.[76]

Investors (*INVESTOR*=YES) should be less mobile than owner-occupants, with lower

[76] See Bruekner and Follain, *op. cit.*, and Capone and Cunningham, *op. cit.*

payoff incentives. In the FHA data, there is little evidence either way. The investor effect is small but positive in the purchase-mortgage equation, and smaller and negative in the refinance equation. FHA policy is to not insure investor ARMs, so there are no effects to measure there.

The largest property age effects are for refinance loans, where borrowers who are refinancing new homes (*PROP_AGE_CLASS*=NEW)–refinance in the first year of owning the home--continue to have higher motivations to prepay again in the future.

Cohort-year effects. While cohort-year effects (*COHORT_YEAR)* for loan default are quite clearly associated with major changes in underwriting procedures, no such patterns exist for prepayments. Capturing idiosyncratic elements associated with each cohort of business increases the efficiency of estimates on more pertinent loan characteristics and economic environment coefficients.

Prepayment Regression Results: Economic Conditions

House-price-cycle stage. House-price-cycle stage (CYCLE_STAGE) models local economic cycles. As such, the regression results reveal increased incentives for household mobility during the first half of local recessions (*CYCLE_STAGE=1*), as laid-off workers find work elsewhere and sell their homes. Mobility is then depressed as the cycle approaches and passes through the trough (*CYCLE_STAGE=*{2,3}). During these stages, declines in housing market activity make it more difficult for laid-off workers to sell, even if they do take jobs in other areas. Thus, we see declining incentives to prepay that mirror increased incentives to default.

The largest effects are for ARM purchase mortgages, with much larger payoff incentives when *CYCLE_STAGE=1*, and much lower incentives when *CYCLE_STAGE=3* (initial phase of expansion from trough). ARM refinance loans do not show any statistically valid results here because the data sample for that regression does not have good data on transitions of loans through house-price-cycle-stages.

Market rates on fixed-rate mortgages: ARM effects. Current market interest rates for fixed-rate mortgages (*CUR_MKT_RATE*) are the third most influential variable in both the ARM-purchase and ARM-refinance regressions (Table 7). Low rates incent ARM borrowers to refinance into fixed-rate products, and higher rates increase their incentive to stay with the ARM.

National unemployment rate. As discussed earlier for default regressions, increases in the national civilian unemployment rate (*US_UN_RATE*) depress overall borrower mobility, which then decreases incentives either to default or to prepay mortgages. The largest prepayment effect is in the ARM purchase-mortgage equation, where the coefficient is more than twice the size of those in other equations. Overall, as seen in the ANOVA test results, national unemployment rate changes are a less important influence on prepayment rates than they are on default rates.

Yield curve slope. The slope of the Treasury yield curve (*YLD_SLOPE_CAT*) reveals information about incentives to refinance into adjustable rate products, independent of the value of refinancing into a (new) fixed-rate loan. Regression results show that nearly all of the positive incentive to refinance is associated with steep yield curves, with slopes over 200 basis points, and especially over 300 basis points (*YLD_SLOPE_CAT*=5) . In such environments, new ARM loans can both reduce monthly payments for a number of years, and have low lifetime interest rate caps.

Not only are new FHA ARMs attractive steep yield-curve environments, but conventional market ARMs with initial fixed-rate terms of 3, 5, or 7 years are also attractive to homeowners. FHA recently received statutory authority to offer such intermediate-term products. When they become available, those products should decrease loss of business to the conventional market during steep yield-curve environments.[77]

Market shift. The *MARKET_SHIFT* variable was discussed earlier for the default regressions. Its effects on prepayment incentives are all large and they rank higher in overall influence (ANOVA Wald tests) than do the *MARKET_SHIFT* effects in the default equations.

---

[77]FHA loans with initial periods of fixed interest rates, followed by annual adjustments, are called hybrid ARMs. Lenders are not yet expecting that FHA hybrid ARMs, under current law, will contribute much to FHA origination volumes because the interest-rate adjustment caps at the expiration of the fixed-rate term are too restrictive. The lending community is working on technical changes to the implementing legislation that will make these loans more attractive to mortgage investors.

Prepayment Regression Results: Financial Incentives.

Probability of (20 percent or higher) Positive Equity. This variable (*POS_EQ_CLASS*) is analogous in construction to the probability of negative equity variable (*NEQ_EQ_CLASS*) used in the default regressions. Here, however, it is not a direct option-value variable but rather an indication of sufficient home equity for property sale or cash-out refinancing. It has an important influence on overall prepayment incentives.

Interest rate spreads. As discussed earlier, prepayment incentives are highly influenced by the value of the call option (*SPREAD_CLASS*).

Burnout variables: number of refinance opportunities and new opportunity flag. Analysis here follows the earlier discussion of burnout issues as they relate to the default regressions. The negative coefficient on *BURN_SUM* in the prepayment equations likely reflects the existence of a core group of homeowners who do not respond to call-option incentives. Because FHA has a large presence in the lower end of housing markets, there is a core of borrowers for whom the financial benefits of refinancing can be small or nonexistent because of low loan balances, even when the call option appears to be deeply in-the-money for mortgage borrowers in general. However, the positive coefficients on *NEW_REFI=ON* show some of the same increase-in-credit-quality effect seen for this variable in the default equations.

ARM payment adjustment. Annual payment adjustments (*PMT_ADJ_CLASS*) can influence borrower decisions to prepay, as well as to default. For ARM purchase mortgages, large payment adjustments–either negative or positive–are associated with larger prepayment incentives. Annual interest rate changes on FHA ARMs are limited to one percentage point, so large downward movements in market interest rates only yield a corresponding decrease in mortgage payments if the borrower refinances. That appears to be what is seen in the largest (positive) coefficient being on *PMT_ADJ_CLASS*=1, where downward payment adjustments are larger than 5 percent. When rates drop precipitously, borrowers can get an even larger reduction in payments by refinancing.

USE OF REGRESSION EQUATIONS IN FORECAST SIMULATIONS

In the forecast simulations, data input into the logistic probability equations for default and prepayment are loan-group characteristics and forecasted economic conditions (see Section II). Economic conditions change quarterly and have different trajectories in each of the simulation paths. The probability equations are used to predict termination events only between the fourth quarter of loan-life and the end of the twentieth year of each cohort. Outside these bounds, quarterly default and prepayment rates are predicted as here.

The regressions described earlier in this section also have missing coefficients on some categories of the economic factors, due to relatively short loan-data histories. Most of these are for ARMs, where FHA experience starts in 1991. There is also one instance in which an effect coefficient must be imputed for fixed-rate refinance loans. These special cases are detailed below, by variable.

Loan Performance in First Three Quarters of Loan Life

Defaults in the first year of loan life are rare. They are generally due to faulty underwriting so the costs are often borne by loan originators, rather than by FHA. Prepayment activity in this period is also negligible. In the simulation model, termination rates in the first *three* quarters of loan life are fixed at a 30 basis point annual rate (0.0030) for defaults and a 100 basis point (0.0100) annual rate for prepayments. These rates reflect an average experience of recent cohorts.

Loan Performance After Year Twenty

Assumptions used for post-year-twenty experience have very little effect on lifetime loan performance and estimated subsidy rates. The minimal effect is because most terminations have occurred by the end of year 20, and because discounting cash flows to produce subsidy rates diminishes the effect of any variations in post-year-twenty performance. Pre-1982 insurance cohorts enter the forecasts with 20 or more years of loan seasoning. They are assigned termination rates in line with forecasts made by the FHA actuarial-review contractor.[78] Contractor projections of default rates on these loans show a fairly uniform random pattern within the range of 5 to 15 basis points (.0005 - .0015) per year, so a value of 10 basis points is used in the FHA-BSSS. Prepayment rates predicted by the FHA contractor also appear as fairly uniform random variables, with a range of 5 to 17 percent per year, so a value of 11 percent is used in the FHA-BSSS.

Post-year-twenty loan performance for cohorts entering the forecast period with less than twenty years seasoning (post 1982 cohorts), follows from the year-twenty forecasts made in the FHA-BSSS. They are fixed at the larger of 5 percent per year or the projected rate in the fourth quarter of year twenty. Default rates in the post-year-twenty period start with the projected rate in the final quarter of year twenty and decline with a 10 percent (not percentage points) annual rate of decay after that.

Performance Regression Coefficient Adjustments

There is no one single method of imputing values for missing regression coefficients. In the regressions described above, most missing values are for highest-value classes of categorical variables. Therefore, the primary imputation method used in the FHA-BSSS is to extrapolate beyond the estimated class-level effects using a ratio of class-level effects found in another regression. For example, if variable *X* has a complete set of seven

[78] Deloitte & Touche, U.S. Department of Housing and Urban Development Annual Actuarial Review of the Federal Housing Administration's Mutual Mortgage Insurance Fund As of Fiscal Year 2000. Philadelphia, PA: Deloitte Touche Tohmatsu, December 13, 2000, Appendix I.

coefficient estimates in regression 1, say $\beta 1(1)$ - $\beta 1(7)$, but is missing class 7 in regression 2, $\beta 2(7)$, then $\beta 2(7) = \beta 2(6) \cdot \left(\frac{\beta 1(7)}{\beta 1(6)} \right)$.

In a few cases, this type of imputation is not done because the value of the highest-class effect estimated in the regression analysis does not correspond with the rest of the series, nor to what is seen in the regression coefficient series in other equations. These are cases in which the final class appears to be an outlier class with too few observations for reasonable estimation of the effect.

The following discussion of imputation for individual coefficient values refers to regression results reported in Tables 9 and 10. The imputed values are identified in those Tables by boldfaced type.

Mortgage Age Class. The refinance regression for ARM defaults only measured effects for the first five age classes (years 1-10). Given the large (negative) size of the category five effect, category six and seven effects are simply set equal to that value. For the purchase regression for ARM defaults, the category seven age effect is set to that same value (from the ARM refinance equation).

For ARM prepayment regressions, the category seven age effect in the purchase equation is extrapolated using the class-seven-to-six ratio found in the fixed-rate purchase equation. Category six and seven age effects in the refinance equation are set equal to the effects computed for the ARM purchase equation.

House Price Cycle Stage. ARM refinance loans in the observation period did not pass through any complete house price cycles. As a result, all observations are in *CYCLE_STAGE*=4, leaving even that effect subsumed in the constant term of the regression. Thus, when imputing values, instead of the four stage coefficients summing to zero, as would be the case if all stages entered the regression, the *CYCLE_STAGE*=4 coefficient value must be set to zero.

For default regressions, effects in the fixed-rate refinance equation are nearly identical to those in the fixed-rate purchase equation. It is then reasonable to use effects from the ARM purchase default equation directly in the ARM refinance default equation, for cycle stages one through three. Effects for the prepayment equation start with those from the ARM purchase equation, which are scaled according to the relative size of effects in the fixed-rate refinance prepayment equation compared with the fixed-rate purchase prepayment equation. Again, the stage four effect is set to zero.

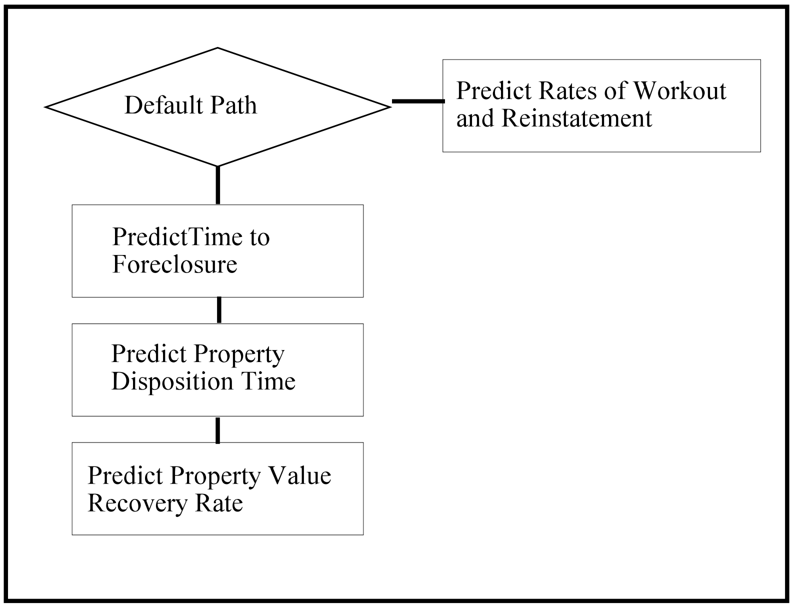
Interest Rate Spreads. Effects are imputed for the last two interest rate classes in the ARM-purchase-default equation as well as for both the first two and last three classes in the ARM-refinance-default equation. In each case, missing class effects are extrapolated from the existing (intermediate class) series using relative cross-classes effects from the fixed-rate-default equations, purchase and refinance, respectively. However, before the calculations can be performed for the ARM-refinance-default coefficients, the final spread-class effect for fixed-rate refinance default equation is extrapolated from the cross-class relationship (ratio of class-14-to-class-12 effects) found in the fixed-rate-purchase-default equation.

For prepayment equations, effects for the last two spread classes in the fixed-rate-refinance and ARM-purchase equations are set equal to the effects in the fixed-rate-purchase equation. For ARM-refinance loans, effects for the first and last three classes are set equal to one-half the effects of the fixed-rate-refinance equation. This choice follows the relationship seen for classes in which both of these equations have coefficient estimates .

Lowest Loan-To-Value Class. In all performance regressions, the lowest loan-to-value (LTV) class–loans with original LTVs below 80 percent–is excluded from the estimation because of low volumes. They average between one and two percent of total FHA insurance per year. To allow for their inclusion in the simulations, the effect for LTV class 1 is imputed as follows:

a. Prepayment equations: zero (average) effect for the two purchase equations and class 2 (81-90 LTV) effect for the two refinance equations.
b. Default equations: class 2 effect used for class 1in all equations.

IV. POST-DEFAULT OUTCOME REGRESSION MODELS



Default rate regression models described in section III provide only part of what is needed to measure timing and dollar amounts of credit-cost cash flows for FHA loan guarantees. The actual cost of paying insurance claims, and time lags in the process, depend on many additional factors. These factors include mortgage terms, state foreclosure laws, economic conditions, and FHA policies. It is important to estimate statistical models that relate cost factors to economic conditions, and to measure systematic differences across mortgage types and property locations. Four such regression models are presented in this section. Cost factors insensitive to such influencing factors are treated as fixed amounts and percentages, and are discussed in section V.

The four regression models described here are for:

- Probability (relative frequency) of defaulted loans receiving workouts;
- Time-to-foreclosure for loans not receiving workouts, and those that fail in attempted workouts;
- Time-to-disposition (sale) for foreclosed properties; and
- Recovery rates (net sale proceeds) on foreclosed properties.

A timeline showing how these and other cost factors fit together to determine amount and timing of credit-cost cash flows is provided in Table 11.

LOGISTIC REGRESSION OF WORKOUT ASSISTANCE OFFERS

Once delinquent borrowers miss three payments and a fourth is due (90-days delinquency), FHA loan servicers are required to evaluate each case for use of tools to avoid foreclosure. Loan servicers themselves are evaluated annually by FHA on how well they do in reinstating defaulted borrowers and in minimizing the costs of resolving all defaults. Those that score high are rewarded with higher foreclosure-cost reimbursements (75 percent versus 67 percent) and are given more flexibility in the administrative rules governing eligibility of borrowers for workout assistance.

This workout program, called *Loss Mitigation*, is relatively new, having just begun full operations in 1997, but having grown steadily since then.[79] Loss mitigation activities have significant effects on default-related insurance-claims costs. Whereas a full foreclosure can cost FHA, on net, 35 percent or more of the outstanding loan balance, successful loss mitigation cases have an average cost of just 1.5 percent of outstanding loan balances. Therefore, it is important to understand how the program is being used by loan servicers, and how rates of use vary across borrowers, loan types, and economic conditions.

Regression Model

The probability of loan workout (versus foreclosure) model is based on a binary logistic regression using records on over 400,000 non-cured 90-day defaults, from 1998Q1 to 2001Q3.[80] The regression estimation is on individual loan activity. In simulation forecasts the estimated probability equation is used to predict the percent of defaults that receive workout offers, by loan group and quarter. The regression estimation uses an effects parameterization, as do the loan termination regressions (see section III). Coefficient estimates from the logistic regression are reported in Table 12.

The time-frame of the regression analysis is also the phase-in period of the FHA loss-

[79]This program was approved by Congress and announced by FHA April, 1996. Interim regulations were issued in July, 1996. However, given the time necessary for FHA and lenders to retool information systems and retrain staff, there was no measurable activity until 1997.

[80]The term “workout” is used here, rather than loss mitigation, because workouts also include forbearances offered by loan servicers, outside the FHA loss mitigation tracking system. These forbearances accounted for as much as three-quarters of all workouts during the late 1990s. However, FHA now directs the servicers to report these as loss-mitigation forbearances, so the rate of outside forbearances has declined.

mitigation program. Workout offers increased dramatically from 18 percent of all non-cured defaults in early 1998 to over 75 percent by mid-year 2001. Numerous ways of capturing changes in the baseline rate of workout offers during the phase-in period were tested here. These tests involved using a time-varying constant term in the regression, or else adding some function of elapsed time since 1997Q1. While regressions with these effects increased the statistical fit of the estimated model, none provided satisfactory results when applied in forecast simulations; the variability of simulated outcomes was too small, as forecast results were dominated by the program phase-in effects. In contrast, industry analysts who work with conventional-market, default-workout programs indicate that a large variance of outcomes across economic conditions should be expected.

Estimating the regression with economic variables and without any phase-in effects produced a model with larger coefficients on the economic variables, allowing for a wider range of outcomes in the simulations. The final regression model then excludes any phase-in effects, yet has a constant term that is consistent with program activity in 2001.

