# Mortgage default and low downpayment loans: The costs of public subsidy

Yongheng Deng[a,*], John M. Quigley[b], Robert Van Order[c], Freddie Mac[c]

[a]Office of Federal Housing Enterprise Oversight, 1700 G Street, NW, Washington, DC 20552, USA
[b]University of California, Berkeley, CA 94720-7320, USA
[c]Federal Home Loan Mortgage Corporation, 8200 Jones Branch Drive, McLean, VA 22102, USA

## Abstract

This paper presents a unified model of the default and prepayment behavior of homeowners in a proportional hazard framework. The model uses the option-based approach to analyze default and prepayment, and considers these two interdependent hazards as competing risks. The results indicate the sensitivity of default to the initial loan-to-value ratio of the loan and the course of housing equity. The latter is a measure of the extent to which the default option is in the money. The results also indicate the importance of trigger events, namely unemployment and divorce, in affecting prepayment and default behavior. The empirical results are used to analyze the costs of a current policy proposal–stimulating homeownership by offering low downpayment loans. We simulate default probabilities and costs on zero-downpayment loans and compare them with conventional loans with conventional underwriting standards. The results indicate that if zero-downpayment loans were priced as if they were mortgages with 10% downpayments, then the additional program costs would be 2–4% of funds made available–when housing prices increase steadily. If housing prices remained constant, the costs of the program would be much larger indeed. Our estimates suggest that additional program costs could be between $74,000 and $87,000 per million dollars of lending. If the expected losses from such a program were not priced at all, the losses from default alone could exceed 10% of the funds made available for loans.

Keywords: Mortgage default; Low downpayment; Public subsidy; Proportional hazard model

JEL classification: G14; R20; H21

* Corresponding author. Fax: (202) 414-3826; e-mail: deng@econ.berkeley.edu.

## 1. Introduction

This paper presents a unified model of the default and prepayment behavior of homeowners in a context which recognizes these hazards as competing risks. The model uses the option-based approach to analyzing default and prepayment, i.e. it views default as a put option and prepayment as a call option. The analytical model also recognizes that the borrower, by exercising one of these two options, gives up the right to exercise the other option.

The model is applied to analyze a current policy proposal–transferring resources and stimulating homeownership by offering low downpayment loans. For instance, it has been proposed recently that the federal government, through FHA, insure zero downpayment mortgage loans for low income families.[1] We do not estimate the effectiveness of this policy at stimulating homeownership, but we do use our model to estimate program costs. We do this by comparing default probabilities and costs on zero downpayment loans with conventional loans requiring a 5% or 10% downpayment. Obviously the subsidy inherent in zero downpayment loans compared with conventional loans depends on the pricing of these loans. We can estimate subsidy costs directly if the difference in risk is not priced. The FHA has, historically, not varied price significantly by downpayment.[2]

The starting point for option-based models is the contingent claims model, developed by Black and Scholes (1972) and Cox et al. (1985). A number of studies have applied this model to the mortgage market (see, for example, Dunn and McConnell, 1981; Buser and Hendershott, 1984; Brennan and Schwartz, 1985; Kau et al., 1992, 1995; Quigley and Van Order, 1995). Hendershott and Van Order (1987) and Kau and Keenan (1995) provide surveys of these models and results.

The key result of these models is that the exercise of *each* option depends on *both* house prices and interest rates. The default option is in the money when a borrower's equity is negative, and the prepayment option is in the money when the present value of the remaining payments is less when discounted at the current interest rate rather than the mortgage coupon rate. The contingent claims model describes how far into the money these options must be in order for it to be optimal for the borrower to exercise one of

---

[1] Under the Clinton administration's fall 1994 initiative, mortgages would be targeted to first-time homebuyers with incomes below 115% of the median for the metropolitan area of residence. Mortgage amounts would be limited to the lesser of 75% of the FHA 203(b) loan limit for the location or 100% of property value.

[2] Prior to 1991, the FHA did not vary price by initial loan-to-value ratio (LTV). Currently all borrowers pay an up-front premium of 2.25%. They also pay 0.5% of the loan balance each year for seven years (in the case of LTV <90%) to 30 years (in the case of LTV >95%). As indicated by our estimates below, the current variation does not nearly match the variation in risk.

them. If there are any transactions costs, the mathematics of all this gets quite complicated. Empirical work exploiting the option perspective has generally taken a probabilistic approach: each option is more likely to be exercised the further it is 'in the money'.

Several recent empirical studies have applied the Cox proportional hazard model (Cox and Oakes, 1984) to evaluate mortgage default or prepayment risk (see, for example, Green and Shoven, 1986; Schwartz and Torous, 1989; Quigley and Van Order, 1990, 1995). Instead of solving for the unique critical values of the state variables in the contingent claims model, the proportional hazard model assumes that at each point in time during the mortgage contract period the mortgage has a certain probability of termination, conditional upon the survival of the mortgage. The hazard function in this model is defined as the product of a baseline hazard and a function of time-varying covariates. These covariates need not be limited to the option value itself; they may include other important determinants of behavior. The proportional hazard model can thus incorporate reasonable mortgage prepayment and default behavior that would be considered 'suboptimal' under the pure contingent claims framework.

However, few of the existing empirical models have treated the inter-dependence of borrowers' prepayment and default options. For example, Schwartz and Torous (1989) analyzed Government National Mortgage Association (GNMA) mortgage prepayment experience by using a model with fixed covariates[3] and assuming the mortgages were free of default risk. More general models using time-varying covariates; for example, Green and Shoven (1986) and Quigley and Van Order (1990) made analogous assumptions in the analysis of prepayment behavior. Cunningham and Hendershott (1984) analyzed default costs in an optimizing context with transactions costs and a single hazard. Foster and Van Order (1984) and Quigley and Van Order (1995) studied default behavior using the model of a single hazard as well. Foster and Van Order (1985) did estimate both default and prepayment, but they used highly aggregated data on mortgage pools. Cunningham and Capone (1990) estimated mortgage default and prepayment using a limited sample of regional data in a multinomial logit framework.

## 2. The model

In this paper we model mortgage termination in a competing risks framework. The competing risks of mortgage termination consist of two parts: a prepayment risk and a default risk. The function specifying

---

[3] The option-related financial and economic variables relevant to mortgages clearly vary over time, so it would have been more appropriate to specify the model with time-varying covariates.

prepayment risk estimates the probability that a mortgage loan is prepaid during any period, conditional upon survival to that particular period. Similarly, the default function estimates the conditional probability of default during each period. The model assumes that the borrower makes the prepayment or default decision on the basis of market conditions to maximize net wealth. Following the contingent claims model, we assume that the probability of exercising these options is a function of the extent to which they are 'in the money' and of 'trigger events' that affect the decision about how far the option needs be into the money in order for exercise to be optimal. For instance, an increase in the probability of negative equity will increase the probability that the put option is in the money, and hence increase the probability of default. Examples of trigger events include such economic variables as employment or divorce.[4]

One particular feature of mortgage prepayment and default is the possibility of right censoring. There are two sources of censoring. First, some mortgage loans may simply mature under the contract, and some mortgages may not terminate by the end of the data collection period. For those mortgage loans, we simply never observe their actual durations to default or prepayment. Second, and more importantly, if a mortgage loan has defaulted, it cannot be prepaid in the future. Thus defaulted loans are treated as censored data for the prepayment function, and prepaid loans are treated as censored data for the default functions.

Both of these mechanisms are random censoring. It is reasonable to assume that the former random censoring mechanism (e.g. the maturation of loans) is independent of the default or prepayment failure time series. However, the latter mechanism need not be an independent random censoring mechanism at all. Indeed, for this reason we should expect there to be a correlation between defaults and prepayments.[5]

## 2.1. A semiparametric estimation approach to the proportional hazard model

A major concern in actually estimating hazard models of mortgage prepayment and default behavior is the computational difficulty involved. Useful models must be specified in short time intervals (e.g. months or quarters), but mortgage terms are typically written for 30 years. Computational time can become a real constraint when the model involves time-varying covariates. To estimate a useful model for the housing market

---

[4] Job loss or household dissolution leads to reduced ability to fulfill monthly payment obligations and thus increases the likelihood of mortgage termination by default.

[5] For example, households with negative equity may be more likely to default, but negative equity also makes it less likely that a household will choose to prepay.

requires either dramatically limiting sample sizes, arbitrarily and unreasonably aggregating time intervals, or else finding a way to aggregate observations on individual behavior.[6]

In this paper we use a semiparametric estimation approach (SPE) to estimate the proportional hazard model with competing risks and time-varying covariates. This approach is described in detail elsewhere (Deng et al., 1995); here we merely sketch out the major points.

Let us define $T \in R^+$ as a duration variable. Let $T_i$ $(i = 1, 2, ..., q)$ be the discrete time intervals that partition the support of $T$. Let

$$h_j(t, Z) = h_{0j}(t)[\exp(Z_j(t)\beta_j)]\eta_j \, , \quad j = 1, 2 \, , \tag{2.1}$$

be the conditional probability of a mortgage terminating at $t$. Here $j = 1$ is the prepayment function and $j = 2$ is the default function.

A log integrated hazard function for risk type $j$ can be constructed as

$$\log \left[ \int_{T_{i-1}}^{T_i} h_j(t, Z_j) \mathrm{d}t \right] = Z_j(T_i)\beta_j + \gamma_j(T_i) + \varepsilon_j \, , \tag{2.2}$$

where

$$\gamma_j(T_i) = \log \left[ \int_{T_{i-1}}^{T_i} h_{0j}(t) \mathrm{d}t \right] , \tag{2.3}$$

and

$$\varepsilon_j = \log \eta_j \, , \quad j = 1, 2, ..., J \, , \quad i = 1, 2, ..., q,$$

given that $Z_j(t)$ is constant between $T_{i-1}$ and $T_i$.

The left-hand side of Eq. (2.2) is not directly observable in micro data. We can, however, use the 'local smoothing' technique, developed in the literature on non-parametric methods, to estimate individual hazard functions based on the empirical distribution of the hazard functions. We partition the covariate matrix $Z$ into $K$ distinct matrices $Z_1, ..., Z_K$. The $k$th subgroup contains $M_k$ observations. $M_1 + M_2 + ... + M_K = N$, where $N$ is the total sample size. For each subgroup, we estimate the hazard rate such that $\hat{h}_{jkt} = n_{jkt}/S_{kt}$, where $n_{jkt}$ is the number of loans terminated in the $t$th

---

[6] See Deng et al. (1995) for a discussion of alternative models and computational methods.

period with type $j$ in the $k$th subgroup, and $S_{kt}$ is the total number of loans surviving to the $t$th period in the $k$th subgroup.[7]

Now, if we replace the left-hand side of Eq. (2.2) with the smoothed log hazard function,

$$\log \int_{T_{i-1}}^{T_i} \hat{h}_{jk}(t, Z_{jk})\mathrm{d}t,$$

then we obtain

$$\log \left[ \int_{T_{i-1}}^{T_i} \hat{h}_{jk}(t, Z_{jk})\mathrm{d}t \right] = Z_{jk}(T_i)\beta_j + \gamma_j(T_i) + \varepsilon_j + u_{jk}(T_i), \qquad (2.4)$$

$$j = 1, 2, \ldots, J, \quad k = 1, 2, \ldots, K, \quad i = 1, 2, \ldots, q,$$

where

$$u_{jk}(T_i) = \log \left[ \int_{T_{i-1}}^{T_i} \hat{h}_{jk}(t, Z_{jk})\mathrm{d}t \right] - \log \left[ \int_{T_{i-1}}^{T_i} h_{jk}(t, Z_{jk})\mathrm{d}t \right].$$

For a model in which time is measured in discrete intervals, say quarters, Eq. (2.4) reduces to

$$\log \left[ \hat{h}_{jk}(T_i, Z_{jk}) \right] = Z_{jk}(T_i)\beta_j + \gamma_j(T_i) + \varepsilon_j + u_{jk}(T_i). \qquad (2.5)$$

The covariance of the $\varepsilon_j$'s captures the correlation among competing risks. Eqs. (2.4) and (2.5) are seemingly unrelated regression systems that can be analyzed using the approach proposed by Zellner (1962). It has been shown elsewhere (Deng et al., 1995) that the coefficient vector $\hat{\beta}$ estimated from Eq. (2.4) is consistent.