Regression Results

ANOVA test statistics calculated for this regression indicate that *HPI_CLASS* has the largest effect on workout utilization rates, with *PRICE_CLASS* second, and *HPI_GROWTH_CLASS* third. The effect coefficient for *HPI_CLASS* ranges from -2.20 for price index values below 0.90 (more than 10 percent cumulative, to-date property value depreciation) to +1.26 for price index values above 1.50 (more than 50 percent cumulative, to-date property value appreciation). Given an average probability of workout of near 33 percent across the historical sample, the range of workout rates resulting just from differences in house price growth is from 5 to 64 percent.[81]

When viewing workout rates by loan characteristics, one sees the greatest rate of loan workout offers for high-priced homes (*PRICE_CLASS* = 3), loans with high initial downpayments (LTV_*CLASS* = 1), and loans with 15-year amortization periods (*PROD_CLASS* = 2). These outcomes all suggest that households in higher income classes who, in general, will have more resources to draw upon to effect a workout, do indeed have higher rates of use of workouts. In general, workouts can only be offered if the borrower has regained income sufficient for maintaining the mortgage and property within a short period of time.[82]

[81]These changes in workout probability (p) are calculated using the relationship between logit coefficients and the workout odds-ratio. The odds-ratio at the sample mean is $p/(1- p) = 0.33/(1-.33) = 0.50$. When the effect for *HPI_CLASS* =1 or 6 is turned on, the odds-ratio will change proportionately, according to the value of each *HPI_CLASS* coefficient (β): $p_1/(1-p_1) = \exp\{\beta\}*(.33/.67)$, where p_1 is the new probability of workout. Inserting the values for β and solving for p_1 gives the new probabilities. While the data sample is centered around a workout probability of p=.33, simulations using the regression equation are centered around p=.50.

[82]However, in October, 2002, FHA issued new directives to loan servicers that allow long-term forbearance plans for borrowers unemployed but with good prospects of near-term employment.

Use of Model in Forecast Simulations

The regression results reported in Table 12 are missing effects for *CYCLE_STAGE* = {1,2}. This omission is because the short history of the FHA loss mitigation program yields no observations in these stages in any Census Division or MSA. In forecast simulations, the effect coefficient for stage 4 is used also for stage 1, and the effect for stage 3 is used also for stage 2. This choice is based on an assumption that ease of finding new work will be more similar between stages 2 and 3, representing the bottom of a recessionary period, and also between stages 1 and 4, which represent early recession and expansion.

TIME-TO-FORECLOSURE REGRESSION

Foreclosure is the legal process of taking property title from homeowners to satisfy unpaid mortgage debts. As this process falls into the general category of property rights and law, foreclosure rules are specific to each state. However, there are two principal methods in use: power-of-sale and judicial proceedings. Power-of-sale, also known as trustee sale, allows a trustee appointed at the time of mortgage origination to sell the property in an orderly auction, as soon as the lender has complete homeowner notification requirements. These sales can be done quickly. The usual time frame for an uncontested power-of-sale foreclosure is between six and 16 weeks.

Judicial foreclosure, on the other hand, requires court filings and proceedings that can last for many months. States that require use of judicial proceedings often also have statutory post-foreclosure redemption periods, to give homeowners even more time to find the funds necessary to reinstate their mortgages and reclaim their homes. Once a judge decides that the lender is within its rights to obtain judgment, the property is auctioned by a court magistrate. These are often referred to as Sheriff's Sales.

Foreclosure times also vary according to economic conditions and the number of homes in foreclosure. As recessions deepen, and defaults escalate, the institutional time necessary to process individual cases increases. Also, as recessions deepen, more borrowers file for bankruptcy court protection. Such filings cause an automatic stay on all debt collections, including pending foreclosures. Attorneys for the lenders must then file with the courts for release from the stay in order to proceed with the foreclosure. All of these events tend to extend average foreclosure times as economic conditions worsen.

Regression Model

Because foreclosure times are inherently institutional processes, they are best modeled with waiting-time theory. In waiting-time models, the focus of empirical estimation is the distribution of times from entry into a queue until service. Here, the foreclosure time distribution is modeled using an accelerated failure time (AFT) regression of months from default to foreclosure, using individual loan data.[83] AFT is a semi-log regression where time-to-foreclosure, in log form, is regressed on explanatory variables that act as proportional adjustments to a baseline, stochastic foreclosure time.[84] That underlying, baseline stochastic distribution is found in the error term of the model. The distributional form used here is the Weibull, which allows for periods of increasing and then decreasing rates of foreclosures, over time. This simply means the foreclosure/waiting time probability distribution has a single-peak at some point in time greater than zero. The estimation procedure chooses the best-fit Weibull shape for the data.

Regression Results

Full regression results are reported in Table 13, which includes regression coefficients, the standard errors of those estimates, and a column of proportionality factors. These factors show the effects of changes in the explanatory variables on foreclosure times. In the AFT model, a one unit change in a continuous variable, or turning on a categorical effect, multiplies the base foreclosure time by $\exp\{\beta\}$, where β is the regression coefficient. Negative coefficients will have $\exp\{\beta\}$ values less than one, and positive

[83] After studying the data and estimation results, all available data was used, dating back to defaults in 1975. The only restriction placed on the data was that loans assigned to FHA prior to foreclosure were removed from the sample. These loans could have extended foreclosure timeframes because of forbearances provided in assignment. The assignment program is no longer operating, so whatever information could be gained from studying that history is no long valid for forecasting the future.

[84] The AFT is related to the proportional hazard model, where the event of interest is the rate of exit from the sample each period, due to foreclosure. Proportional hazard models estimate conditional rates of exit each period, with the values of explanatory variables and effect parameters scaling those rates up or down. Proportionality in the AFT comes from the way that values of explanatory variables scale time-to-exit up or down. This scaling indirectly changes the hazard (exit) rates.

coefficients will have exp{β} values greater than one.[85]

Unlike logit regressions used for mortgage performance and default workouts, categorical variables in the AFT regression do not use an effects parameterization.[86] The effects of the largest value-class for all categorical variables are combined into the constant term of the regression. The effects shown in Table 13 are then all relative to that of the combined highest-value classes of all variables.

The regression explanatory variables capture differences in foreclosure times by economic conditions and state property laws. Economic conditions enter the model through house-price cycle stage (*CYCLE_STAGE*), local house price growth in the first four quarters after default (*HPI_GROWTH*), and the market interest rate for long-term fixed-rate mortgages at the time of default (*CUR_MKT_RATE*). Property-law effects enter the regression through a Census Division indicator (*CENSUS_DIV*). This measures the average effect of the various state laws in that Division, weighted by actual FHA activity in each state. The regression results show moderate differences across Divisions.

---

[85] The extremely large proportionality factors on some *DEFAULT_YR* effects, where those effects lack statistical significance, only highlight that there were few observations with these effects and that their foreclosure times differed markedly. Removing these years from the analysis did not change the regression results. The underlying coefficient effects for these years are not used in the forecast simulations.

[86] All regression analysis is performed using the SAS System statistical software. The AFT procedures in SAS (PROC LIFEREG) do not allow for an effects parameterization. The last value category of each variable is automatically removed from the data used in estimation.

The last explanatory variable in the time-to-foreclosure regression is default year (*DEFAULT_YR*). This variable controls for institutional factors affecting foreclosures. Prior to 1998, foreclosure time including processing of assignment requests by HUD field offices. The purpose of (mortgage) assignment was for HUD to buy defaulted mortgages and give borrowers additional time to work out their financial difficulties. However, a large percentage of applications were denied and they appear here through extended foreclosure times. The *DEFAULT_YR* variable captures assignment processing delays and differences in HUD rules on the time lenders have to initiate foreclosure, over time.

To avoid problems associated with outlier foreclosure time values that may represent data errors rather than true event times, time-to-foreclosure on individual cases is censored at 36 months. All values greater than 36 are treated as though they were observed as open cases at 36 months, with no additional information available. The Weibull shape parameter is estimated assuming that the shape of the waiting time (foreclosure time) function as it approaches the 36-month mark continues after that point.

The regression results in Table 13 show significant variations across Census Divisions, and that foreclosure times are extended by 27 percent (average effect of 3 months) when local housing cycles are passing through a trough (*CYCLE_STAGE*=3). Default year effects show a downward trend over time. The *DEFAULT_YR* effect used in the simulation forecasts is 0.50, which corresponds with the effects of 1996 and 1999.

Though the coefficient for 2000 is lower, the volatility of this parameter over the final years of the historical period calls for caution when choosing a value to use as a baseline level for forecast simulations.

Use of Model In Forecast Simulations

When predicting foreclosure times on grouped data in the forecast simulations, all estimates begin with the mean-value Weibull baseline foreclosure time, using the estimated shape parameter, rather than randomly assigning values from the Weibull distribution. Given the shape parameter, *c*, the mean of the Weibull distribution is

$$\overline{W} = \Gamma\left(\frac{1/c + 1}{1/c}\right)$$, where Γ is the Gamma function.

Starting with the mean value saves computation time without sacrificing any portfolio-level variability. The mean baseline time, W-bar, is scaled for each forecast observation according to values of the explanatory variables.[87] Forecast foreclosure time is then

$\overline{W} \cdot \exp(X\beta)$, where W-bar is the mean-value Weibull variable, X is the vector of explanatory variables and β is the vector of estimated coefficients. The $exp(X\beta)$ term is known as the scale parameter for the Weibull distribution.

[87]For more information on the Weibull distribution, see Merran Evans, Nicholas Hastings, and Brian Peacock, *Statistical Distributions*, 2nd *ed.* New York: John Wiley & Sons, 1993, chapter 41.

TIME-TO-DISPOSITION REGRESSION

A second time-frame of interest for credit-cost cash flows is the delay between property foreclosure and property sale. The purpose of foreclosure sales is ostensibly to liquidate properties and payoff the defaulted loans. However, sales to third parties at foreclosure auctions are rare because the properties are generally deteriorated to where they need repairs and significant maintenance before they can be sold at a price close to the outstanding loan balance. Third-party sales are also inhibited by the foreclosure sale process, one in which potential buyers cannot inspect the property beforehand and must provide large deposits at the time of auction. Thus, FHA generally takes title to foreclosed properties by instructing its loan servicers to bid in the auction, on its behalf. Servicers then file claims for insurance reimbursement for the loan balance and foreclosure expenses, and HUD takes the properties and works with contractors to sell them. The properties are known as *real estate owned* or, simply, REO.

Regression Model

Like time-to-foreclosure, time-to-disposition is modeled from waiting-time theory, using an AFT regression on months from foreclosure to disposition. Time-to-disposition is primarily a function of economic conditions at the time of foreclosure, but it has institutional influences as well. Explanatory variables used in this regression are the same as were used for time-to-foreclosure, only here, the economic variables are observed at time of foreclosure, rather than time of default, and house price growth is in categorical form (*HPI_GROWTH_CLASS*). In addition, time-to-disposition is censored at 24 months, rather than 36. To account for various problems FHA has had with property-management contractors over the years, a series of foreclosure-year variables are added to the analysis (*FORECLOSE_YR*).

<u>Regression Results</u>

Estimation results are found in Table 14. The proportionality factors reported there indicate that REO disposition times are extended during housing cycle downturns, as prices and sales volumes fall throughout the market. The housing-cycle proportionality effect alone is around 10 percent (stages 1 and 2), while the added effect of negative appreciation rates *(HPI_GROWTH_CLASS = 1*) is another 14 to 19 percent. Together, these results indicate that periods of housing market declines extend property sales times by 25 to 30 percent, meaning 1.5 to 2 months, or one quarter.

Interest rates prevailing after foreclosure (*CUR_MKT_RATE*) also affect disposition times, but the effect is modest. Each one percentage point increase in mortgage rates leads to a 3 percent increase in disposition time. With an average disposition time of 6 months, that amounts to just under one week. So, it takes a five percentage point increase in interest rates to extend disposition times by one month.

Use of Model in Forecast Simulations

The only free parameter to be chosen for forecast simulations is the *FORECLOSE_YR* effect. As with calendar year effects in the foreclosure time regressions, there has been a downward trend over time, with some volatility in recent years. The value chosen is 0.40, which equals the average value of the 1997-2000 effects.

REO RECOVERY RATE REGRESSION

The final stage in the process of settling accounts on loan defaults is disposing of foreclosed properties. Disposition produces receipts that help offset the cost of paying full insurance claims at the time of foreclosure. Just how large the receipts are, relative to the unpaid balance of the loan at default, is a function of housing market conditions after foreclosure, property and owner characteristics, and regional differences. The ratio of net sales receipts to the unpaid loan balance is the foreclosed-property recovery rate.

Regression Model

Recovery rates are modeled with ordinary least-squares regression. The dependent variable in the regression is sales price less sales expense, divided by the outstanding loan balance. A recovery ratio measure net of property and holding expenses was tested, but it fit the data less well. Likewise for a recovery ratio net out repairs expenses. These additional disposition related expenses are then added in the forecast simulations as fixed ratios of loan balances (See section V, Table 16).[88]

Following work by OFHEO on conventional-market recovery rates, the best variable found here for modeling economic volatility in recoveries on FHA loans is a statistical measure of the price decline necessary to make the default put option in-the-money (*PRICE_DROP_STAT*).[89] This variable is the same standardized normal statistic used to calculate the probability of negative equity (*NEG_EQ_CLASS*) variable used in the logistic default regressions. Only here, the statistic is the additive inverse of that used to compute *NEQ_EQ_CLASS*.[90] *PRICE_DROP_STAT* is the sum of market generated property equity (at the average appreciation rate) and borrower generated equity via the downpayment and regular payments toward loan principal. It is then a measure of how much property-value depreciation must be on an individual property to wipe out market and borrower generated equity, and thus to make foreclosure an attractive option to the borrower/homeowner.

---

[88] An additional issue arises because HUD has programs where it sells properties at below-market prices to non-profit organizations, local governments, policemen, and school teachers. Eliminating observations with very-low sales prices did not change the regression results materially. Therefore, since below-market sales are expected to continue, we allowed these loans to remain in the regression sample set.

[89] Department of Housing and Urban Development, Office of Federal Housing Enterprise Oversight, Risk Based Capital, *Federal Register*, Vol. 64 no. 70, Tuesday, April 13, 1999, pp. 18188-18192, esp. p. 18190.

[90] *PRICE_DROP_STAT* is computed as, $\frac{\log\left(\frac{HPI_t}{LTV_t}\right)}{\sigma_t}$, where HPI_t is the market house price index value from loan origination to default, LTV_t is the original loan-to-value ratio, amortized to time of default, t, and σ_t is the standard deviation of cumulative house price growth rates around $log(HPI_t)$.

PRICE_DROP_STAT indicates how good or bad are housing market conditions at the time of foreclosure. A high value of *PRICE_DROP_STAT* indicates that market conditions were good so that there had to be a large amount of property depreciation to make foreclosure in the best financial interest of the mortgage borrower. In such cases, property depreciation on foreclosed properties is likely to be just barely enough to make the foreclosure beneficial to the borrower/homeowner so that property recovery rates are likely to be high. On the other hand, a low (or negative) value of *PRICE_DROP_STAT* indicates that housing market conditions are fair-to-poor, so that it does not take much property-specific depreciation to make foreclosure in-the-money to the borrower/homeowner. In such cases it is likely that typical property depreciation levels will cause actual recovery rates to be low.

Additional effects found to be important in the regression analysis are the property price class (*PRICE_CLASS*), owner-type (*INVESTOR*), and Census Division (*CENSUS_DIV*).[91] Over time, some Divisions consistently had recovery rates that were either higher or lower than expectations based on economic and property characteristics. Mortgage rates were excluded from this regression because their influence was too small to merit including them in the simulation forecasts.

[91] The large differences between property-price classes are clearly seen in frequency distributions of recovery rates, by price class. The distributions are normally distributed with similar variances, but with means of 0.67, 0.79, and 0.88 for classes 1, 2 and 3, respectively.

A control for disposition year (*DISP_YEAR*) was added to the regression to account for variations in HUD procedures and contracting methods over time.

Regression Results

Regression results are reported in Table 15. Additional analysis-of-variance tables (not shown) indicate that *PRICE_DROP_STAT* is the largest influence on recovery rate outcomes, with *PRICE_CLASS* second and *CENSUS_DIV* third. The positive coefficient on *PRICE_DROP_STAT* suggests that, in healthy housing markets (*PRICE_DROP_STAT* >> 0), the average recovery rate will be greater than the 79 percent indicated by the constant term of the regression. In declining markets (*PRICE_DROP_STAT* < 0), the recovery rate will be lower. The 79 percent average recovery on REO exists when average *depreciation* in the market would make the value of the average property equal to its outstanding loan balance. Defaulted properties normally have appreciation/depreciation worse than the market average, leading to the additional (average) 21 percent loss of value in these circumstances.[92]

Loans in the highest property price class (above 100 percent of area median price) have recovery rates 23 percentage points higher than those in the lowest class (under 50 percent of area median price). Investment properties have recovery rates 11 percent below owner-occupied properties. By region of the country, the lowest recovery rates over time

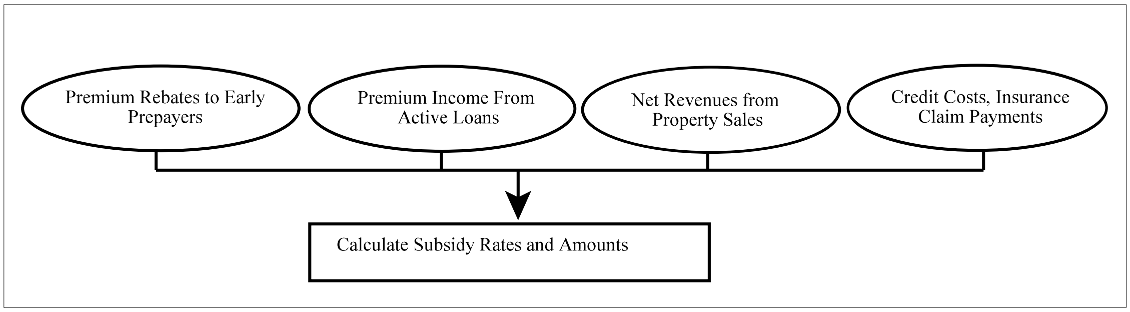
[92] Declines in *PRICE_DROP_STAT* can be thought of as indicating that exercised put options will be deeper in the money, yielding lower recovery rates.

have been consistently in the New England and MidAtlantic Census Divisions.

Use of Model in Forecast Simulations

To forecast recovery rates in the subsidy-rate simulations, the *FORECLOSE_YR* effect chosen is an average of the most recent five-year period (0.040).