## 3. The empirical analysis

### 3.1. The data

The empirical analysis is based on individual mortgage history data maintained by the Federal Home Loan Mortgage Corporation (Freddie Mac). The database contains 1,489,372 observations on single family

---

[7] Note the risk set of conditional hazard rates includes not only the loans that have the same termination type, but also all those loans that have a different termination type, as long as the age of termination is greater than the current one. Furthermore, the risk set also includes those right-censored observations if the censored time point is greater than the current termination time.

mortgage loans issued between 1976 to 1983 and purchased by Freddie Mac. All are fixed-rate, level-payment, fully amortized loans, most of them with 30-year terms. The mortgage history period ends in the first quarter of 1992. For each mortgage loan, the available information includes the year and month of origination and termination (if it has been closed), indicators of prepayment or default, the purchase price of the property, the original loan amount, the initial loan-to-value ratio (LTV), the mortgage contract interest rate, the monthly principal and interest payment, the state, the region and the major metropolitan area in which the property is located. The data set also reports the borrower's monthly gross income at loan origination. For the mortgage default and prepayment model, censored observations include all matured loans as well as the loans active at the end of the period.

The analysis is confined to 30-year fixed-rate mortgage loans issued for owner occupancy, and includes only those loans that were either closed or still active at the first quarter of 1992.[8] The analysis is confined to loans issued in 26 major metropolitan areas (MSAs). The data set contains 780,443 observations. Loans are observed in each quarter from the quarter of origination through the quarter of termination, maturation, or through 1992:I for loans still active.

To estimate the model, the entire sample of 780,443 loans was partitioned into 312 groups, according to 26 major MSAs, four household income levels, and three LTV groups.[9] For each group there are 64 cells, reflecting failure time periods (measured in quarters, from 1976:II to 1992:I). We use the Kaplan–Meier approach to fit the empirical hazard rates of prepayment and default based on the entire sample. Then the estimated empirical hazard rates were mapped to 11,866 mortgage loans that were randomly drawn from the total sample.[10] We assume that the randomly drawn subsample has the same distribution as the population.

The appendix provides some descriptive statistics of the subsample of 11,866 observations used in the regression. A majority of the mortgage loans have an equity ratio greater than 0.3; only 1.6% of loans have an equity ratio less than 0.1. About 2.2% of households had incomes below 60% of the MSA median at loan origination, and 43% of households had

---

[8] It excludes those observations that were in delinquency or foreclosure at the time data were collected.

[9] The three LTV groups are: LTV less than or equal to 0.8, LTV greater than 0.8 and less than 0.95, and LTV greater than or equal to 0.95.

[10] We confine the random drawing within the subsample of those cells that have non-zero empirical hazard rates. It should be pointed out that if the cell partition is too small, then it might increase the occurrence of zero empirical hazard rates, especially for default hazard rates, for certain cells. Consequently, the drawing will over-sample default loans by eliminating those loans assigned zero empirical default hazard rates. (To obtain an optimal partition size, we could have used the cross-validation technique, which is familiar in the literature on non-parametric estimation.)

incomes above 150% of the MSA median. About 12% of the loans were issued with an LTV below 0.6, and 15% of the loans had an initial LTV at or above 0.95.

Figs. 1 and 2 summarize the raw data on mortgage terminations which underlie the empirical analysis. Fig. 1 displays the conditional prepayment rate, separately by LTV, as a function of duration. Conditional prepayment rates are slightly higher for higher LTV loans. Rates increase substantially after the first 15 quarters. Fig. 2 displays raw conditional default rates by LTV. Note again that default rates increase substantially after about 15 quarters. Note also that the default rates increase very substantially with initial LTV. Default rates for loans with LTV above 95% are three or four times higher than default rates for 90–95% LTV loans. The default rates for these latter loans are, in turn, about five times as high as for those with an LTV below 80%.

Finally, note that conditional default rates are quite low. Even for the riskiest class of loans, conditional default rates are no higher than four in a thousand per quarter. Residential mortgages are relatively safe investments (and simple random samples of mortgages are likely to contain very few observations on default).

The key variables in the theoretical model are those that measure the extent to which the put and call options are in the money. To value the call option, the current interest rate, and the initial contract terms, are sufficient. We compute a variable *POPTION* that measures the ratio of the present discounted value of the unpaid mortgage balance at the current
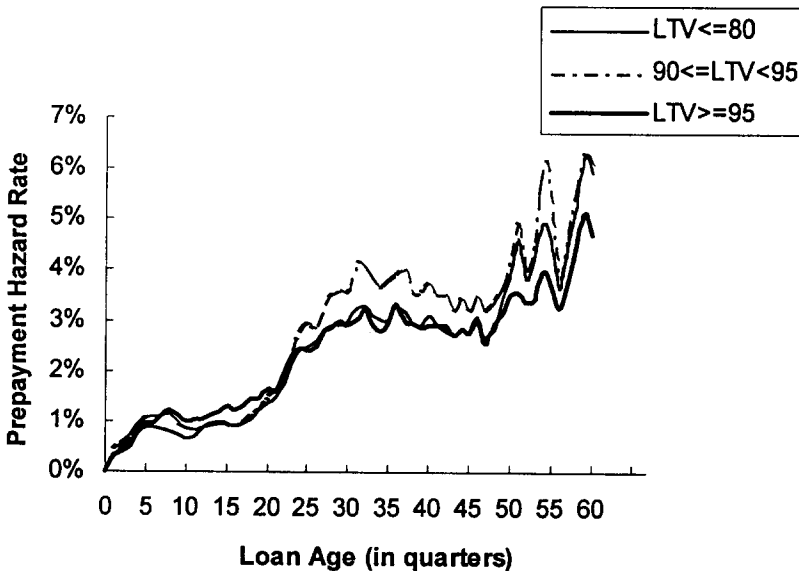


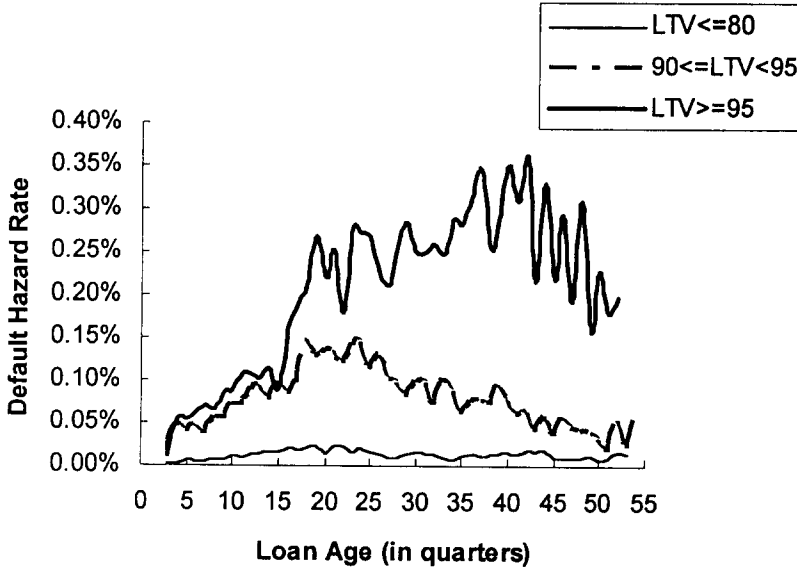Fig. 1. Conditional prepayment rates by LTV.

Fig. 2. Conditional default rates by LTV.

quarterly market interest rate[11] relative to the value discounted at the contract interest rate.[12]

To value the put option analogously, we need to measure the market value of each house quarterly and to compute homeowner equity quarterly. Obviously, we do not observe the course of price variation for individual

[11] The rate used is the average interest rate charged by lenders on new first mortgages reported by Freddie Mac's quarterly market survey (the 'effective Freddie rate'). This mortgage interest rate varies by quarter across five major U.S. regions.

[12] Specifically, *POPTION* for the *l*th loan observation is defined as

$$POPTION_l = \frac{\displaystyle\sum_{t=1}^{term_l - \tau_i} \frac{mopipmt_l \times 3}{(1 - mktrate_{\omega_l,\kappa_l+\tau_i}/400)^t} - \sum_{t=1}^{term_l - \tau_i} \frac{mopipmt_l \times 3}{(1 + noterate_l/400)^t}}{\displaystyle\sum_{t=1}^{term_l - \tau_i} \frac{mopipmt_l \times 3}{(1 + mktrate_{\omega_l,\kappa_l+\tau_i}/400)^t}}$$

$$= 1 - \frac{mktrate_{\lambda,\kappa_l+\tau_i} \times \left(1 - \left(\frac{1}{1 + noterate_l/400}\right)^{term_l-\tau_i}\right)}{noterate_l \times \left(1 - \left(\frac{1}{1 + mktrate_{\omega_l,\kappa_l+\tau_i}/400}\right)^{term_l-\tau_i}\right)}, \qquad (3.1)$$

where $\tau_i$ is loan age measured in quarters, $\omega_l$ is a vector of indices for geographical location, $\kappa_l$ is the loan origination time, $mopipmt_l$ is the monthly principal and interest payment, $noterate_l$ is the mortgage note rate, $mktrate_{\omega_l, \kappa_l+\tau_i}$ is the current local market interest rate, and $term_l$ is the mortgage loan term calculated by

houses in the sample. We do, however, have access to a large sample of repeat (or paired) sales of single family houses in 26 MSAs. This information is sufficient to estimate a weighted repeat sales (WRS) housing price index separately for each of the 26 MSAs. The WRS index provides estimates of the course of housing prices in each metropolitan area. It also provides an estimate of the variance in price for each house in the sample, by metropolitan area and elapsed time since purchase.[13]

$$term_t = \frac{\log \left( \dfrac{mopipmt_t}{origamt_t \times (noterate_t / 1200) + mopipmt_t} \right)}{\log (1 + noterate_t / 1200) \times 3}, \tag{3.2}$$

where $origamt_t$ is the original loan amount.

[13] Housing price indices and their volatilities are estimated according to the three-stage procedure suggested by Case and Shiller (1987) and modified by Quigley and Van Order (1995). The model assumes that the log price for the $i$th house at time $t$ is given by

$$P_{it} = I_t + H_{it} + N_{it}, \tag{3.3}$$

where $I_t$ is the logarithm of the regional housing price level, $H_{it}$ is a Gaussian random walk, such that

$$E[H_{i, t+\tau} - H_{it}] = 0,$$

$$E[H_{i, t+\tau} - H_{it}]^2 = \tau \sigma_{\eta_1}^2 + \tau^2 \sigma_{\eta_2}^2,$$

and $N_{it}$ is white noise, such that
$$E[N_{it}] = 0, \qquad E[N_{it}]^2 = (1/2)\sigma_\nu^2.$$

The model is estimated on paired sales of owner-occupied housing. In the first stage, the log price of the second sale minus the log price of the first sale is regressed on a set of dummy variables, one for each time period in the sample except the first period. The dummy variables have values of zero in every quarter except the quarter in which the sales occurred. For the quarter of the first sale, the dummy is $-1$, and for the quarter of the second sale, the dummy is $+1$. (This follows Bailey et al., 1963, exactly.)

In the second stage, the squared residuals ($e^2$) from each observation in the first stage are regressed upon $\tau$ and $\tau^2$:

$$e^2 = A + B\tau + C\tau^2, \tag{3.4}$$

where $\tau$ is the interval between the first and second sales. The coefficients $A$, $B$, and $C$ are estimates of $\tau_\nu^2$, $\sigma_{\eta_1}^2$, and $\sigma_{\eta_2}^2$, respectively.