V. FINAL FORECAST ASSUMPTIONS AND SUBSIDY RATE CALCULATIONS



Forecasts of economic conditions and mortgage performance provide the foundation for setting up calculations of government cash flows from FHA loan guarantees. The two primary cash-flow streams that define subsidy estimates on FHA guarantees are premium income and credit expenses. These, in turn, are sums of individual calculations for:

- Premium income inflow on new mortgage originations (up-front premium);
- Premium income each quarter on outstanding loans, within contractual time frames (ongoing premiums are assessed annually, paid monthly);
- Partial refunds of up-front premiums due to borrowers that payoff mortgages in the early years (based on premium earning rates and schedules determined by FHA);
- Claim payments to loan servicers for workout efforts;
- Claim payments to loan servicers for foreclosed properties; and
- Recoveries on foreclosed properties, net of all property and sales expenses.

Lifetime cash-flow time series for each loan group are the grafting of historical cash flows with future projections. They are created for each simulated economic path.

Regression models outlined in Sections II-IV provide all of the meaningful variation in subsidy rates, yet there are several additional cost and revenue factors needed to complete the cash flows. These are discussed here. A final issue addressed here is determination of volumes and characteristics of future insurance cohorts.

All historical data on cash flows is maintained in databases used as starting positions for the simulation analysis. These records are aggregations, following the same scheme used for mortgage performance regression analysis data bases, as outlined in Section III. The one difference here is that historical data used as forecast starting positions are aggregated by cohort (fiscal) year, rather than by loan origination quarter. To make quarterly predictions of mortgage events, all loans are assumed to have been originated in the second quarter of the calendar year of their cohort (third quarter of fiscal year). The main reason for quarterly forecasts is to model the effects of intra-year movements in interest rates. Because the loan aggregation scheme separates loans by original interest-rate class, the assumption of a common origination quarter is not restrictive. After quarterly forecasts are made, the cash flows are aggregated by fiscal year of experience before subsidy rates are calculated.

PREMIUM INCOME

Premium income is the more straightforward of the two primary cash-flow series. FHA-insured borrowers pay an up-front premium, which is nearly always financed with the mortgage, and pay ongoing, monthly fees (based on an annually assessed premium) for a certain length of time. The up-front fee is refundable, in part, if the borrower pays off the mortgage early. Both premium and refund rates have changed over time and so are specific to each origination cohort. Current rates are used for all future cohorts.[93]

Because the mean/average prepayment in any quarter will occur at the end of month two, premium income on loans paid off in a given quarter is assumed for two of the three months. For loans defaulting in each quarter, no premium income is recorded. Loans that receive workout assistance return to paying premiums (in the simulations) in the quarter after default.

OTHER PARAMETERS NEEDED TO CALCULATE NET DEFAULT COSTS

There are numerous instances in which default cost components are not particularly sensitive to economic conditions or loan characteristics. Recent trends in FHA business activity provide reasonable values for these cost parameters in forecast simulations. Each

[93]Readers who desire the exact details are referred to Deloitte & Touche LLP, *Annual Actuarial Review of the Federal Housing Administration's Mutual Mortgage Insurance Fund Fiscal Year 2001*. Philadelphia: Deloitte & Touche, December 2001, Tables D.2 and D.3.

parameter is listed and described in Table 16.

FHA has an additional default workout option that has seen relatively little use: the preforeclosure sale. It permits borrowers to sell their properties, with FHA paying for any loss-on-sale. These sales generally result in measurable savings to FHA, over the net cost of foreclosure and property disposition. However, use of this tool has been limited to about five percent of the number of foreclosures each year. Because it has not been an important tool for FHA, and its usage has been fairly consistent over time, the FHA-BSSS focuses on other aspects of default resolution and cost. In the forecast simulations, five percent of foreclosures are treated as preforeclosure sales, and are given a cost saving of 8 percentage points. This cost saving rate was the average found in HUD's evaluation of its preforeclosure sale pilot.[94]

FUTURE COHORT CHARACTERISTICS AND VOLUMES

Forecast simulations provide estimates of experience on future loan cohorts, as well as projections of lifetime performance for existing cohorts. Volumes and characteristics of future cohorts are generated in four parts:

---

[94] U.S. Department of Housing and Urban Development, Office of Policy Development and Research, *Evaluation of the Federal Housing Administration Preforeclosure Sale Demonstration*. Washington, DC: U.S. Department of Housing and Urban Development, June 1994.

1. All key variables other than the interest rate are taken from a chosen historical cohort;
2. Interest rates are assigned based upon values generated in the vector autoregression (section II), by quarter of loan origination;
3. Refinance dollar volumes are fixed in the first two years of the forecast period and adjustable rate mortgage (ARM) volumes are fixed throughout the forecast period as a percent of all originations; and
4. Dollar volumes of purchase and refinance loans grow over time with population and inflation factors.

In the FHA-BSSS, these four sets of factors are specified as follows:

1. The most recent FHA originations available at the time of FHA-BSSS development were for FY2002. Those loans are used as a basis for starting future originations. First, the loans are aggregated using eight of the nine classification variables outlined in section III (interest rates are excluded). Numbers and dollars of loans are recorded for each aggregate loan group. All future years use the distribution of loan-group characteristics (i.e., percentage of dollar volume found in each loan group) from the FY2002 cohort of FHA insurance endorsements.
2. Within each quarter's originations, interest rates are assigned by product type. Three product types are used for coupon interest rate determination: fixed-rate 30-year, fixed-rate 15-year, and adjustable rate mortgages. No graduated payment or graduated

equity loans are assumed to originate in the forecast period. The VAR model outlined in section II generates a fixed-rate 30-year mortgage rate directly. A 50 basis point deduction is taken to arrive at a 15-year fixed-rate. ARM contract rates are the prevailing one-year Treasury rate plus an index margin of 2.75 percent. This margin is the most common used for FHA loans. Because teaser rate discounts have not been permitted since 1998, the index-adjusted Treasury rate provides the best estimate of rates on newly originated ARMs.

3. The dollar volume of refinance loans is fixed at $50 billion in FY2003, and at $20 billion in FY2004. These correspond with FHA predictions mid-year 2002. All loan-group dollar volumes and numbers are adjusted to reflect differences between what was calculated based on 2002 origination shares of refinance and purchase loans and these new totals. The ARM share of total originations, both purchase and refinance, is fixed at 4 percent, which is representative of ARM shares in recent years.[95]
4. Population-growth induced increases in mortgage volumes (and numbers) are set at 2 percent per year and the inflationary change in mortgage volumes is set to 4 percent. The 4 percent inflation factor begins for purchase loans in FY2003 and for refinance loans in FY2004.

The resulting origination volumes for future cohorts are not tied to either house price appreciation or prepayment rates generated by the forecast simulations. They are, rather,

[95] ARM shares have been higher in years with significant refinance activity. Since the current version of the FHA-BSSS does not predict refinance waves, the lower rate of ARM activity is assumed to exist in future years.

illustrative of continuing current business into the future.

SUBSIDY CALCULATIONS

Cash flows resulting from mortgage performance projections are grafted to historical cash flows associated with each loan group to complete life-of-loan cash-flow series. Net premium income is subtracted from net credit expense in each year, and the difference is discounted to the year of loan origination (insurance endorsement) to compute subsidy amounts. The final subsidy rate is the (present value) subsidy amount divided by dollar amount of loans insured. The subsidy rate is then a function of the dollar mix of loan and property types in each cohort, as well as economic conditions in the forecast simulations.

Discounting procedures were discussed in section II. Each historical origination year cohort uses one set of interest rates and discount factors for all simulations, coming from the year of loan origination. These rates are determined and maintained by OMB. The underlying interest rates used to calculate discount factors for 1992-2003 are provided here in Table 4. For all future cohorts, discount rates are specific to each simulated economic scenario. Because the President's budget is formulated in the first quarter of each calendar year, the FHA-BSSS uses those interest rates to calculate discount rate factors in each simulation. Thus, interest rates prevailing in 2003Q1 are used for the FY2004 cohort discount factors, and 2004Q1 rates are used for the FY2005 cohort discount factors.

VI. SIMULATION RESULTS FOR THE 1992-2007 COHORTS

The FHA-BSSS model identified in Figure 1 and described in sections II-V generates a large number of outputs. These outputs range from databases with cash-flow summaries to tables of descriptive statistics for subsidy-rate and mortgage-termination-rate outcomes, and to graphs of frequency distributions of subsidy rate outcomes and their component parts. However, the ultimate objective of the FHA-BSSS is to improve budget preparation, so the results described here focus on the distribution of subsidy rate outcomes by budget-year cohort of loan guarantees.

The simulations merge historical data through FY 2002 with forecasts that start in FY 2003. All historical loan origination, termination, and cash-flow data are current through 2002Q3 (FY 2002). Economic data (house price growth, interest rates, inflation rates, and unemployment rates) are current through 2002Q4. The simulations described here were performed in March 2003.

Simulations at a given point in time provide budget re-estimates of outstanding cohorts, initial estimates for new cohorts, and projections for future cohorts. All of these can be summarized together in graphic and tabular form. Figure 4 maps subsidy rate means and confidence bounds (1, 5, 95, and 99 percentiles) by cohort year. The exact values mapped in Figure 4 are provided in Table 17.

SUBSIDY RATE CONFIDENCE BOUNDS

Figure 4 is most helpful for highlighting the rate at which confidence bounds tighten with the seasoning of individual cohorts. Cohorts from the early and mid-1990s have very tight confidence bounds today, which indicates that any future changes in mortgage termination patterns will have little effect on the measured value of these loans in the federal budget. In contrast, as we move toward more recent and then to future cohorts, projections of confidence bounds increase dramatically, indicating there still could be sizable future revisions to current budget estimates.

For recent cohorts, this increased variance of potential outcomes reflects uncertainty with respect to economic conditions during the crucial early years of mortgage life. Default and prepayment rates tend to peak in the three-to-seven-year time frame, and what those peak rates will be depend on house price growth and interest rates during the formative years of cohort existence. For future cohorts, the continued increase in confidence bounds is in proportion to the increase in range of mortgage coupon rates and existing house price conditions that could exist when those loans originate.

Following the time-line of cohorts in Figure 4, one also sees that mean-value subsidy rates approach zero in future years. Indeed the mean subsidy rate for FY2007 resembles the low rates of the 1995-1997 cohorts. This movement toward zero reflects both the increasing uncertainty in economic conditions and the movement of the means of the distribution of each economic variable to CBO's long-run average predictions, over a five-year period. Today, interest rates are at historical lows and real (national) house price growth is at an historically high level. Taken together, they portend relatively low prepayments and low defaults for the 2002-2004 cohorts. As mean levels of the economic variables in each quarter move toward their long-run averages (7.43 percent 30-year mortgage rates and 1.2 percent annual, real, national house price growth), conditions will be less favorable for new FHA guarantees. The implication of this movement is that subsidy rate estimates for outstanding FHA single-family loan guarantees should not be used as a basis for long-run policy regarding premiums and underwriting standards.

Subsidy Rate Estimates as a Basis for Policy Making

Economic conditions in 2003 are very favorable for new mortgage guarantees–interest rates are low and house-price growth remains healthy throughout most of the nation. This favorable climate means that simulation results for 2003 do not provide an appropriate basis to set policy on insurance premiums and underwriting standards. Simulation results for 2007 are more instructive for policy making because they embed a long-range view of plausible initial economic conditions. The 1000 simulations run for 2007 allow for initial house-price growth and interest rates to be high or low (See Figures 2 and 3) and to be moving either up or down.

The low mean-value subsidy rate (-0.34 percent) and high probability of positive values

(30 percent) for 2007 suggest that the current premium structure is marginally sufficient for long-run solvency of Mutual Mortgage Insurance Fund programs. Actuarial soundness generally requires a larger margin of error, meaning a long-run mean subsidy rate farther from zero and a correspondingly lower probability of losses.

While there is no one measure of actuarial solvency, the term means that the underlying (insurance) business enterprise has a limited probability of sustaining cumulative losses on outstanding insurance contracts large enough to jeopardize its viability. It is generally calculated without consideration of the value of new business in the future. For a private firm, actuarial soundness means controlling loss exposure so that the probability of firm insolvency and failure are kept within bounds acceptable to investors.

FHA enjoys the full faith-and-credit of the U.S. government and so does not face insolvency or failure, as do privately owned firms. For FHA the issue is then two-fold: the frequency at which losses might be experienced, and the potential size of those losses. Policy makers have time horizons of one, five, and ten years when formulating the federal budget. Actuarial solvency for FHA then must be defined as net receipts outweighing net outlays within those time horizons.

The 30-percent probability of positive subsidy rates estimated for 2007 means that, going forward, FHA can expect to have net losses on three-out-of-ten years books-of-business.[96] The -0.34 percent average expected long-term subsidy rate implies that, over a ten-year period, net receipts will outweigh net outlays by an amount equal to about 3.4 percent of an average year's business volume. The size of FHA's total outstanding portfolio averages between four and five times the size of any one year's business volume, so the implied "capital" ratio achieved over a ten-year period, as defined in law for FHA, will be somewhere around 0.75 percent.[97] Yet, that result too is an average that has a frequency distribution around it. It is based on having the average expected number of years with net losses (outlays), and the size of those losses also being of average expected size.

FREQUENCY DISTRIBUTIONS AND MEAN-VALUE ESTIMATES

The mean-value, simulated, subsidy rate outcome for each cohort yields an unbiased estimate of the true dollar subsidy amount to be revealed once all loans have terminated and all program cash flows are known. Unbiasedness means the expected value of the sum of all future revisions, given the information available at the time the initial estimate/forecast is made, is zero. At the same time, the mean is subject to misunderstanding because there is a better-than-even chance that revisions will be

[96] Or a six-year loss experience, as occurred in 1980-1985, happening every twenty years.

[97] FHA's statutory capital cumulates over time so that the twenty-year capital ratio can be expected to be around 1.5 percent. The actual capital ratio reported by FHA today is a summary measure of expected net receipts and net outlays on all past and presently outstanding insurance policies; it does not begin with a zero balance at any given point in time. Therefore, it is not a good measure of the solvency or actuarial soundness of FHA's insurance programs as they exist today. See USC 1711(f) for a description of how capital is measured for FHA.

downward rather than upward.

To see this potential for misunderstanding, let us start with an illustration of the distribution of outcomes for the FY2003 cohort of FHA guarantees, shown here in Figure 5. If one thinks of Figure 4 as the top-view of a probability mapping (like a contour map), then Figure 5 is a side-view, cross-section, or cut-away of that mapping, taken at one horizontal point–the year 2003. Figure 5 is a density plot, with relative frequencies of events (subsidy rate outcomes) measured vertically. The vertical dashed line marks the mean of the distribution (-1.85 percent).

The right skewness of the distribution implies that the mean is to the right of the median. The median is the estimate for which there is equal probability that the actual outcome will be either higher or lower: half of all potential outcomes are higher and half are lower. If the mean is used as the initial budget estimate, there is a better than even chance of future downward revisions, as actual performance becomes known. In the distribution for FY2003, the mean subsidy rate is at the 63rd percentile of outcomes, meaning there is a 63 percent probability that revisions will be downward and just a 37 percent probability that revisions will be upward.

If policy makers see a preponderance of downward revisions, they could mistakenly conclude that subsidy rate estimates used in budget preparation are too conservative, so that there is more room for lowering premium charges or loosening underwriting standards. The mean estimate may appear overly conservative in the short run (under-predicting actual budget receipts from FHA), but that is only because it makes allowance for the chance of large losses in future years. Choosing the mean rather than the median then provides something like a self-insurance premium against the potential for bad outcomes. Unlike the median, the mean considers the actual size of those bad outcomes in addition to their relative frequency.

This additional factor in mean-value calculations becomes meaningful when annual re-estimates are made. Re-estimates of subsidy rates are made for all outstanding cohorts of guarantees, and the netting of resulting adjustments are either passed on budget as new receipts (net downward dollar adjustment to subsidies) or require new budget outlays to provide for presently unfunded, but now expected, claim liabilities (net upward dollar adjustment to subsidies). The calculation of the dollar budget effect involves multiplying changes to subsidy rate estimates by dollars of loan originations. Even if the number of cohorts with downward revisions to subsidy rates is greater than the number of cohorts with upward revisions, it will likely be that the upward revisions are of a larger size and thus will balance the net budget effects of the re-estimates back toward zero.

Such balancing may not always occur in one set of (annual) re-estimates, but will more likely occur over time. That is because both good and bad economic surprises will affect all outstanding cohorts. Good economic surprises could mean a net of downward subsidy rate re-estimates for outstanding cohorts, while bad economic surprises could mean a net of upward re-estimates. The point of using the mean-value estimate (rather than the median) when making re-estimate adjustments is that when bad surprises happen, they will tend to be of larger size than are the good surprises when they occur, and thus budget effects of re-estimates will tend to balance out over time even though bad surprises happen less often than good surprises.

Even then, the mean value estimate may still appear to be overly conservative over a period of several years. Because by using mean-value estimates there will be more good surprises than bad surprises, upward subsidy rate re-estimates may not fully offset downward subsidy rate re-estimates over a five or ten year period. If that turns out to be the case, it is because the economic forecasts used in the FHA-BSSS to generate the distribution of subsidy rate outcomes include extremely bad events, meaning national recessions and depressions (see section II). Because they are small probability events, by definition, such worst-case outcomes may not appear for ten, twenty, thirty, or more years. In current subsidy rate forecasts, those bad events serve to help pull the mean away from the median. If they do not occur for many years, then net annual re-estimates over any ten year period could provide more new budget receipts than new budget outlays.

While it is valuable to include worst-case events to understand the shape of the entire frequency distribution of outcomes, it would also be a reasonable policy position to ignore outlier events when making decisions on program design. Such a position would lead to using the median- rather than the mean-value estimate when deciding on such policy options as premiums and underwriting standards. Choice among these options can be greatly influenced by the measured budget impact of any changes. Given the twin but competing goals for FHA's single family insurance programs of encouraging affordable homeownership opportunities while maintaining actuarial soundness, policy makers may want to focus on median in addition to mean outcomes.

While the distinction between mean and median can be important when budgets are formulated, it disappears fairly quickly as loan cohorts age. At present, mean and median predictions are virtually identical for the 1992-2001 cohorts. The differences for 2002-2007 are:[98]

Cohort Year	Mean Subsidy Rate Estimate	Median Subsidy Rate Estimate
2002	-1.73	-1.83
2003	-1.85	-2.07
2004	-1.53	-1.81
2005	-1.05	-1.38
2006	-0.65	-1.18
2007	-0.34	-0.94

[98]Mean outcomes are reported for all years, 1992-2007, in Table 17.