In the third stage, the stage one regression is re-estimated by GLS with weights $\sqrt{A + B\tau + C\tau^2}$.

The estimated log price level difference $(\hat{I}_{t+\tau} - \hat{I}_t)$ is normally distributed with mean $(I_{t+\tau} - I_t)$, and variance $(\tau \sigma_{\eta_1}^2 + \tau^2 \sigma_{\eta_2}^2 + \sigma_\nu^2)$. We denote by $msa_\tau = \exp(\hat{I}_\tau)$ the estimated regional housing price index; then $\log (msa_{\kappa+\tau}/msa_\kappa)$ is normally distributed with mean $(I_{\kappa+\tau} - I_\kappa)$ and variance $(\tau \sigma_{\eta_1}^2 + \tau^2 \sigma_{\eta_2}^2 + \sigma_\nu^2)$.

Means and variances are estimated for each of 26 major MSA regions using samples of paired sales. There are about four million paired sales in the Freddie Mac database.

Estimates of the mean and variance of individual house prices, together with the unpaid mortgage balance (computed from the contract terms), permit us to estimate the distribution of homeowner equity quarterly for each observation. In particular, *EQR* is the estimate of equity ratio assuming prices of all houses in the MSA grow at the mean rate, *PNEQ* is the probability that the equity ratio is negative, i.e. the probability that the put option is in the money.[14]

For each mortgage loan observation we calculate the ratio of household reported income (at origination) to the MSA median income level. We then create four dummy variables, *INCLL*, *INCL*, *INCH*, and *INCHH* to separate household income level into four groups.[15]

As proxies for other 'trigger events', we include measures of the quarterly

---

[14] Specifically, the equity ratio for the *l*th loan observation is defined as

$$
EQR_l = \frac{mktvalue_l - pdvunpblc_l}{kmtvalue_l}
$$

$$
= \frac{purprice_l \times \dfrac{msa_{\omega_l, \kappa_l + \tau_i}}{msa_{\omega_l, \kappa_l}} - \displaystyle\sum_{t=1}^{term_l - \tau_i} \dfrac{mopipmt_l \times 3}{(1 + noterate_l / 400)^t}}{purprice_l \times \dfrac{msa_{\omega_l, \kappa_l + \tau_i}}{msa_{\omega_l, \kappa_l}}}
$$

$$
= 1 - \frac{(LTV/100) \times \left(1 - \left(\dfrac{1}{1 + noterate_l / 400}\right)^{term_l - \tau_i}\right)}{\left(\dfrac{msa_{\omega_k, \kappa_l + \tau_i}}{msa_{\omega_l, \kappa_l}}\right) \times \left(1 - \left(\dfrac{1}{1 + noterate_l / 400}\right)^{term_l}\right)}, \tag{3.5}
$$

where *purprice_l* is the purchasing price of the house at the time of loan initiation, and *pdvunpblc_l* is the present discounted value of the remaining loan balance.

The probability of negative equity, *PNEQ*, is thus

$$
PNEQ_l = ncdf\left(\frac{log\,(pdvunpblc_l) - log\,(mktvalue_l)}{\sqrt{e^2_{\omega_l, \, \kappa_l + \tau_i}}}\right), \tag{3.6}
$$

where *pdvunpblc_l* and *mktvalue_l* are defined above, ncdf( · ) is the cumulative standard normal distribution function, and $e^2_{\omega_l, \, \kappa_l + \tau_i}$ is the estimated volatility of the housing price index using the WRS procedure.

[15] Specifically, *INCLL* takes the value one if the ratio of the household reported income to the MSA median level is less than or equal to 0.6, and zero otherwise; *INCL* takes the value one if the ratio lies between 0.6 and 1.0, and zero otherwise; *INCH* takes the value one if the ratio lies between 1.0 and 1.5, and zero otherwise; and *INCHH* takes the value one if the ratio is above 1.5, and zero otherwise.

unemployment rate and the annual divorce rate.[16] Unfortunately, we do not have access to borrowers' credit history, which might be important in determining default. Hence, we assume that the model has heterogeneous error terms owing to these omitted variables.

## 3.2. The empirical results

Table 1 presents a variety of models estimated by the semiparametric estimation (SPE) method, specifying the prepayment and default functions as a seemingly unrelated regression system.

The results from all four models show that financial motivation is of paramount importance in affecting prepayment and default behavior. When the call option is in the money, the prepayment hazard increases substantially. Similarly, a higher probability of negative equity increases the default hazard substantially for all income groups. Note that a higher probability of negative equity also reduces the prepayment hazard, reflecting the negative relationship between the values of these two options.

The coefficients estimated for the interaction between household income level and the equity variables from models 3 and 4 convey consistent information. Lower income households are more sensitive to lower equity values when making decisions to exercise the put option–that is, lower income households are at greater risk for default than high-income households when equity values decline. This is true for each comparison if we exclude the highest income households. The results show that the very wealthy households (i.e. those with a ratio of household income to the MSA median income above 1.5) are apparently more likely to behave in a ruthless fashion in the face of equity declines.

The results are consistent with simple tabulations indicating that both low-income and high-income borrowers tend to default more than borrowers with moderate incomes (see, for example, Van Order and Schnare, 1994). However, the results also reveal a pattern of different responses toward declining equity among different income groups.

The results from all four models show that higher unemployment and divorce rates will lower the prepayment hazard, indicating that liquidity constraints (which make refinancing more difficult for unemployed and divorced households) keep them from exercising in-the-money call options.