BIASES IN CURRENT FHA SUBSIDY RATE ESTIMATES

While mean and median subsidy rate outcomes are identical for pre-2002 cohorts, that does not mean that budget estimates now in use for those years are accurate. As mentioned in section I, current subsidy rates reported by FHA and published by OMB are not based on the type of analysis used in the FHA-BSSS.[99] While FHA uses a similar model of mortgage default and prepayment rates, it projects loan terminations and cash flows over one, smooth economic path. The result is that FHA consistently predicts subsidy rates far to the left even of median outcomes and thus consistently overstates expected net budget receipts from its loan guarantees. The annual re-estimation process only produces a slow convergence to mean/median expectations because it uses the same type of economic forecasts as used in initial budget-year estimates.

Other reasons for a downward bias in subsidy rate estimates currently used in FHA budget formulation include using fixed factors for rates of use of loss mitigation (foreclosure avoidance tools) and foreclosure losses. In the FHA-BSSS, these important elements of default costs are dynamic with respect to mortgage characteristics and economic conditions (see section IV).

FREQUENCY DISTRIBUTIONS OF SUBSIDY RATE COMPONENTS

---

[99]Office of Management and Budget, *FY2004 Federal Credit Supplement,* Table 8.

Subsidy rates are simply the sum of prepayment revenue and default cost component rates. Though these two components are interdependent, meaning that default and prepayment terminations are substitute outcomes for borrowers, it can be instructive to look at their separate frequency distributions. The right-skewness of the subsidy rate distribution comes principally from skewness of the default cost component, which itself reflects the magnitude of potential bad outcomes. (Each subsidy rate is just the sum of premium revenue and default cost components.) In contrast, the premium-revenue component is very symmetrical. Distributions of these components for FY2003 are shown in Figures 6 (premium revenue) and 7 (default cost). Following budget accounting rules, default costs are positive values while premium revenues are negative.

Not only is the premium revenue distribution more symmetrical, but it has a smaller variance as well. The distribution of premium outcomes falls within a 2.35 percentage point range, while the range of default cost outcomes is over 10 percentage points.[100] When combined, the entire range of net subsidy rate outcomes in Figure 5 is from -3.8 to +7.6, or over 11 percentage points.[101]

VALUE AT RISK

[100] Figure 7 does not show the last few default cost component observations, which extend to +10.26 percent.

[101] To avoid loss of detail, the far right extreme of the tail of the subsidy rate distribution is not shown in Figure 5. In 3 of 1000 outcomes, the subsidy rate exceeds the 4 percent maximum rate shown in Figure 5.

Private investors use various metrics to define risk of loss on a given investment. One common measure used for credit risk on lending positions is value-at-risk (VaR). It is a measure of the dollars that could be lost were a given bad event to occur within a defined period of time. That event is defined by a critical value along the frequency distribution of outcomes, where that critical value identifies a risk-tolerance level. For example, given the long-run nature of FHA loan guarantees, it would be reasonable to start with the critical value for the 99^{th} percentile event. That value is the subsidy rate for which only 1 percent of possible outcomes for one year's book-of-business are worse. The dollar amount of outlays implied by that subsidy rate would be the implied capital requirement for that percentile-rated risk tolerance. Holding such capital reserves would assure the solvency of the enterprise so long as events worse than that did not happen. Capitalizing against all possible eventualities is not practical, nor is it generally profitable, and so VaR critical values are always something less than 100%..

The 99^{th} percentile subsidy rate for 2003 is 2.07 percent. If FHA could and did hold reserves of 2.07 times its guarantee volume then taxpayers would be protected from losses with a 99 percent confidence level. There would be only a 1 percent chance that the outcome of these guarantees could be so bad as to require net budget outlays to enable FHA to fulfill its obligations to its lender partners, given current contractual terms for premium payments and refunds. However, 2003 is a very good year for mortgage insurance originations. If we go from the 2003 to the 2007 cohort, and thus examine a long-run expectation of FHA outcomes under a wide range of initial economic conditions, the VaR at the 99^{th} percentile is nearly 7 percent.

Figure 1. Structure of the FHA Budget Subsidy Simulation System

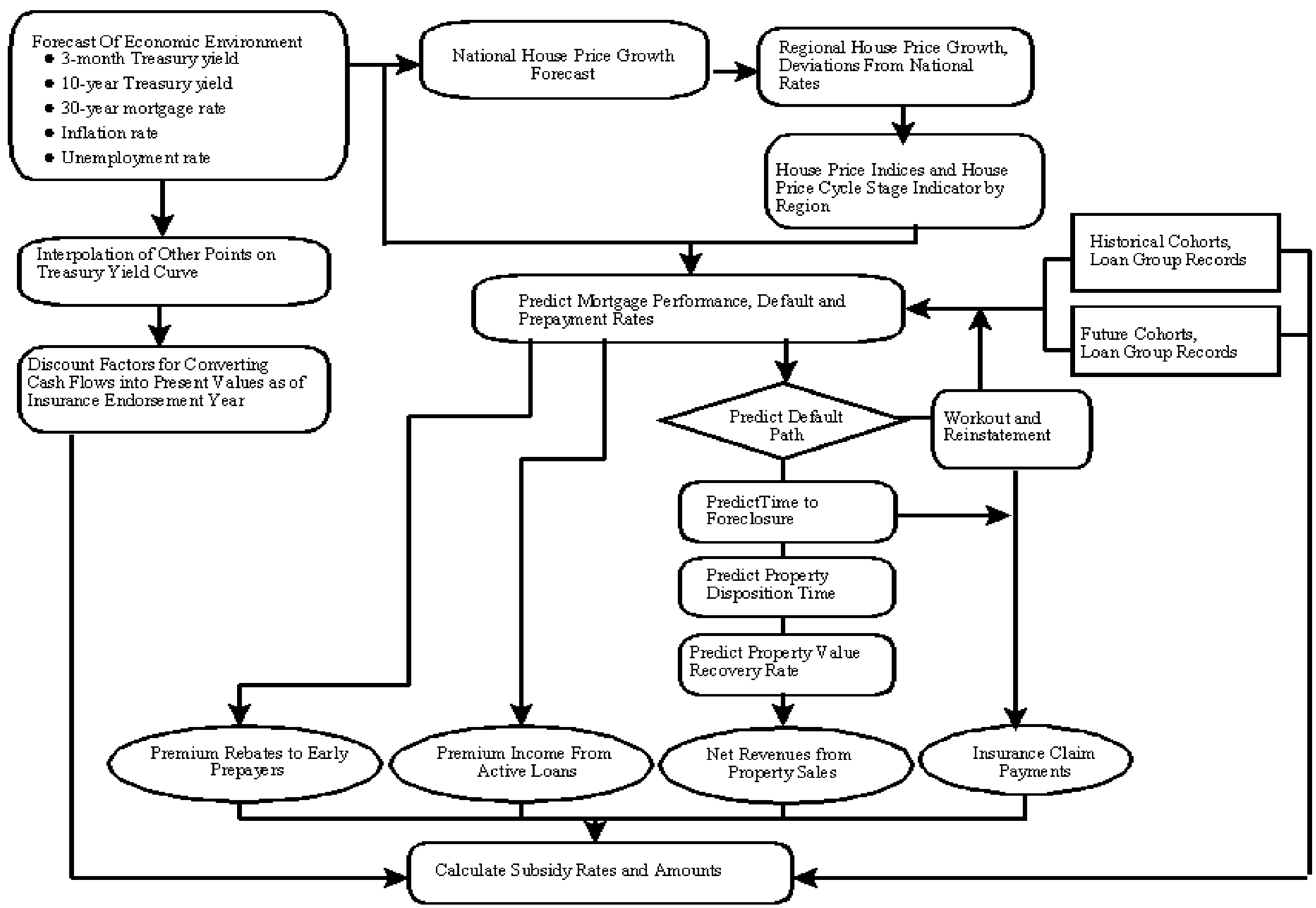


Figure 2. Mortgage Interest Rates, Historical Series and Forecast Confidence Bounds (percent)

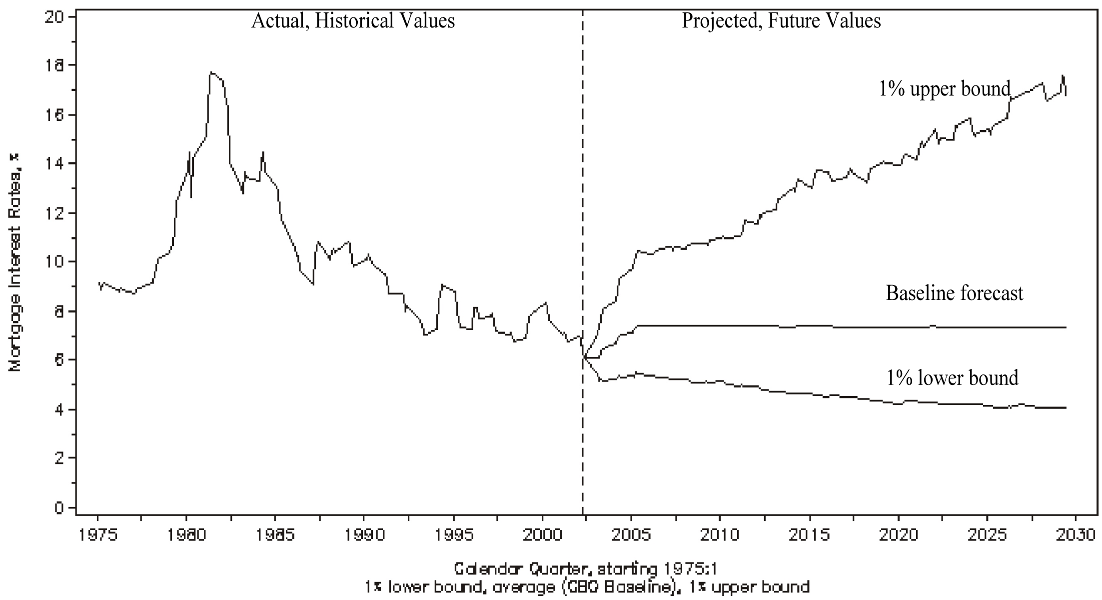


Figure 3. House Price Growth Rates, Historical and Forecast Confidence Bounds (percent, annual rate)

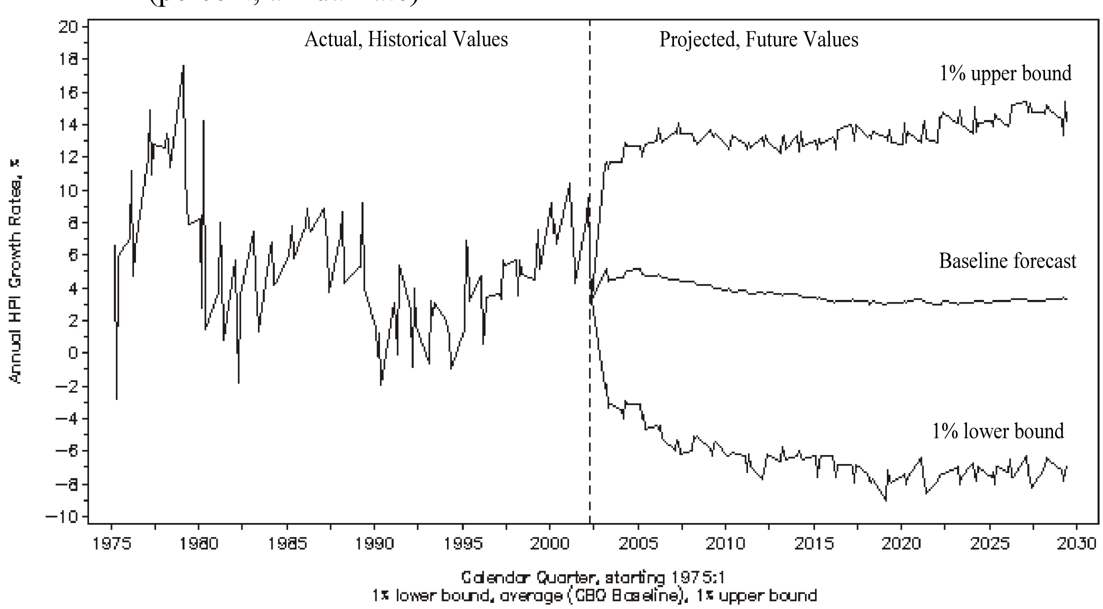


Figure 4. Confidence Bounds for Subsidy Rate Estimates, by Cohort

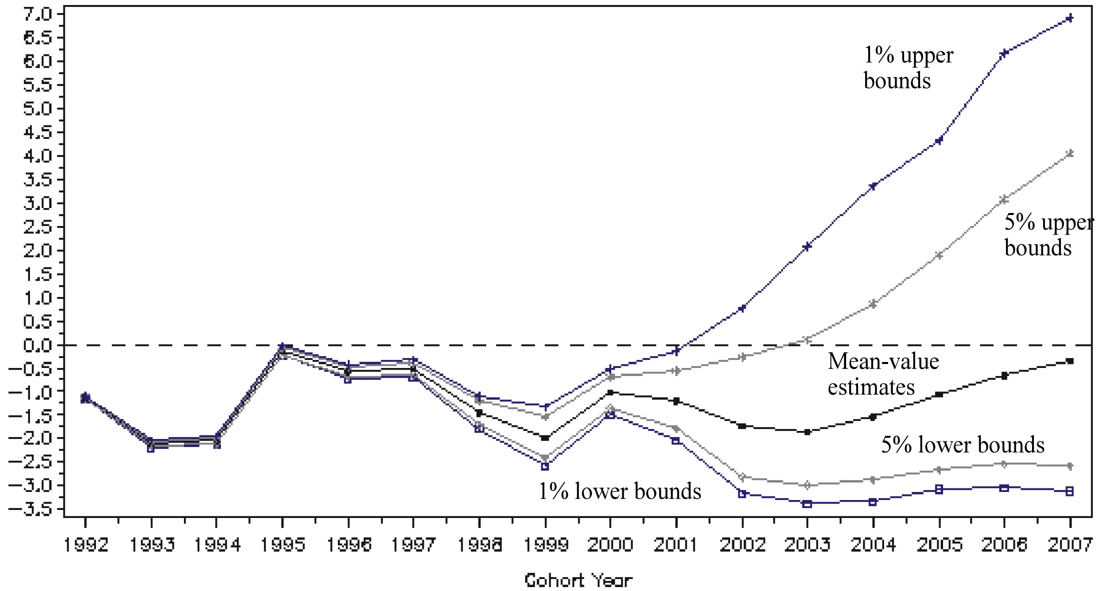


Figure 5. Frequency Distribution of Subsidy Rate Estimate for FY2003 Loan Guarantees (percentage)

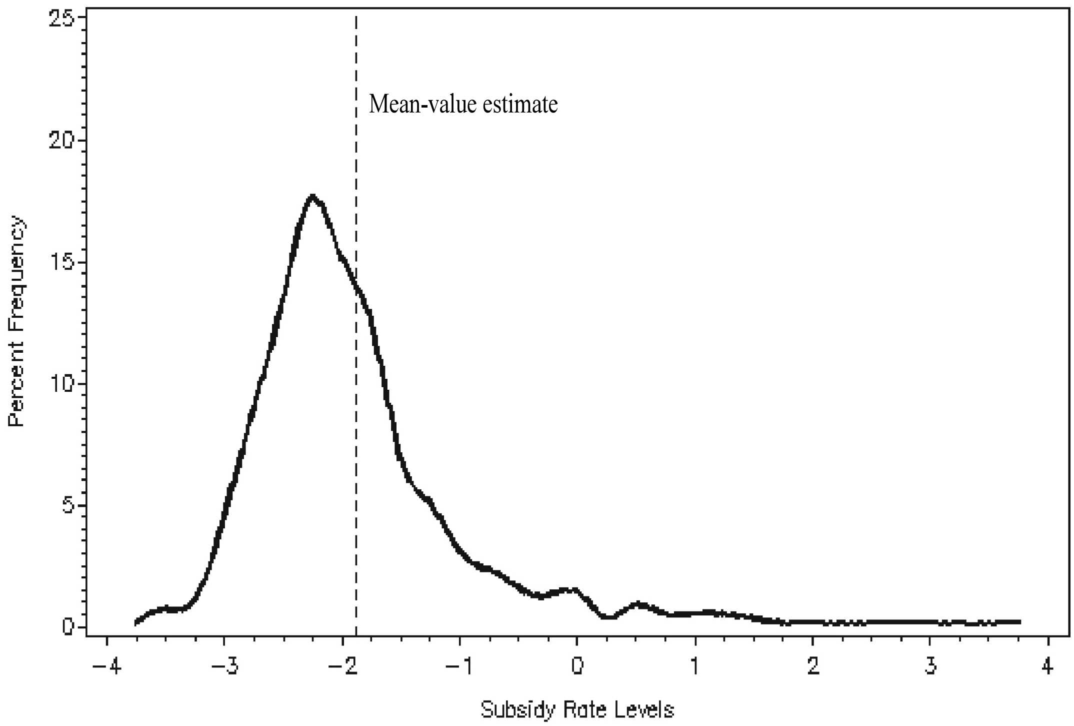


Figure 6. Frequency Distribution of Premium Revenue Component of Subsidy Rate Outcomes for FY2003 (percent)

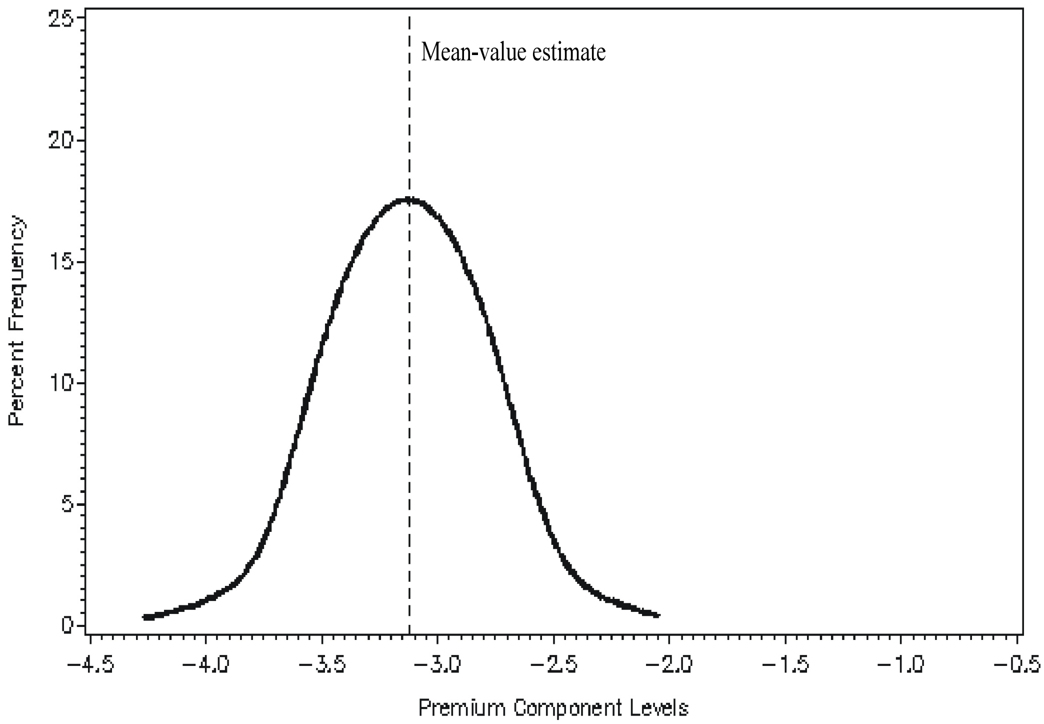
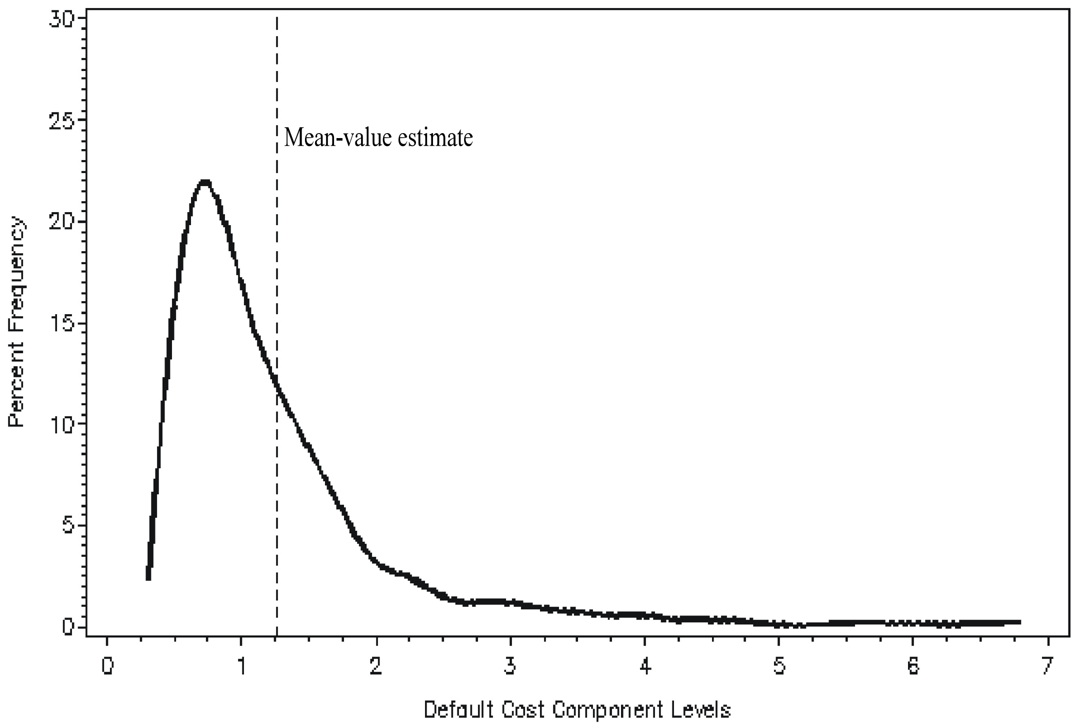
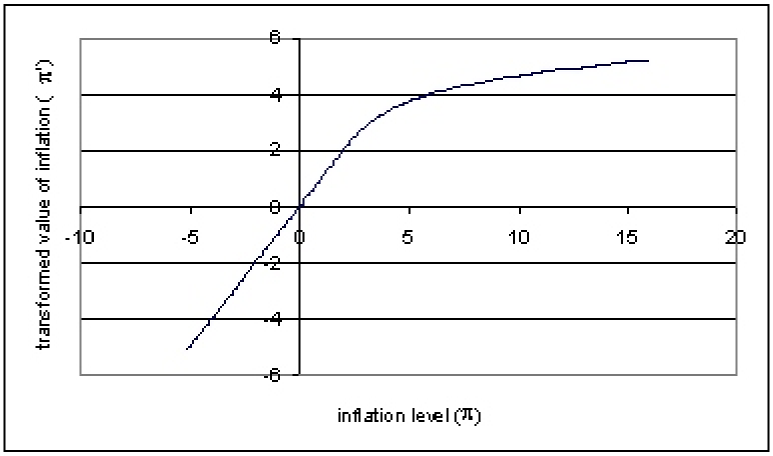


Figure 7. Frequency Distribution of the Default Cost Component of Subsidy Rate Outcomes for FY2003 (percent)



Box 1. Transformation of Inflation Variable for Use in the Vector Autoregression Model

Economic variables used in the vector autoregression (VAR) model (see Table 1) have larger variances at higher levels. To achieve uniform variances in the regression, four of the five variables are log transformed. Using a log transformation also eliminates negative values from the forecasts, which is too restrictive for inflation. Thus, a new transformation is derived here to provide a regression variable with a constant variance and which allows for negative values of inflation forecasts. The relationship between actual inflation rates and transformed values is shown in the following figure:



Derivation of the transformation function begins with a regression of the variance of the quarterly inflation time-series. A second-order regression of squared deviations of actual from expected inflation is performed on expected inflation:

$$(\pi - \pi^e)^2 = \alpha + \beta\pi^e + \gamma\,(\pi^e)^2 + \varepsilon\ ,$$

where π is actual inflation (annual rate), π^e is expected inflation, and ε is the residual (error) term. Expected inflation is a weighted moving average of past inflation rates (see description in Table A.2).

Generalized least squares estimates of the three regression parameters are:

$$\hat{\alpha} = 0.3355 \qquad \hat{\beta} = -0.2780 \qquad \hat{\gamma} = 0.0726$$

If we define $\sigma^2(\pi|\pi^e) = E[(\pi - \pi^e)^2]$ as the estimated variance of possible inflation rates in each period, conditional on recent experience as measured in π^e, the regression measures how that variance changes with the underlying level of inflation, indicated by π^e.

The estimated regression equation has a minimum value at $\pi_0 = -\hat{\beta}/2\hat{\gamma} = 1.9$ percent, and the standard deviation of inflation at that minimum value is, $\sigma(\pi_0) = \sqrt{\hat{\alpha} - \frac{\hat{\beta}^2}{4\hat{\gamma}}} = 0.26$ percent. The variance is held constant for values of $\pi < \pi_0$, rather than allowing it to rise. Thus, the general equation for the variance of inflation is:

$$\sigma^2(\pi) = \begin{cases} \hat{\alpha} + \hat{\beta}\pi^e + \hat{\gamma}(\pi^e)^2 & \pi^e > \pi_0 \\ 0.26 & \pi^e \le \pi_0 \end{cases}$$

(Continued)

Box 1 (continued)

The goal of the transformation process is to create a new variable, π', with a uniform variance for use in the VAR model. Resultingly, the transformation function, $f(\pi) = \pi'$, should exhibit the property that $\sigma(\pi)\dfrac{df(\pi)}{d\pi} = C$, where C is an arbitrary constant. $f(\pi)$ would then neutralize changes in the variance of π' otherwise caused by changes in the variance of the underlying level of π.

Rearranging terms yields $\dfrac{df(\pi)}{d\pi} = C\dfrac{1}{\sigma(\pi)}$, which implies, $f(\pi) = C\int\dfrac{1}{\sigma(\pi)}d\pi$. A closed-form solution for $f(\pi)$ is: $f(\pi) = C\cdot\left(\dfrac{1}{\sqrt{\hat{\gamma}}}\right)\cdot\sinh^{-1}\left(\dfrac{\pi-\pi_0}{\sigma_0/\sqrt{\hat{\gamma}}}\right) + D$,

where D is the constant of integration. Values of C and D are chosen so that $f(\pi_0) = \pi_0$ and $f'(\pi_0) = 1$, which assures that the implied, complete transformation function across all values of inflation will be continuous at $\pi = \pi_0 = 1.9$. Thus, $C = \sigma_0$ and $D=\pi_0$.

The final transformation function used to create the inflation variable in the VAR model is then:

$$\pi' = f(\pi) = \pi_0 + \left(\sigma_0/\sqrt{\hat{\gamma}}\right)\sinh^{-1}\left(\frac{\pi'-\pi_0}{\sigma_0/\sqrt{\hat{\gamma}}}\right)$$

The inverse transformation function needed to obtain inflation values from predictions of the VAR (transformed) inflation equation is:

$$\pi = f^{-1}(\pi') = \begin{cases} \pi_0 + \left(\sigma_0/\sqrt{\hat{\gamma}}\right)\sinh\left(\dfrac{\pi'-\pi_0}{\sigma_0/\sqrt{\hat{\gamma}}}\right) & \ldots\quad \pi' \geq \pi_0 \\ \pi' & \ldots\quad \pi' < \pi_0 \end{cases}$$

While values of $\pi < \pi_0$ do not appear in the regression sample, they may appear in the VAR predictions, especially after stochastic shocks are added (that is, $\pi' < \pi_0 = 1.9$). Applying no inverse transformation to predictions in this range creates the one-to-one correspondence between π' and π for values at or below 1.9. The lack of any inverse transformation on low and negative values of π' also imposes a constant standard deviation on inflation forecasts below 1.9 percent at $\sigma_0 = 0.29$ percent. The standard deviation of inflation forecasts above 1.9 percent will rise with π^e, as indicated by the regression equation shown near the top of this Box.

Box 2. Deriving Discount Factors By Interpolating Forward Rates

Discount factors used to create present-values of cash flows are specified as $1/(1+r)^t$, where r is an appropriate interest rate and t is the number of (semi-annual) time periods. FHA cash flows are accumulated annually, and discounted to year of insurance endorsement (loan origination) as if all cash flows occurred at mid-year. Thus, discount rates are needed for 6, 18, 30, 42 months, and so on. The VAR equations (Table 1) provide 3-month and 10-year rates, and the historical slope relationships (Table 5) provide rates for 6 months, 1 year, 2 years, 5 years, and 30 years. The intermediate points are arrived at by first calculating successive 6-month yields that complete each missing interval. These yields are called forward rates. Refer to these as f_1, f_2, and so on, each representing forward rates over successive six month intervals. We start with $f_1 = r_{6m}$, where r_{6m} is derived in Table 5. The resulting discount factor for interval one is: $\delta_1 = \frac{1}{\left(1 + f_1/2\right)}$. Successive discount factors are products of the individual forward-rate present-value factors. Because constant maturity Treasury (CMT) rates give yields for securities sold at par, the one-year yield (r_1) must be equivalent to what earnings on two successive 6-month investments, and f_2 can be solved for in the following equation:

$$1 = \delta_2 + (r_1/2)(\delta_1 + \delta_2) \text{ , where } \delta_2 = \left(\frac{1}{1+f_1/2}\right)\left(\frac{1}{1+f_2/2}\right)$$

This equation says the par yield (1=100%) must equal the present value of the principal payment after one year ($1 discounted with δ_2), plus the sum of the discounted value of two semi-annual interest payments at the one-year CMT yield, r_1. With r_1 and δ_1 now known, we can solve for the second period discount factor, δ_2 as:

$$\delta_2 = \left(1 - \left(\frac{r_1}{2}\right)\cdot \delta_1\right) \Big/ \left(1 + \frac{r_1}{2}\right)$$

and the forward rate for the second 6-month time period is: $f_2 = \left(\frac{\delta_1}{\delta_2} - 1\right)\cdot 2$

Beyond the second interval, multiple forward rates must be solved for simultaneously, each to complete the next interval along the yield curve. For the one-to-two-year interval, there are two forward rates, f_3 and f_4. Assuming forward rates follow a logarithmic relationship, we have: $f_3 = f_2 + (f_4 - f_3)\cdot\left(\frac{\ln(3/2)}{\ln(4/2)}\right)$.

This equation has two unknowns, f_4 and f_3, and is solved together with the market yield equivalence condition:

$$1 = \delta_1\left(\frac{r_2}{2}\right) + \delta_2\left(\frac{r_2}{2}\right) + \delta_3\left(\frac{r_2}{2}\right) + \delta_4\left(1 + \frac{r_2}{2}\right) \text{ , where } \delta_4 = \prod_{i=1}^{4}\left(1 \Big/ \left(1 + \frac{f_i}{2}\right)\right)$$

(continued)

Box 2 (continued)

An interval-bisection grid search is used to solve for values of f_3 and f_4 that best fit these two equations. The same procedure is used to compute forward rates between two-and-five years, five-and-ten, and ten-and-thirty. The series of forward rates is then used to compute spot rates that will be used for discounting cash flows. For example, the 18-month spot rate, r^s_{18m}, is computed from the first three forward rates: $r^s_{18m} = \left(2 \cdot \prod_{i=1}^{3} \left(1 + \frac{f_i}{2}\right)^{1/n}\right) - 2$.

And the resulting discount factor for 18-month cash flows is: $\delta^s_{18m} = \left(1 \Big/ \left(1 + \frac{r^s_{18m}}{2}\right)\right)^{3/2}$

Table 1. Vector Autoregression Equations

Explanatory Variables	Mortgage Rates Coefficient Estimate	Mortgage Rates Standard Error	Unemployment Coefficient Estimate	Unemployment Standard Error	Inflation Coefficient Estimate	Inflation Standard Error	Yield Curve Slope Coefficient Estimate	Yield Curve Slope Standard Error	Spread of Mortgage to Treasury Rate Coefficient Estimate	Spread of Mortgage to Treasury Rate Standard Error
Constant	0.0720*	0.0492	0.0329	0.0468	0.8062**	0.3956	-0.0349	0.0992	0.0679*	0.0465
MORTGAGE_RATE{1}	1.1341***	0.0946	0.1076	0.0900	0.6388	0.7603	-0.1163	0.1906	0.0555	0.0894
MORTGAGE_RATE{2}	-0.1864*	0.1377	-0.0738	0.1310	0.0927	1.1066	0.0606	0.2774	-0.1457	0.1302
MORTGAGE_RATE{3}	-0.0370	0.0840	-0.0042	0.0798	-0.6750	0.6747	-0.0430	0.1692	0.1278*	0.0794
US_UNEMP_RATE{1}	0.0759	0.1103	1.3698***	0.1049	-2.3991***	0.8862	0.6353***	0.2222	0.0965	0.1042
US_UNEMP_RATE{2}	-0.2587*	0.1780	-0.3629**	0.1693	3.7319***	1.4305	-0.7494***	0.3586	-0.0556	0.1682
US_UNEMP_RATE{3}	0.2287**	0.1056	-0.0815	0.1005	-1.2539*	0.8488	0.2310	0.2128	-0.0631	0.0998
INFLATION_RATE{1}	0.0112	0.0111	0.0143*	0.0106	0.3435***	0.0892	-0.0344*	0.0224	0.0089	0.0105
INFLATION_RATE{2}	0.0067	0.0088	0.0012	0.0084	0.6794***	0.0706	0.0075	0.0177	-0.0023	0.0083
INFLATION_RATE{3}	-0.0032	0.0112	-0.0018	0.0107	-0.2203***	0.0903	0.0311*	0.0226	-0.0241**	0.0106
YLD_CURVE_SLOPE{1}	-0.0502	0.0479	0.0082	0.0456	0.4151	0.3850	1.1383***	0.0965	-0.0568	0.0453
YLD_CURVE_SLOPE{2}	-0.0494	0.0737	0.0347	0.0700	-0.8023*	0.5918	-0.2685**	0.1484	0.0088	0.0696
YLD_CURVE_SLOPE{3}	0.0256	0.0542	-0.0663*	0.0515	-0.1698	0.4351	-0.0389	0.1091	0.0511	0.0512
SPREAD_TO_TREAS{1}	-0.6220***	0.1010	0.0735	0.0960	-1.7646**	0.8114	0.5650***	0.2034	0.6857***	0.0954
SPREAD_TO_TREAS{2}	0.6474***	0.1311	-0.1103	0.1246	0.5877	1.0531	-0.8242***	0.2640	-0.0297	0.1239
SPREAD_TO_TREAS{3}	0.0522	0.1138	0.0066	0.1082	-0.3371	0.9144	0.6609***	0.2292	0.0683	0.1075
SUMMARY STATISTICS										
Time Period	1968:1-2001:4									
Adjusted Coefficient of Determination (R-bar^2)	97.17		97.64		80.78		86.69		57.38	
Standard error of regressions	0.0382		0.0363		0.3066		0.0769		0.0361	

LEGEND

MORTGAGE_RATE	30-year fixed-rate mortgage coupon rate, Freddie Mac series, percent, log form
US_UNEMP_RATE	Unemployment rate, civilian, noninstitutionalized population, age 16 and older, percent, log form.
INFLATION_RATE	Quarterly change in the consumer price index, urban consumer, percent, transformed variable (see Box 1).
YLD_CURVE_SLOPE	Ratio of 10-year Treasury yield to 3 month Treasury bill rate, log form
SPREAD_TO_TREAS	Ratio of 30-year mortgage rate to10-year constant maturity Treasury yield, log form
{ i}	Lagged values, i = {1,2,3} quarters
Statistical Significance in one-sided tests	*** .01 level, ** .05 level, * .10 level

Table 2. National House-Price-Growth-Rate Regression

Variable Name	Coefficient Estimate	Standard Error	Description
Constant	1.2405***	0.3816	The model is estimated with seasonal effect dummies. The results are averaged to produce the intercept effect shown here.
Δ_*EXPECTED_INFLATION*	-0.6899***	0.1914	Expected inflation, forecast by CBO, change from previous quarter, multiplied by 4 to get annual rate
Δ_*REAL_MORT_RATE*	-0.6694***	0.0994	Real mortgage rate, Freddie Mac commitment rate, 30-year mortgages, less expected inflation, change from previous quarter, multiplied by 4 to match house price-growth which appears on the left-hand-side of the equation in an annualized rate.
Δ_*REAL_MORT_RATE{1}*	-0.2879***	0.0778	lagged difference in the real mortgage rate, multiplied by 4
Δ_*US_UNEMP_RATE*	-1.4522***	0.3712	US civilian, noninstitutionalized population, unemployment rate, age 16 and older, change from value four quarters previous

SUMMARY STATISTICS	
Time Period	1975Q2 - 2001Q1
Standard error of the regression	3.3740
Statistical significance in one-sided tests	*** .01 level, ** .05 level, * .10 level

Notes: The house price growth rate is four times the quarterly growth rate found in the OFHEO U.S. HPI series. All variables in regression are in percent form. See Table 1 for more details on variables used to create the difference-variables in this regression. Expected inflation comes from the CBO Philips Curve equation. It is a weighted average of the previous 12 quarters inflation (change in CPI), where the weights (starting with the most recent quarter) are: 0.4014, 0.2425, 0.1296, 0.0555, 0.0136, -0.0032, -0.0019 ,0.0108, 0.0277, 0.0420, 0.0468, 0.0351. All inputs for forecast simulations are generated using the vector autoregression model found in Table 1, along with the inverse transformation of inflation rates described in Box 1.