---

[16] Unemployment and divorce rates are measured at the state level. State unemployment data are reported in various issues of: US Department of Labor, *Employment and Unemployment in States and Local Areas* (monthly) and in the *Monthly Labor Review*. State divorce data are reported in various issues of U.S. National Center for Health Statistics, *Vital Statistics of the United States, Vol. III, Marriage and Divorce* and in *Statistical Abstracts of the U.S.*

Table 1
Estimates of SPE for mortgage prepayment and default

| | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|
| | Prepmt | Default | Prepmt | Default | Prepmt | Default | Prepmt | Default |
| POPTION | 0.145 (7.52) | | 0.152 (7.82) | | 0.190 (10.28) | | 0.196 (10.57) | |
| PNEQ | −2.179 (49.66) | 4.55 (34.52) | −2.229 (49.54) | 3.497 (27.45) | | | | |
| INCOME | 0.065 (17.57) | −0.138 (12.67) | 0.066 (17.73) | −0.121 (11.76) | | | | |
| PNEQ × INCLL | | | | | −0.639 (2.04) | 9.154 (9.42) | −0.681 (2.18) | 8.197 (8.94) |
| PNEQ × INCL | | | | | −3.567 (54.42) | 5.692 (27.97) | −3.619 (54.50) | 4.501 (23.15) |
| PNEQ × INCH | | | | | −1.862 (30.28) | 3.388 (17.81) | −1.902 (30.66) | 2.503 (13.84) |
| PNEQ × INCHH | | | | | −0.786 (8.86) | 5.393 (19.71) | −0.844 (9.44) | 4.091 (15.72) |
| LTV | | | 0.115 (5.05) | 2.477 (38.17) | | | 0.105 (4.73) | 2.491 (38.33) |
| UNEMPLOY | −0.067 (28.90) | −0.008 (1.10) | −0.006 (28.08) | 0.036 (5.42) | −0.070 (30.92) | 0.019 (2.77) | −0.069 (30.14) | 0.046 (7.02) |
| DIVORCE | −0.126 (23.01) | −0.021 (1.30) | −0.126 (23.01) | −0.024 (1.57) | −0.116 (21.82) | −0.046 (2.77) | −0.116 (21.79) | −0.045 (2.91) |
| Var. of residual | 0.132 | 1.221 | 0.132 | 1.087 | 0.127 | 1.227 | 0.127 | 1.092 |
| Cov. of residual | 0.016 | | 0.009 | | 0.011 | | 0.006 | |
| $R^2$ | 0.633 | 0.229 | 0.634 | 0.314 | 0.647 | 0.225 | 0.648 | 0.312 |

Note: All models estimated by seemingly unrelated regressions using observations on 11,866 mortgages. t-ratios are in parentheses.

However, the coefficient estimates for these two variables are less stable in the default function.[17]

For each of the models reported in Table 1, we have also estimated a baseline hazard for prepayment and default; this estimation is discussed in detail in Quigley et al. (1994). We now apply these models and the estimated baselines to simulate the effects of low downpayment loans.

## 4. Simulations of default losses and program costs

The simulations of default losses are based on model 4 in Table 1. We use Freddie Mac's existing simulation model to generate 300 paths of mortgage market rates and 300 paths of house price inflation rates according to a joint stochastic mean-reverting process. We consider three alternative patterns of housing price change: average annual appreciation rates of 5% and 10%, and a benchmark case of 0% change, on average. We fix the divorce rate at the mean and vary the aggregate unemployment rate between 4% and 8%. For each scenario, we calculate 10,000 default hazard rates for zero downpayment loans, 5% and 10% downpayment loans for the first 15 years of the mortgage loan period. We do this by the repeated sampling of housing prices and interest rates from the joint distributions and the calculation of conditional default and prepayment probabilities for each draw.

Table 2 presents estimates of the average default rates associated with these economic conditions. The table presents the default rate, cumulated over 15 years, for zero downpayment loans and also for conventional mortgage loans covering 90% and 95%, respectively, of house purchase. Not surprisingly, the probability of ever defaulting is sensitive to variations in housing prices and also to the aggregate unemployment rate. Default probabilities are also quite sensitive to the required downpayment as well as household income level. Consider 'lower income home purchases', i.e. households with incomes below 60% of the MSA median level. The simulations suggest that, with zero down payment loans, these households would have cumulative default rates about *twice* as high as those whose mortgages require 10% down–when house prices appreciate at 10% annually and the unemployment rate is 8%. With zero downpayment loans, these households would have cumulative default rates about *four* times as high as those whose mortgages require 10% down–when house price levels are constant.

Figs. 3–6 summarize the predicted loss severities associated with these mortgage loans, as well as the timing of those losses. We use the loss

---

[17] These two variables are measured at the state level. Thus, it would be an understatement to observe that they are prone to measurement error.

Table 2
Average cumulative default in percent after 15 years

| Unemployment rate (%) | Average annual house price change (%) | Household income at initiation to MSA median income | Downpayment rate | | |
|---|---|---|---|---|---|
| | | | 10% | 5% | 0% |
| 8 | 10 | Ratio ≤ 0.6 | 4.65% | 6.25% | 9.40% |
| | | 0.6 < Ratio ≤ 1.0 | 4.34 | 5.31 | 6.71 |
| | | 1 < Ratio ≤ 1.5 | 4.08 | 4.79 | 5.70 |
| | | Ratio > 1.5 | 4.16 | 4.98 | 6.14 |
| | 5 | Ratio ≤ 0.6 | 5.80 | 9.07 | 16.18 |
| | | 0.6 < Ratio ≤ 1.0 | 4.89 | 6.35 | 8.62 |
| | | 1 < Ratio ≤ 1.5 | 4.35 | 5.25 | 6.48 |
| | | Ratio > 1.5 | 4.54 | 5.72 | 7.50 |
| | 0 | Ratio ≤ 0.6 | 9.12 | 17.35 | 34.82 |
| | | 0.6 < Ratio ≤ 1.0 | 6.15 | 8.72 | 12.88 |
| | | 1 < Ratio ≤ 1.5 | 4.90 | 6.20 | 8.00 |
| | | Ratio > 1.5 | 5.42 | 7.39 | 10.51 |
| 4 | 10 | Ratio ≤ 0.6 | 3.33 | 4.60 | 7.17 |
| | | 0.6 < Ratio ≤ 1.0 | 3.09 | 3.84 | 4.95 |
| | | 1 < Ratio ≤ 1.5 | 2.87 | 3.40 | 4.09 |
| | | Ratio > 1.5 | 2.92 | 3.54 | 4.44 |
| | 5 | Ratio ≤ 0.6 | 4.28 | 6.94 | 12.87 |
| | | 0.6 < Ratio ≤ 1.0 | 3.56 | 4.72 | 6.58 |
| | | 1 < Ratio ≤ 1.5 | 3.09 | 3.79 | 4.74 |
| | | Ratio > 1.5 | 3.24 | 4.15 | 5.57 |
| | 0 | Ratio ≤ 0.6 | 7.02 | 13.88 | 28.96 |
| | | 0.6 < Ratio ≤ 1.0 | 4.62 | 6.73 | 10.22 |
| | | 1 < Ratio ≤ 1.5 | 3.56 | 4.58 | 6.03 |
| | | Ratio > 1.5 | 3.96 | 5.53 | 8.06 |