<table>
<tr><td>Table 3.</td><td colspan="2">Parameter Values for Stochastic Cumulative, Regional House-Price-Path Deviations from Cumulative, National House-Price Growth</td></tr>
<tr><th>Parameter</th><th>Value</th><th>Conditions</th></tr>
<tr><td colspan="3">Ornstein-Uhlenbeck Process: $\eta(t, k) = \eta(t, k\text{-}1)*(1- \alpha/92) + N(0, 2/92)$
$\eta(0,0) = N(0,1/\alpha)$ and $\eta(t,0) = \eta(t\text{-}1,92)$
Realized values of growth factors: $\gamma(t) = \eta(t, 92)*\varepsilon$,
Definitions: t = quarters, k= daily increments (k = {1,...,92})
$N(\ )$ = normal (Gaussian) random number</td></tr>
<tr><td>α for Divisions</td><td>.08</td><td rowspan="2">The MSA value is the most common value for describing 1980-2000 data movements. Division value increases impact of past values, smoothing the series.</td></tr>
<tr><td>α for MSAs</td><td>.10</td></tr>
<tr><td>ε for Divisions</td><td>.001</td><td rowspan="2">The MSA value is the most common for describing 1980-2000 data movements. Division value is set equal to MSA and differences in volatilities are a function of σ, below.</td></tr>
<tr><td>ε for MSAs</td><td>.001</td></tr>
<tr><td colspan="3">Cycle Phase add-on: $\sigma \cdot \cos(\Omega \cdot t + \rho)$</td></tr>
<tr><td>σ for Divisions</td><td>.05*(1+ N(0, .01))</td><td rowspan="2">Scale amplitude of cosine function, which cycles between -1 and +1. Scale factors are drawn as normal random variates with means of .05 or .09, and standard errors of one-tenth of those amounts. Actual realized values of σ are unique for each simulation iteration and locality.</td></tr>
<tr><td>σ for MSAs</td><td>.09*(1+ N(0, .01))</td></tr>
<tr><td>Ω for Divisions</td><td>.15*(1+ N(0, .01))</td><td rowspan="2">Average length of cycle is 10 years, with random deviations having standard error of 10 percent (1.5 years)</td></tr>
<tr><td>Ω for MSAs</td><td>.15*(1+ N(0, .01))</td></tr>
<tr><td rowspan="3">ρ</td><td>$U(0,\pi/2) + \pi$

Uniform random variate over the $(0,\pi/2)$ interval, and shifted by π. The phase parameter, ρ, is randomly chosen at the start of each simulation (iteration), within the segments defined here for each region.</td><td>when $(\dot{HPI}_r - \dot{HPI}_{us}) \leq -2$

$\dot{HPI}_r$ is average annual house price growth in region/locality r, over the most recent five year period (1997Q2- 2002Q1). $\dot{HPI}_{us}$ is the same for the US.</td></tr>
<tr><td>$U(0,\pi/2) + 0.50*\pi$</td><td>when $-2 < (\dot{HPI}_r - \dot{HPI}_{us}) \leq -1$</td></tr>
<tr><td>$U(0,\pi/2) + 0.75*\pi$</td><td>when $-1 < (\dot{HPI}_r - \dot{HPI}_{us}) \leq 0$</td></tr>
</table>

(continued)

Table 3.	Parameter Values for Stochastic Cumulative, Regional House-Price-Path Deviations from Cumulative, National House-Price Growth (continued)	
Parameter	Value	Conditions
ρ	$U(0,\pi/2) + 1.25*\pi$	when $0 < \left(H\dot{P}I_r - H\dot{P}I_{us}\right) \leq 1$
	$U(0,\pi/2) + 1.50*\pi$	when $1 < \left(H\dot{P}I_r - H\dot{P}I_{us}\right) \leq 2$
	$U(0,\pi/2) + 1.75*\pi$	when $2 < \left(H\dot{P}I_r - H\dot{P}I_{us}\right)$

Table 4. Treasury Rates Used in Credit Subsidy Calculations

Fiscal Year	1 year or less	More than 1 year and less than 5 years	5 years or more, and less than 10 years	10 years or more, and less than 20 years	20 years or more
1992	4.06	5.52	6.93	7.59	7.59
1993	3.25	4.58	5.90	6.65	6.65
1994	4.06	5.36	6.21	6.77	6.77
1995	5.99	6.78	7.11	7.26	7.39
1996	5.45	5.96	6.33	6.64	6.77
1997	5.50	6.19	6.51	6.77	6.89
1998	5.34	5.57	5.68	5.85	5.98
1999	4.76	5.11	5.36	5.67	5.81
2000	6.04	6.38	6.39	6.40	6.36
2001	4.52	4.76	5.22	5.58	5.75
2002	1.94	3.37	4.63	5.32	5.62

Source: U.S. Executive Office of the President, Office of Management and Budget (OMB).

Notes:

1. All rates are stated as simple annual rates, which differ from bond equivalent rates and bank discount rates. The 10-20 year and the 20+ year intervals were a single interval in the 1992-1994 Budgets. The separate intervals were effective with the 1995 Budget.
2. These actual fiscal-year averages derived from the prevailing market yields during the fiscal year, excluding the last five business days, on outstanding fixed-rate Treasury securities, categorized on the basis of their remaining maturity. The rates for 2002 are actual fiscal year averages derived from the prevailing market yields during the fiscal year, excluding the last ten business days, on outstanding fixed-rate Treasury securities, categorized on the basis of their remaining maturity.
3. The rates above are used to calculate credit subsidy cost, to reestimate credit subsidy cost, and to calculate the interest income and expense of the financing accounts. They can be used in the OMB credit subsidy calculator without modification.

Table 5. Computing Major Points on the Constant Maturity Treasury Yield Curve From the 3-month and 10-year Rates and Regression Equations

Yield Curve Point/Rate	Equation	Notes
3-month, bond equivalent yield	$r_{3m}^{b} = \left(\frac{r_{3m}}{100} * 365\right) \Big/ \left(360 - \frac{r_{3m}}{100} * 91\right)$	The 3-month yield must first be converted to a bond-equivalent form (365 day year, rather than 360 days) to compute longer-term yields.
Bound on the 10-year yield	$r_{10} = Max(r_{10}, 0.0001)$	Assures that any negative or near zero interest rates calculated in the VAR model (Table A.1) are converted to positive numbers.
Bound on the 3-month yield	$if\ r_{3m}^{b} < 0.0001\ then\ r_{3m}^{b} = 0.50 \cdot r_{10}$	Assures nonnegative results and that yield curves close to zero remain upward sloping.
6-month yield	$r_{6m} = \left(\frac{1036 * r_{3m}^{b}}{100} * 365\right) \Big/ \left(360 - \frac{1.036 * r_{3m}^{b}}{100} * 182\right)$	1.036 is the estimated coefficient found by regressing the 6-month yield on the 3-month yield.
1-year yield	$r_{1}^{b} = \left(\frac{1.054 * r_{3m}^{b}}{100} \Big/ 2 + 1\right)^{2} - 1$	1.054 is the estimated coefficient found by regressing the 1-year yield on the 3-month yield.
	$r_{1} = 100 * \left(\frac{360 * r_{1}^{b}}{365 + 364 * r_{1}^{b}}\right)$	
2-year yield	$r_{2} = 1.091 * r_{3m}$	1.091 is the estimated coefficient found by regressing the 2-year yield on the 3-month yield.
5-year yield	$r_{5} = Max\left\{1.054 * r_{3m}^{b}, \left(r_{3m}^{b} + 0.83 * \left(r_{10} - r_{3m}^{b}\right)\right)\right\}$	0.83 1 is the estimated coefficient found by regressing the 5-year yield on the 10-year yield.
30-year yield	$r_{30} = 1.029 * r_{10}$	1.029 is the estimated coefficient found by regressing the 30-year yield on the 10-year yield.

Source: U.S. Congressional Budget Office

Table 6. Explanatory Variables Used in Mortgage Performance Regressions

Variable Names	class	Category Values	Descriptions
		Loan Group Characteristics	
LTV_CLASS	1	up to 80%	Original loan-to-value ratio (*LTV*), where value is the lesser of purchase price or appraised value. Streamline refinance loans rarely have appraisals, and so *LTV*s are computed as either 97 (single-family) or 100 (condominiums) plus the up-front premium. In the few cases where value is missing for purchase mortgages, the same values are used, plus the 100 *LTV* is assigned to new construction.
	2	81-90%	
	3	91-95%	
	4	96-100%	
	5	over 100%	
PRICE_CLASS	1	under 50%	Based on ratio of house price to area (MSA) median house price.
	2	51-100%	
	3	over 100%	
PROP_AGE_CLASS	1	under 1 year	Age of house at time of loan origination.
	2	1-15 years	
	3	16-30 years	
	4	over 30 years	
COHORT_YEAR		1975-2000	Fiscal year of loan insurance endorsement.
INVESTOR		NO/YES	Investor properties are identified in FHA data bases since 1992. Before that time, investor loans are identified by *LTV*s between 84.5 and 85.5 because investors were required to have a minimum 15 percent downpayment. This criterion was first used in the original (1989) actuarial study of the FHA MMIF portfolio, performed by PriceWaterhouse (now PriceWaterhouseCoopers). Beyond this, if a loan cannot be identified as an owner-occupied property using the FHA data fields, then it is assigned to the investor class if the *LTV* is under 90 percent.
AGE_CLASS	1	1-2 years	Age of mortgage, from quarter of origination.
	2	3-4 years	
	3	5-6 years	
	4	7-8 years	
	5	9-10 years	
	6	11-15 years	
	7	16+ years	
		Economic Conditions	
CYCLE_STAGE	1	initial recession	House price cycles are defined as periods in which prices decline at least 3 percent, using a five-quarter moving-average of the OFHEO HPI. The moving average removes the significant quarterly volatility in the HPI series. Stage 1 includes all observation quarters during the first half of a price decline, Stage 2 covers the time period of the second half of the price decline, and Stage 3 is the time during which prices recover out of Stage 2. Stage transitions are based on movements in the smoothed HPI, not time. Cyclical peaks and troughs are identified by comparing quarterly (smoothed) HPI values with five quarters of leads and lags. Cycles are based on HPI paths from quarter of loan origination, by MSA and Census Division. Observations not in one of these three stages are labeled as stage 4 (out-of-cycle).

(continued)

Table 6. Explanatory Variables Used in Mortgage Performance Regressions (continued)

Variable Names	class	Category Values	Description
CYCLE_STAGE	2	latter recession	
	3	early expansion	
	4	out of cycle	
US_UNEMP_RATE			National unemployment rate, civilian, non-institutionalized population, age 16 and over (in percent).
CUR_MKT_RATE			Median contract rate on FHA originations in the current quarter, decimal form. Used only in ARM regressions to model incentives to prepay/refinance into fixed-rate products.
YLD_SLOPE_CAT	1	up to 0%	Ten-year constant maturity Treasury (zero coupon bond) yield, minus the one-year constant maturity Treasury yield.
	2	0.1 - 1%	
	3	1.1 - 2%	
	4	2.1 - 3.0%	
	5	over 3.0%	
MKT_SHIFT		OFF/ON	Observations starting 1993Q1.
		Financial Incentives	
NEG_EQ_CLASS	1	over 50%	Probability of negative equity for properties in a loan group, each quarter. It is the probability that an individual property could have value erosion large enough to offset both the original downpayment and additional loan amortization, over time. Values are computed as integrals of the standard normal density function, up to the limit defined by: the log of the ratio of current loan balance to original-price-times-HPI, divided by the standard deviation of *log*(HPI). The *log*(HPI) is cumulative house price growth implied by the HPI, and its standard deviation is computed from so-called volatility parameters published by OFHEO in its quarterly *HPI Report*. The ratio described here is a standard-normal random variate, which follows from an assumption that house price growth rates following a random walk.
	2	26-50%	
	3	11-25%	
	4	2-10%	
	5	up to 1%	
POS_EQ_CLASS	1	over 75%	Probability of having at least 20% equity in a property, each quarter. It is the probability that an individual property could have experienced enough price appreciation, given the initial downpayment and on-going loan amortization, to reach a 20% equity threshold. It is analogous to *NEG_EQ_CLASS*, only measured as the density mass in the upper tail of the cumulative house-price-growth distribution, and where the limit of integration is: log of, HPI divided by loan-balance-over-original-house-price plus 0.20, divided by the standard deviation of *log*(HPI).
	2	51 - 75%	
	3	26 - 50%	
	4	11 - 25%	
	5	0 - 10%	
SPREAD_CLASS	1	under -.04	Current rate on mortgages minus the mortgage coupon rate, in decimal. Current rates are median values from FHA loan originations in each calendar quarter. Separate median values are found for (30-year) fixed-rate and adjustable-rate mortgages. An implied median value for 15-year loans is computed as the 30-year rate less 0.0050 (one-half percent).
	2	-.04 - -.03	
	3	-.029 - -.02	
	4	-.019 - -.015	
	5	-.014 - -.010	
	6	-.090 - -.005	
	7	-.005 - 0.00	
	8	.001 - .005	
	9	.005 - .010	
	10	.011 - .015	

(continued)

Table 6. Explanatory Variables Used in Mortgage Performance Regressions (continued)

Variable Names	class	Category Values	Description
SPREAD_CLASS	11	.016 - .020	
	12	.021 - .030	
	13	.031 - .040	
	14	over .04	
BURN_SUM			Number of past quarters of loan life in which *SPREAD_CLASS* = {1,2,3}.
NEW_REFI		OFF/ON	Turned on when *SPREAD_CLASS*={1,2,3,4} in current quarter, but had higher values in lagged quarters 3-7.
PMT_ADJ_CLASS	1	under -5%	Percent change in monthly payments of principal and interest for adjustable-rate mortgages, from previous year. Once *PMT_ADJ_CLASS* is established, that value remains for the entire 4-quarter rate period. Adjustments are based on interest rates prevailing in the quarter preceding the adjustment quarter (the fully indexed market rate), and annual and lifetime interest rate change caps on FHA ARMs (1% and 5%, respectively).
	2	-4 - 0%	
	3	1 - 5%	
	4	6 - 10%	
	5	over 10%	

Table 7. ANOVA Tests of the Statistical Significance and Relative Influence of Explanatory Variables in Default-Rate Regression Analysis

		Wald χ^2 Statistic Values, by Regression Equation			
Variable	DoF	Fixed Rate, Purchase	Adjustable Rate, Purchase	Fixed Rate, Refinance	Adjustable Rate, Refinance
AGE_CLASS	6	6958	793	1017	180
COHORT_YEAR	25	16631	1105	970	108
INVESTOR	1	890	-	178	-
MKT_SHIFT	1	925	9	117	**2**
LTV_CLASS	3	535	117	373	**1**
PRICE_CLASS	2	4705	1660	52	121
PROP_AGE_CLASS	3	575	315	147	100
NEG_EQ_CLASS	4	29928	7598	7251	1412
CYCLE_STAGE	3	2745	81	146	-
US_UNEMP_RATE	1	3302	910	1333	83
SPREAD_CLASS	13	18794	2611	2390	120
PMT_ADJ_CLASS	4	-	279	-	103
CUR_MKT_RATE	1	-	185	-	**1**
BURN_SUM	1	864	70	14	21
NEW_REFI_OPP	1	346	50	15	**1**

Notes: These are ANOVA Type III, partial sums of squares tests. Bold faced values are *NOT* statistically significant at the 0.01 level. DoF represents degrees-of-freedom in the test. It is the number of free parameters that are estimated for each variable. For classification variables, this is the number of classes/categories minus one. For single-class (on/off or dummy) and continuous variables, DoF equals one. Ranking of Wald Statistic values within each regression equation indicates ranking of importance of each variable in measuring differences in default rate outcomes. See Table 6 for a description of variables.

Table 8. ANOVA Tests of the Statistical Significance and Relative Influence of Explanatory Variables in Prepayment-Rate Regression Analysis

		Wald χ^2 -Squared Statistic Values, by regression equation			
Variable	DoF	Fixed Rate, Purchase	Adjustable Rate, Purchase	Fixed Rate, Refinance	Adjustable Rate, Refinance
AGE_CLASS	6	53000	2677	7236	462
COHORT_YEAR	25	94455	5219	2799	1486
INVESTOR	1	170	-	8	-
MKT_SHIFT	1	31114	360	7186	**0**
LTV_CLASS	3	1906	734	1615	**113**
PRICE_CLASS	2	48982	3952	4345	531
PROP_AGE_CLASS	3	2104	247	169	44
POS_EQ_CLASS	4	55365	17793	14274	641
CYCLE_STAGE	3	3241	89	60	
US_UNEMP_RATE	1	2890	2715	344	59
SPREAD_CLASS	13	353507	17996	42135	943
YLD_SLOPE_CLASS	4	27579	3800	4714	537
PMT_ADJ_CLASS	4	-	2030	-	171
CUR_MKT_RATE	1	-	9232	-	723
BURN_SUM	1	27002	2590	730	55
NEW_REFI_OPP	1	8817	214	119	**3**

Notes: These are ANOVA Type III, partial sums of squares tests. Bold faced values are *NOT* statistically significant at the 0.01 level. DoF represents degrees-of-freedom in the test. It is the number of free parameters that are estimated for each variable. For classification variables, this is the number of classes/categories minus one. For single-class (on/off or dummy) and continuous variables, DoF equals one. Ranking of Wald Statistic values within each regression equation indicates ranking of importance of each variable in measuring differences in prepayment rate outcomes. See Table 6 for a description of variables.