*Note:* Estimates are based on model 4 of Table 1. Each rate is based upon 10,000 replications.

severities, as a fraction of mortgage balances and as a function of initial LTV, reported by Van Order and Zorn (1993). These severities, together with the time path of (unconditional) defaults, yield the time paths of losses.

Table 3 summarizes the estimated costs of the subsidy program. The table reports the present value of the losses over a 15-year period, using an interest rate of 10%. The first column of Table 3 reports the appropriate price, ex post, for mortgage loans with 10% downpayments made under each of the set of economic conditions described. For example, for a
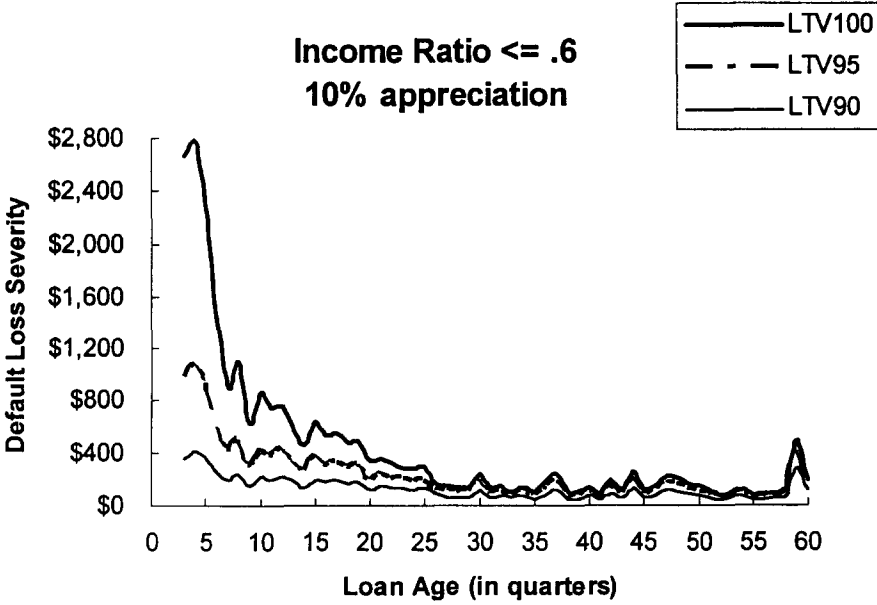
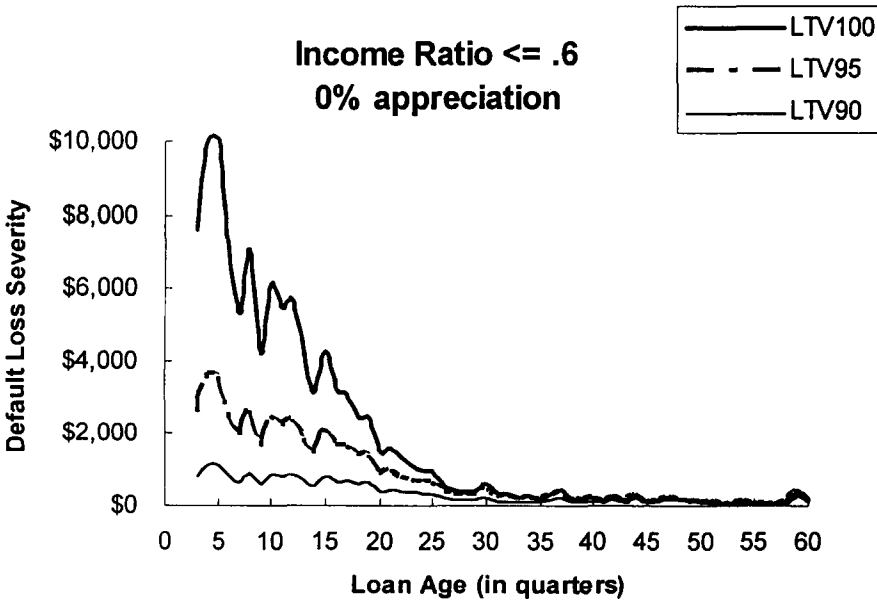Fig. 3. Simulated time path of present values of default losses.