Table 9. Results of Default- Rate Logistic Regressions

Variable Names	class	Category Values	Coefficient Estimates By Loan Type: Fixed Rate, Purchase	Adjustable Rate, Purchase	Fixed Rate, Refinance	Adjustable Rate, Refinance
CONSTANT		1	-4.0295***	-5.0167***	-3.9235***	-3.7708***
		Loan Group Characteristics				
LTV_CLASS	1	up to 80%	**0.0450**	**-0.2235**	**0.1821**	**-0.1689**
	2	81-90%	0.0450***	-0.2235*	0.1821***	-0.1689
	3	91-95%	-0.1601***	-0.2501***	-0.0255	0.1840
	4	96-100%	0.0477***	0.2273***	0.0424***	0.0249
	5	over 100%	*0.0674****	*0.2463****	*-0.1990****	*-0.0400*
PRICE_CLASS	1	under 50%	0.2368***	0.2779***	0.0874***	0.3660***
	2	51-100%	0.0100***	0.0485***	-0.0143*	0.0322
	3	over 100%	*-0.2468****	*-0.3264****	*-0.0731****	*-0.3982****
PROP_AGE_CLASS	1	under 1 year	0.1138***	0.1443***	0.2195***	0.3868***
	2	1-15 years	-0.0388***	-0.0184**	-0.0006	0.0750
	3	16-30 years	-0.0550***	-0.0968***	-0.1793***	-0.2694***
	4	over 30 years	*-0.0200*	*-0.0291***	*-0.0396*	*-0.1924****
COHORT_YEAR		1975	-0.1718***		-2.7644	
		1976	-0.289***		-2.8447	
		1977	-0.3044***		0.9507	
		1978	-0.1080***		1.1248***	
		1979	-0.0785***		0.9082	
		1980	0.1224***		0.2626	
		1981	0.3761***		0.5844	
		1982	0.5486***		-0.0156	
		1983	0.1125***		0.3017	
		1984	0.3703***		0.8382***	
		1985	0.2634***		0.5301*	
		1986	-0.0691***	-0.2151***	0.1641	
		1987	-0.2956***	-0.1629***	0.1511	
		1988	-0.3268***	-0.4212***	0.1765	
		1989	-0.3813***	-0.3892***	0.1994	
		1990	-0.4769***	-0.5009***	0.2233	
		1991	-0.4364***	-0.1627***	0.1266	-0.5174
		1992	-0.4578***	-0.1243***	-0.4104	-0.2607*
		1993	-0.3955***	-0.2053***	-0.4553	-0.1271
		1994	-0.1973***	-0.1530***	-0.4001	-0.0522
		1995	0.0759***	0.2098***	-0.0235	0.3585***
		1996	0.1882***	0.3560***	0.0013	0.1976*
		1997	0.2929***	0.3984***	0.1039	0.3068**
		1998	0.3560***	0.3505***	-0.0452	-0.2340*
		1999	0.5672***	0.5188***	0.0335	-0.0432
		2000	*0.7149****	*0.5011****	*0.2788*	*0.3717****

(continued)

Table 9. Results of Default- Rate Logistic Regressions (continued)

Variable Names	class	Category Values	Coefficient Estimates By Loan Type: Fixed Rate, Purchase	Adjustable Rate, Purchase	Fixed Rate, Refinance	Adjustable Rate, Refinance
INVESTOR		NO	-0.3350***		-0.1588***	
		YES	*0.3350****		*0.1588****	
AGE_CLASS	1	1-2 years	-0.2725***	0.8015	-0.0104	0.2409***
	2	3-4 years	0.1445***	0.9904	0.3789***	0.6148***
	3	5-6 years	0.2165***	0.7253	0.3075***	0.1967***
	4	7-8 years	0.1216***	0.4653	0.2375***	-0.3973***
	5	9-10 years	0.0705***	0.2563	-0.0369	-0.6551
	6	11-15 years	-0.0103	0.0451	-0.3586***	**-0.6551**
	7	16+ years	*-0.2703****	**-0.6551**	*-0.5180****	**-0.6551**
		Economic Conditions				
CYCLE_STAGE	1	initial recession	0.1999***	0.3381*	0.1878***	**0.3381**
	2	latter recession	0.2004***	0.5985***	0.2123***	**0.5985**
	3	early expansion	-0.1314***	-0.2289	-0.1037	**-0.2289**
	4	out of cycle	*-0.2689****	*-0.7077****	*-0.2964****	0
US_UNEMP_RATE..................................			-0.1497***	-0.4444***	-0.3106***	-0.6430***
CUR_MKT_RATE....................................				17.1784***		4.3519
MKT_SHIFT		OFF	0.1210***	-0.0747***	0.1515***	-0.4279
		ON	*-0.1210****	*0.0747****	*-0.1515****	*0.4279*
		Financial Incentives				
NEG_EQ_CLASS	1	over 50%	0.9087***	1.1734***	1.1452***	1.3355***
	2	26-50%	0.3191***	0.1693***	0.4761***	0.0731
	3	11-25%	0.0112***	-0.2680***	0.0157	-0.2901***
	4	2-10%	-0.3141***	-0.2893***	-0.4798***	-0.5007***
	5	up to 1%	*-0.9249****	*-0.7854****	*-1.1572****	*-0.6178****
SPREAD_CLASS	1	under -.04	0.9473***	1.2732***	1.7767	**1.509**
	2	-.04 - -.03	0.8160***	1.0779***	1.3324	**1.2775**
	3	-.029 - -.02	0.6505***	0.7143***	0.9943	0.8466***
	4	-.019 - -.015	0.5663***	0.5343***	0.9342	0.7196***
	5	-.014 - -.010	0.3691***	0.3629***	0.7632	0.6618***
	6	-.090 - -.005	0.1963***	0.1857**	0.4903	0.4296***
	7	-.005 - 0.00	-0.0188***	-0.0881	0.3082	0.3489***
	8	.001 - .005	-0.1281***	-0.4622***	0.1207	-0.0574
	9	.005 - .010	-0.1987***	-0.5035***	-0.0147	-0.1815*
	10	.011 - .015	-0.3444***	-0.6199***	-0.2222	-0.0652
	11	.016 - .020	-0.4290***	-0.9807***	-0.4633	-0.7254***
	12	.021 - .030	-0.4393***	-1.1820***	-0.7316	**-1.1455**
	13	.031 - .040	-0.7759***	**-2.0877**	-0.4946	**-0.7744**
	14	over .04	*-1.2113****	**-3.2592**	**-2.0173**	**-3.1585**
		BURN_SUM	-0.0214***	0.0214***	-0.0168***	0.0513***
NEW_REFI		OFF	0.0969***	0.0567***	0.0699***	-0.0285
		ON	*-0.0969****	*-0.0567****	*-0.0699****	*0.0285*

(continued)

Table 9. Results of Default- Rate Logistic Regressions (continued)

Variable Names	class	Category Values	Fixed Rate, Purchase	Adjustable Rate, Purchase	Fixed Rate, Refinance	Adjustable Rate, Refinance
			Coefficient Estimates By Loan Type			
PMT_ADJ_CLASS	1	under -5%		0.2276***		0.0810
	2	-4 - 0%		-0.0976***		-0.3042***
	3	1 - 5%		-0.0090		-0.0328
	4	6 - 10%		-0.0416***		0.0986***
	5	over 10%		*-0.0794****		*0.1574****
Summary Statistics						
Loan-group observations			1,977,847	458,595	357,064	60,347
Likelihood ratio $\chi 2$			166,722	26,757	18,039	3250
Degrees of Freedom (DoF)			(DoF=64)	(DoF=56)	(DoF=64)	(DoF=43)

Notes: Italics represent coefficient values for classes omitted from regressions, calculated as the negative sum of estimated class effects; boldface represent imputed coefficient values, as described in Section III. Standard indicators of statistical significance are marked as: *** for 0.01 level (one-sided test), ** for 0.05 level, and * for 0.10 level. These indicators, however, are less meaningful for categorical variables. What matters is the significance of differences between the coefficients, and not differences of the coefficients from zero. Missing cells represent a lack of loan records with variable values in these categories. The fixed-rate-purchase equation was estimated on a one-in-three sample of loan groups, where each started with at least 10 loans. Each sampled loan group provides quarterly time series of observations, through 2000Q4, the end of year twenty of loan life, or until all loans terminate, whichever comes first. The number of actual loan quarters represented by the loan-group observations listed in the Summary Statistics is typically very large. For example, for the fixed-rate-purchase equation, the number of loan quarters in the regression estimation is over 65 million, so that, on average, each loan-group provides 33 quarters of observations.

Table 10. Results of Prepayment-Rate Logistic Regressions

Variable Names	class	Category Values	Coefficient Estimates By Loan Type: Fixed Rate, Purchase	Adjustable Rate, Purchase	Fixed Rate, Refinance	Adjustable Rate, Refinance
CONSTANT		1	-3.4581***	1.6821***	-3.409***	-0.0143
		Loan Group Characteristics				
LTV_CLASS	1	up to 80%	**0.0000**	**0.0000**	**-0.1615**	**-0.2936**
	2	81-90%	-0.0621***	-0.0052	-0.1615***	-0.2936***
	3	91-95%	0.0199***	0.0337***	-0.0382***	0.2006***
	4	96-100%	0.0553***	0.0451***	0.0746***	-0.0426
	5	over 100%	*-0.0131*	*-0.0736*	*0.1251****	*0.1356***
PRICE_CLASS	1	under 50%	-0.3430***	-0.2428***	-0.2352***	-0.2749***
	2	51-100%	0.0215***	0.0489***	-0.00206	0.0118
	3	over 100%	*0.3561****	*0.3164****	*0.1101****	*0.1393****
PROP_AGE_ CLASS	1	under 1 year	0.0572***	-0.0025	0.1484***	0.2972***
	2	1-15 years	0.0262***	0.0269***	-0.0249***	-0.1091***
	3	16-30 years	-0.0555***	0.0055**	-0.0566***	-0.1164***
	4	over 30 years	*-0.0279****	*-0.0299****	*-0.0669*	*-0.0717*
COHORT_YEAR		1975	0.6621***		0.2594	
		1976	0.6276***		0.6529***	
		1977	0.6306***		0.7489***	
		1978	0.3238***		0.7621***	
		1979	-0.1602***		-0.0112	
		1980	-0.3728***		-0.0471	
		1981	-0.3676***		0.6156***	
		1982	0.0207**		0.1851	
		1983	-0.2774***		0.0572	
		1984	-0.2826***		-0.2459***	
		1985	-0.1790***		-0.148***	
		1986	-0.4499***		-0.4321***	
		1987	-0.3914***	0.1718***	-0.4508***	
		1988	-0.3807***	0.2944***	-0.5589***	
		1989	-0.2674***	0.1352***	-0.4388***	
		1990	-0.1615***	0.0065	-0.3724***	
		1991	-0.0971***	-0.1827***	-0.2693***	-0.4734
		1992	0.0281***	-0.2265***	-0.1405***	0.0849
		1993	0.0505***	-0.2526***	-0.2093***	0.0453
		1994	-0.0213***	-0.3190***	-0.2276***	-0.017
		1995	-0.0842***	-0.0599***	-0.2681***	0.3498
		1996	-0.0811***	0.0024	-0.123***	0.7846
		1997	-0.0235***	-0.0556***	-0.00654	0.8410*
		1998	-0.0833***	-0.1624***	0.0377	0.8014
		1999	-0.1192***	-0.0228	-0.1761***	0.6762
		2000	0.6925***	0.1216***	0.1024***	0.6159
		2001	*0.7643****	*0.5496****	*0.7043****	*-3.7087*

(continued)

Table 10. Results of Prepayment-Rate Logistic Regressions (continued)

			Coefficient Estimates By Loan Type			
Variable Names	class	Category Values	Fixed Rate, Purchase	Adjustable Rate, Purchase	Fixed Rate, Refinance	Adjustable Rate, Refinance
INVESTOR		NO	0.0807***		-0.0137***	
		YES	*-0.0807****		*0.0137****	
AGE_CLASS	1	1-2 years	0.1717***	0.3795***	0.8303***	0.3247***
	2	3-4 years	0.5187***	0.3199***	0.5625***	-0.0709
	3	5-6 years	0.3698***	0.0306	0.3548***	-0.2664***
	4	7-8 years	0.1566***	-0.0066	0.1766***	-0.1374
	5	9-10 years	-0.1138***	-0.0947***	-0.3868***	-0.1897**
	6	11-15 years	-0.3387***	-0.3473***	-0.4611***	**-0.3473**
	7	16+ years	*-0.7643****	**-0.7837**	**-1.0763**	**-0.7837**
		Economic Conditions				
CYCLE_STAGE	1	initial recession	0.2717***	0.8743***	0.1716***	**0.5522**
	2	latter recession	-0.0633***	0.0490	-0.0418	**0.0324**
	3	early expansion	-0.1927***	-1.0322***	-0.1623***	**-0.8694**
	4	out of cycle	*-0.0157*	*0.1089*	*0.0325*	**0.0000**
US_UNEMP_RATE..			-0.1005***	-0.3813***	-0.1314***	-0.1926***
CUR_MKT_RATE..				-49.3482***		-45.4862***
YLD_SLOPE_CAT	1	up to 0%	-0.3950***	-0.0557***	-0.3650***	-0.095
	2	0.1 - 1%	-0.3110***	-0.1514***	-0.4058***	-0.2171
	3	1.1 - 2%	-0.0629***	0.0652***	-0.2977***	0.0483
	4	2.1 - 3.0%	0.1979***	-0.3956***	0.1900***	-0.612
	5	over 3.0%	*0.5710****	*0.5375****	*0.8785****	*0.8758****
MKT_SHIFT		OFF	-0.3175***	-0.2638***	-0.4861***	-1.0753
		ON	*0.3175****	*0.2638****	*0.4861****	*1.0753****
		Financial Incentives				
POS_EQ_CLASS	1	over 75%	0.5080***	0.5615***	0.6144***	0.3686***
	2	51 - 75%	0.1950***	0.2469***	0.2520***	0.1672***
	3	26 - 50%	0.0006	0.0302***	-0.0830***	-0.00451
	4	11 - 25%	-0.1284***	-0.1934***	-0.2764***	-0.1766***
	5	0 - 10%	*-0.5752****	*-0.6452****	*-0.5070****	*-0.3547****
SPREAD_CLASS	1	under -.04	2.0141***	1.8872***	1.8986***	**0.9493**
	2	-.04 - -.03	1.6354***	1.0198***	1.6058***	0.4999***
	3	-.029 - -.02	1.2572***	0.9507***	1.3326***	0.5701***
	4	-.019 - -.015	0.8084***	0.8422***	1.0571***	0.5284***
	5	-.014 - -.010	0.6776***	0.6909***	0.8225***	0.3803***
	6	-.090 - -.005	0.2162***	0.4579**	0.4581	0.2787***
	7	-.005 - 0.00	-0.2366***	0.0701	-0.00741	-0.0263
	8	.001 - .005	-0.4589***	-0.4550*	-0.2512	-0.3826***
	9	.005 - .010	-0.5883***	-0.4746**	-0.4684	-0.4496***
	10	.011 - .015	-0.6213***	-0.5769**	-0.6588***	-0.3438***

(continued)

Table 10. Results of Prepayment-Rate Logistic Regressions (continued)

Variable Names	class	Category Values	Coefficient Estimates By Loan Type: Fixed Rate, Purchase	Adjustable Rate, Purchase	Fixed Rate, Refinance	Adjustable Rate, Refinance
SPREAD_CLASS	11	.016 - .020	-0.6473***	-0.9166***	-0.8613***	-0.4333***
	12	.021 - .030	-0.8449***	-0.8028***	-1.1047***	**-0.5524**
	13	.031 - .040	-1.3468***	**-1.3468**	**-1.3468**	**-0.6734**
	14	over .04	*-1.8648****	**-1.8648**	**-1.8648**	**-0.9324**
BURN_SUM			-0.0548***	0.0214***	-0.0444***	-0.0267***
NEW_REFI		OFF	-0.1651***	0.0567***	-0.0574***	0.0154*
		ON	*0.1651****	*-0.0567****	*0.0574****	*-0.0154**
PMT_ADJ_CLASS	1	under -5%		0.1461***		-0.0125
	2	-4 - 0%		-0.1055***		-0.0930***
	3	1 - 5%		-0.0839***		-0.0369***
	4	6 - 10%		-0.0207***		0.0205*
	5	over 10%		*0.0640****		*0.1219****
Summary Statistics						
Loan-group observations			2,135,432	545,235	400,417	72,041
Likelihood ratio $\chi 2$			1,460,078	328,091	178,878	27,720
Degrees of Freedom			(DoF=69)	(DoF=61)	(DoF=69)	(DoF=50)

Notes: Italics represent coefficient values for classes omitted from regressions, calculated as the negative sum of estimated class effects; boldface represent imputed coefficient values, as described in section III. Standard indicators of statistical significance (one-sided test) are marked as: *** for 0.01 level, ** for 0.05 level, and * for 0.10 level. These indicators, however, are less meaningful for categorical variables. What matters is the significance of differences between the coefficients, and not differences of the coefficients from zero. Missing cells represent a lack of loan records with variable values in these categories. The fixed-rate-purchase equation was estimated on a one-in-three sample of loan groups, where each started with at least 10 loans. Each sampled loan group provides quarterly time series of observations, through 2000Q4, the end of year twenty of loan life, or until all loans terminate, whichever comes first. The number of actual loan quarters represented by the loan-group observations listed in the Summary Statistics is typically very large. For example, for the fixed-rate-purchase equation, the number of loan quarters in the regression estimation is over 65 million, so that, on average, each loan-group provides 33 quarters of observations.

Table 11. Default Related Cash-Flow Timing and Regression Models

Significant Points in Time	Events	Use of Regression Models in Simulations	Cash Flow Items
Mortgage payment due but unpaid	Starts delinquency period.		
90-day delinquency (default)	Three missed payments and a fourth due-and-payable defines default, the point at which foreclosure of property rights becomes a viable option.		
Fourth month of delinquency	Determination of eligibility of borrower for workout option.		Administrative costs paid by FHA to loan servicers for workout agreements with borrowers.
Foreclosure of borrower property rights and transfer of property title to HUD/FHA	Path for all defaulted loans without workout agreements in place by sixth month.	Time from default to foreclosure.	Claim expense paid by FHA for loan balance and foreclosure expenses.
Property Disposition/Sale	Sale of property	Time-to-Disposition and Recovery Rate on Sale	Net proceeds on sale, as percent of defaulting loan balance; property holding expenses.

Table 12. Logistic Regression of the Probability of Receiving Loan-Workout Offers

Variable Names	class	Category Values	Estimated Coefficients	Standard Errors	Descriptions
Intercept			-1.0854***	0.0732	Regression constant term
LTV_CLASS	1	up to 80%	0.3885***	0.0258	Origination LTV
	2	81-90%	0.0169	0.0163	
	3	91-95%	0.0729***	0.0163	
	4	96-100%	-0.1708***	0.0092	
	5	over 100%	*-0.3075****		
PROD_CLASS	1	FRM30	0.2161***	0.0269	Fixed-rate 30-year
	2	FRM15	0.6909***	0.0345	Fixed-rate 15-year
	3	ARM	-0.0722*	0.0276	Adjustable rate
	4	GPM	-0.7964***	0.0615	Graduated payment
	5	GEM	*-0.0384*		Graduated equity
PRICE_CLASS	1	under 50%	-0.3224***	0.0064	Purchase price as a percent of area (MSA) median house price
	2	51-100%	0.0074	0.0046	
	3	over 100%	*0.3150****		
CYCLE_STAGE	3	early expansion	-0.8387***	0.0602	See Table A.5 for description
	4	out of cycle	*0.8387****		
HPI_CLASS	1	under .90	-2.2037***	0.1370	Value of HPI at time of default, where value is 1.00 at loan origination.
	2	.90 - .99	-1.3202***	0.0476	
	3	1.0 - 1.09	0.1619***	0.0301	
	4	1.10 - 1.24	0.5768***	0.0292	
	5	1.25 - 1.49	1.2583***	0.0293	
	6	1.5 or over	*1.5269****		
HPI_GROWTH_CLASS	1	under 0	0.0109	0.0097	4-quarter HPI growth leading, prior to quarter of loan default.
	2	0 - .04	-0.3358***	0.0071	
	3	.05 - .09	0.0526***	0.0056	
	4	.10 and over	*0.2723****		
MORT_AGE_Q			-0.0178***	0.0004	Mortgage age at default, in quarters

Summary Statistics	
Observations	405,539 (207,262 are reported in the FHA loss mitigation data base and another 198,277 are lender forbearances reported in the FHA default monitoring system)
Pseudo R-squared	.085
Likelihood ratio Chi-squared	26,688 (Degrees of Freedom=20)

Notes: Coefficients without standard errors are imputed values from estimated class effects and appear in italics. Only *CYCLE_STAGE* = {3,4} in the regression data sample. In the forecast simulations, the effect for *CYCLE_STAGE* = 3 is also used for *CYCLE_STAGE* = 2, and the effect for *CYCLE_STAGE* = 4 is also used for *CYCLE_STAGE* = 1. Standard indicators of statistical significance (one-sided test) are marked as: *** for 0.01 level, ** for 0.05 level, and * for 0.10 level. These indicators, however, are less meaningful for categorical variables. What matters is the significance of differences between the coefficients, and not differences of the coefficients from zero. Probability estimates (P) are computed from variable values (X) and estimated effect coefficients (β) as: $P = \frac{e^{X\beta}}{1+e^{X\beta}}$.

Table 13. Foreclosure-Time Regression: Accelerated Failure Time with Weibull Error Distribution

Variable Names	Category Values	Estimated Coefficients	Standard Errors	Proportionality Factors	Descriptions
CONSTANT	1	3.7011	0.0109		Regression intercept term.
CYCLE_STAGE	1	-0.1182	0.0032	0.89	1st half of price decline
	2	0.1126	0.0024	1.12	2nd half of price decline
	3	0.2421	0.0030	1.27	Recovery from stage 2
HPI_GROWTH		-4.9764	0.0113	0.01	House price growth in first 4 quarters of default (decimal)
CUR_MKT_RATE		-0.1860	0.0008	0.83	Current mortgage rates (%)
CENSUS_DIV	NEW	0.0926	0.0044	1.10	New England
	MAT	0.2773	0.0024	1.32	Middle Atlantic
	SAT	-0.0058	0.0016	0.99	South Atlantic
	ENC	0.0957	0.0020	1.10	East North Central
	WNC	-0.0380	0.0022	0.96	West North Central
	ESC	-0.1367	0.0024	0.87	East South Central
	WSC	-0.2081	0.0016	0.81	West South Central
	MTN	-0.1507	0.0018	0.86	Mountain
DEFAULT_YR	1975	4.8376	338.80	126.16	Year of default
	1976	4.5314	338.80	92.89	
	1977	5.8693	338.80	354.00	
	1978	2.2857	0.4377	9.83	
	1979	6.9335	46.86	1026.06	
	1980	2.7816	0.2190	16.14	
	1981	3.0678	0.0726	21.49	
	1982	2.2192	0.0164	9.20	
	1983	1.6955	0.0111	5.45	
	1984	1.2623	0.0100	3.53	
	1985	0.9656	0.0095	2.63	
	1986	1.0310	0.0095	2.80	
	1987	0.9176	0.0095	2.50	
	1988	0.8383	0.0094	2.31	
	1989	0.7965	0.0094	2.22	
	1990	0.6979	0.0093	2.01	
	1991	0.5987	0.0092	1.82	
	1992	0.5891	0.0092	1.80	
	1993	0.6414	0.0092	1.90	
	1994	0.6136	0.0093	1.85	
	1995	0.6427	0.0092	1.90	
	1996	0.5054	0.0092	1.66	
	1997	0.4383	0.0092	1.55	
	1998	0.4721	0.0092	1.60	
	1999	0.4951	0.0092	1.64	
	2000	0.3910	0.0093	1.48	
WEIBULL Shape parameter		0.4376	0.0004		Defines shape of error distribution
Summary Statistics					
Observations		874790			
Right-censored values		24395			(reported foreclosure times are over 36 months)
Mean time to foreclosure		12.76 months			

(Continued)

Table 13. Foreclosure-Time Regression: Accelerated Failure Time with Weibull Error Distribution (continued)

Notes: Left-hand-side (dependent) variable is months between default and foreclosure completion. Coefficient values for *CYCLE_STAGE*=4 (no cycle), *CENSUS_DIV*=PACIFIC, and *DEFAULT_YR*=2001 are jointly included in the constant term of the regression. Proportionality factors are exp{β}, where β is the regression effect coefficient. Predicted values of foreclosure time (T_f) are computed as: $T_f = (e^{X\beta} \cdot (-\ln(U))^c$, where X represents a vector of zeros and ones that indicate which effect coefficients are on(1) or off (0) for a given observation, and β is the vector of estimated coefficients, U is a uniform random number, and c is the Weibull shape parameter. The forecast simulations in this report use the mean of the Weibull distribution, rather than randomly generated values, so that:

$$T_f = \left(e^{X\boldsymbol{\beta}}\right) \cdot \Gamma\left(\frac{1/c + 1}{1/c}\right)$$, where Γ represents the Gamma function.

Table 14. Foreclosed-Property Disposition (Sales) Time Regression: Accelerated Failure Time with Weibull Error Distribution

Variable Names	Category Values	Coefficient Estimates	Standard Errors	Proportion-ality factors	Descriptions
CONSTANT		1.0825	0.0241		Regression intercept term
CYCLE_STAGE	1	0.1104	0.0102	1.117	Observation in 1st half of HPI decline
	2	0.0945	0.0073	1.099	Observation in 2nd half of HPI decline
	3	-0.0210	0.0062	0.979	Observation in recovery from stage number 2
HPI_GROWTH_CLASS	under 0	0.1764	0.0065	1.193	House price growth in first 4 qtrs of default
	0.0 - .04	0.1351	0.0062	1.145	
	.05 - .09	0.1269	0.0058	1.135	
CUR_MKT_RATE		0.0301	0.0027	1.031	Current mortgage rates (fixed-rate, 30-year, %)
FORECLOSE_YR	1981	0.8949	0.7651	2.447	Year of property foreclosure.
	1983	1.7158	0.2893	5.561	(No observations in 1982)
	1984	2.8698	0.1535	17.633	
	1985	2.4747	0.0452	11.878	
	1986	2.4506	0.0331	11.595	
	1987	2.3847	0.0215	10.856	
	1988	2.0125	0.0136	7.482	
	1989	1.7313	0.0100	5.648	
	1990	1.0122	0.0076	2.752	
	1991	0.8884	0.0072	2.431	
	1992	0.8296	0.0064	2.292	
	1993	0.7641	0.0063	2.147	
	1994	0.5880	0.0064	1.800	
	1995	0.4817	0.0066	1.619	
	1996	0.3966	0.0066	1.487	
	1997	0.4048	0.0063	1.499	
	1998	0.5000	0.0065	1.649	
	1999	0.3758	0.0069	1.456	
	2000	0.3191	0.0061	1.376	
CENSUS_DIV	NEW	0.2117	0.0082	1.236	New England
	MAT	-0.0192	0.0048	0.981	Middle Atlantic
	SAT	-0.1256	0.0034	0.882	South Atlantic
	ENC	-0.0587	0.0044	0.943	East North Central
	WNC	-0.1968	0.0048	0.821	West North Central
	ESC	-0.2093	0.005	0.811	East South Central
	WSC	-0.1565	0.0036	0.855	West South Central
	MTN	-0.1850	0.0042	0.831	Mountain
Weibull scale parameter		0.7650	0.0007		Defines shape of the error distribution

Summary Statistics	
Observations	610,704
Right-censored values	20,991 (reported disposition times are over 24 months)
Mean sale time	5.7 months

(Continued)

Table 14. Foreclosed-Property Disposition (Sales) Time Regression: Accelerated Failure Time with Weibull Error Distribution
(continued)

Note: Left-hand-side (dependent) variable is months between foreclosure completion and property sale. Combined effects for *CYCLE_STAGE*=4, *HPI_GROWTH* $\geq$ 0.10, *FORECLOSURE_YR*=2001, and CENSUS_DIV=Pacific are jointly included in the constant term of the regression. Proportionality factors are exp{β}, where β is the regression effect coefficient. Predicted values of foreclosure time (T_f) are computed as: $T_f = \left(e^{X\beta}\right)\left(-\ln U\right)^{c}$, where X represents a vector of zeros and ones that indicate which effect coefficients are on(1) or off (0) for a given observation, and β is the vector of estimated coefficients, U is a uniform random number, and c is the Weibull shape parameter. The forecast simulations in this report use the mean of the Weibull distribution, rather than randomly generated values, so that: $T_f = \left(e^{X\boldsymbol{\beta}}\right)\cdot\Gamma\left(\frac{1/c + 1}{1/c}\right)$, where Γ represents the Gamma function.

Table 15. Foreclosed Property Recovery Rate Regression

Variables	Category Values	Coefficient Estimates	Standard Errors	Descriptions
CONSTANT		0.7922	0.0025	Regression intercept term
PRICE_DROP_SCORE		0.0713	0.0003	Standard normal statistic measuring distance between price decline needed to make default in-the-money and the median market price growth found in the HPI.
PRICE_CLASS	< 50%	-0.2296	0.0011	Based on ratio of house price to area
	51-100%	-0.1010	0.0008	(MSA) median house price.
CENSUS_DIV	NEW	-0.2097	0.0026	New England
	MAT	-0.1624	0.0014	Middle Atlantic
	SAT	0.0019	0.0010	South Atlantic
	ENC	-0.0698	0.0013	East North Central
	WNC	-0.0778	0.0015	West North Central
	ESC	-0.0449	0.0015	East South Central
	WSC	-0.0367	0.0011	West South Central
	MTN	0.0514	0.0013	Mountain
INVESTOR	NO	0.1133	0.0020	Investor owned property flag
FORECLOSURE_YR	1985	0.0022	0.0119	Calendar year of foreclosure completion
	1986	-0.0565	0.0129	
	1987	-0.0673	0.0095	
	1988	-0.0924	0.0049	
	1989	-0.0506	0.0033	
	1990	-0.1195	0.0018	
	1990	0.0022	0.0024	
	1991	-0.1321	0.0016	
	1992	-0.1173	0.0015	
	1993	-0.0845	0.0015	
	1994	-0.0629	0.0015	
	1995	-0.0455	0.0016	
	1996	-0.0414	0.0016	
	1997	-0.0500	0.0016	
	1998	-0.0560	0.0015	
	1999	-0.0398	0.0015	
	2000	-0.0205	0.0015	
Summary Statistics				
Observations	610,322			
R-Squared	.1878			
Mean of recovery rate	.7918			

Note: Left-hand-side (dependent) variable is the ratio of the sales price less sales expenses to the unpaid loan balance at time of default. Combined effects for *PRICE_CLASS*=3 (over 100%), *CENSUS_DIV* = Pacific Division, *INVESTOR*=YES, and *FORECLOSURE_YR*=2001 are jointly included in the regression constant. This is an ordinary least squares regression, where predicted recovery rates are a linear function of the explanatory variables and estimated coefficients.

Table 16. Miscellaneous Parameters Used to Calculate Insurance-Claim Cost in Simulations

Parameters	Values	Used in/when	Descriptions
Workout Failure Rate	0.20	Computing percent of default workouts. Applied to number of loans receiving workout assistance.	Available data are not yet sufficient for econometric analysis of workout success/failure rates. The rate used here is consistent with recent FHA experience.
Foreclosure time	upper bound of 10 quarters	Timing of claim expenses and ultimate recovery on REO sales	This limit is placed on results of the foreclosure time regression, simply to assure that no out-of-bounds results occur.
Interest expense in claim	75% of current mortgage note rate, from quarter of last-paid-installment to foreclosure quarter	Insurance claim payment on foreclosure	FHA reimburses loan servicers for lost interest on defaulted mortgages, up to the date of foreclosure. This reimbursement is at a note rate equivalent to the average 30-year Treasury yield in the year of loan origination. CBO approximates the mortgage-to-Treasury spread rather than maintaining a lookup table of Treasury rates in the simulation model.
Foreclosure-related expenses	7.7% of unpaid principal balance	Insurance claim payment on foreclosure	This represents the cost to FHA of attorney fees, property taxes due, and other costs of securing title to the property, net of any escrow funds available. FHA generally reimburses servicers for two-thirds of this net expense. The 7.7% rate used here is the average expense rate booked by FHA, 1996-2001.
Preforeclosure Sale Benefits	8% on 5% of foreclosures	Insurance claim payment on foreclosure	FHA maintains a little-used program of assisting certain defaulted borrowers to sell their properties prior to foreclosure. With little data on this program, CBO uses the cost savings gleaned by HUD in its 1994 pilot study evaluation (Charles A. Capone, Jr., *Evaluation of the Federal Housing Administration Preforeclosure Sale Demonstration.* Washington, DC: U.S. Department of HUD, Office of Policy Development and Research, June 1994.) This cost saving (8% of the unpaid principal balance) is applied to 5% of foreclosures in the simulations.

(continued)

Table 16. Miscellaneous Parameters Used to Calculate Insurance-Claim Cost in Simulations (Continued)

Parameters	Values	Used in/when	Descriptions
Cost of loan workouts	1.5% of outstanding loan balance	Credit cost expense booked in default quarter	FHA supports loan servicer workout initiatives with incentive payments. These are generally a few hundred dollars, averaging just 0.5% of the outstanding loan balance (as of January, 2002). However, when FHA pays a partial claim to make the borrower whole (and placing a lien against the property), the cost has been, on average, between 7 and 8% of the loan balance, over time. By applying the incentive payment rate (0.5%) to all workouts, and partial claim cost (8%) to 10 percent of the cases, one arrives at a net charge for each loan workout of 1.3%. CBO rounds this up to 1.5% for use in the simulations.
REO property sale time	upper bound of 8 quarters	time between foreclosure and property sale.	Limitation placed on results of regression equation to prevent the possibility of out-of-bound results.
REO sale recovery rate	constrained to the (.30, 1.30) interval	revenue inflow from property sales, net of sales expenses	Regression results are bounded to remove out-of-bound events. HUD sales of some homes for $1 to nonprofits and local governments is accounted for in the data used to estimate the sale-time regression equation. Thus, the regression equation captures average results for loans with given characteristics.
Other property management expenses	6.25%	adjusts property sale proceeds downward	Expense ratio–against unpaid principal balance–for property management, repairs, and taxes during the property holding period. The 6.45% rate is the average for properties sold in 1999 and represents an average experience of the most recent 5 year period.
Premium Refund Schedule	rates vary by loan origination date and age of mortgage.	Rebates to borrowers who payoff mortgages early.	The rebate schedule corresponds with FHA's determination of how quickly the up-front premium is actually earned. For loans insured prior to 1994, the premium is considered earned over the life of the loan. For 30-year loans, the rebate declines to 50 percent (of the up-front premium) by the end of the fifth year of loan life, and 10 percent by the end of the seventeenth year. For loans insured from January 1994 through December 2000, premiums are earned over just seven years, with reimbursement rate falling to 50 percent after 40 months. Loans insured since January 2001 have a five-year earning schedule, with rebate percentages in years 1-5 being: 85, 65, 45, 25, and 10, respectively.

Table 17. Confidence Bounds for Subsidy Rate Simulation Results, by Cohort Year (percent)

Cohort Year	Percentiles				
	1% lower bound	5% lower bound	mean	5% upper bound	1% upper bound
1992	-1.15	-1.14	-1.12	-1.10	-1.09
1993	-2.18	-2.16	-2.10	-2.06	-2.04
1994	-2.12	-2.09	-2.03	-1.96	-1.94
1995	-0.25	-0.21	-0.14	-0.07	-0.04
1996	-0.74	-0.69	-0.57	-0.47	-0.43
1997	-0.70	-0.64	-0.51	-0.39	-0.33
1998	-1.80	-1.70	-1.44	-1.20	-1.10
1999	-2.57	-2.41	-1.97	-1.53	-1.32
2000	-1.50	-1.34	-1.02	-0.70	-0.52
2001	-2.04	-1.79	-1.19	-0.55	-0.14
2002	-3.17	-2.81	-1.73	-0.26	0.77
2003	-3.37	-2.97	-1.85	0.10	2.07
2004	-3.32	-2.88	-1.53	0.86	3.36
2005	-3.06	-2.66	-1.04	1.90	4.31
2006	-3.04	-2.53	-0.64	3.08	6.16
2007	-3.10	-2.57	-0.32	4.05	6.92