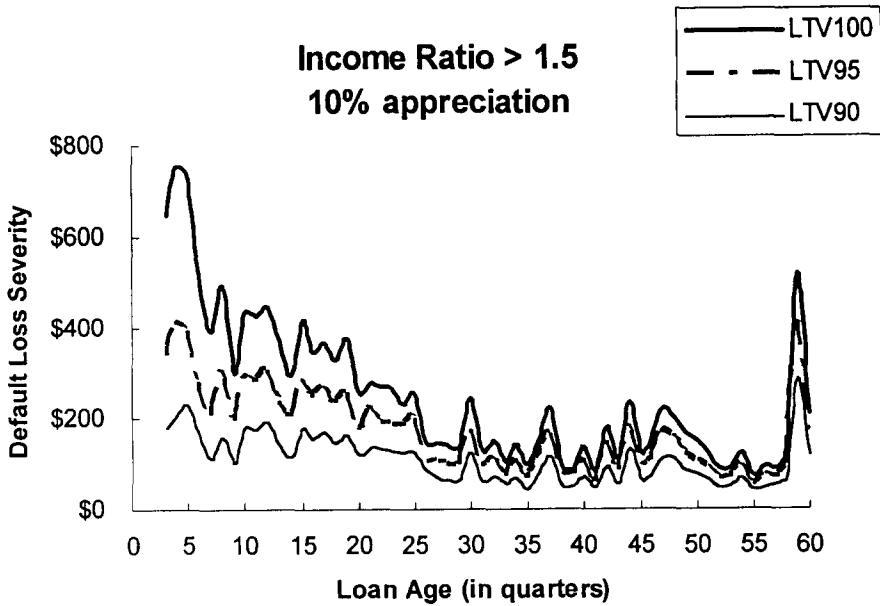Fig. 4. Simulated time path of present values of default losses.

Fig. 5. Simulated time path of present values of default losses.
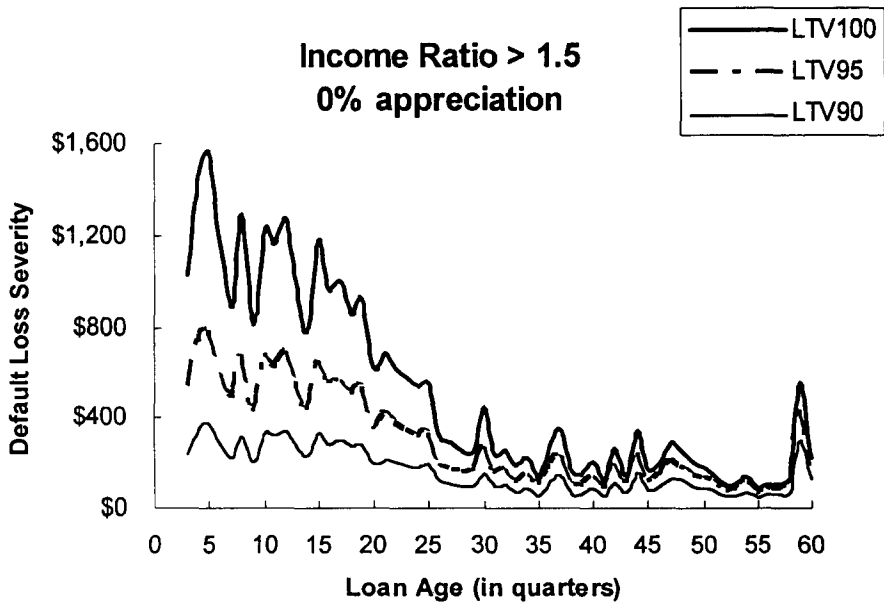
Fig. 6. Simulated time path of present values of default losses.

Table 3
Present value of losses per million dollars of program payments

| Unemployment rate (%) | Average annual house price change (%) | Household income at initiation to MSA median income | Downpayment rate | | |
|---|---|---|---|---|---|
| | | | 10% | 5% | 0% |
| 8 | 10 | Ratio ≤ 0.6 | $7,424 | $13,524 | $25,107 |
| | | 0.6 < Ratio ≤ 1.0 | 6,600 | 10,504 | 15,796 |
| | | 1 < Ratio ≤ 1.5 | 6,030 | 9,224 | 12,675 |
| | | Ratio > 1.5 | 6,278 | 9,902 | 14,424 |
| | 5 | Ratio ≤ 0.6 | 10,372 | 22,554 | 48,288 |
| | | 0.6 < Ratio ≤ 1.0 | 7,779 | 13,574 | 22,420 |
| | | 1 < Ratio ≤ 1.5 | 6,626 | 10,369 | 15,134 |
| | | Ratio > 1.5 | 7,194 | 12,188 | 19,062 |
| | 0 | Ratio ≤ 0.6 | 19,507 | 48,020 | 106,629 |
| | | 0.6 < Ratio ≤ 1.0 | 10,947 | 21,101 | 37,131 |
| | | 1 < Ratio ≤ 1.5 | 7,848 | 13,040 | 19,798 |
| | | Ratio > 1.5 | 9,424 | 17,438 | 29,738 |
| 4 | 10 | Ratio ≤ 0.6 | 5,575 | 10,484 | 19,759 |
| | | 0.6 < Ratio ≤ 1.0 | 4,960 | 8,032 | 12,307 |
| | | 1 < Ratio ≤ 1.5 | 4,500 | 6,844 | 9,473 |
| | | Ratio > 1.5 | 4,630 | 7,313 | 10,953 |
| | 5 | Ratio ≤ 0.6 | 7,974 | 17,811 | 39,069 |
| | | 0.6 < Ratio ≤ 1.0 | 5,967 | 10,765 | 17,802 |
| | | 1 < Ratio ≤ 1.5 | 4,996 | 7,854 | 11,531 |
| | | Ratio > 1.5 | 5,414 | 9,141 | 14,838 |
| | 0 | Ratio ≤ 0.6 | 15,442 | 39,392 | 89,814 |
| | | 0.6 < Ratio ≤ 1.0 | 8,542 | 16,872 | 30,368 |
| | | 1 < Ratio ≤ 1.5 | 5,924 | 10,085 | 15,512 |
| | | Ratio > 1.5 | 7,203 | 13,452 | 23,486 |

borrower who has household income at 60% of the MSA median income, the appropriate risk premium for 90% LTV loans, under circumstances in which there is a 10% increase in housing prices annually, is 0.74%. If we assume that the cost to consumers of zero downpayment mortgage loans is the same as appropriately priced 10%-down mortgage loans, then the unpriced subsidy is about 1.67%.

The costs are quite large if house prices do not appreciate. According to Table 3, if the unemployment rate is 8%, then for loans to households with incomes below 60% of the MSA median, the expected loss from a $50,000 mortgage issued is $5,332. If these mortgages are priced to consumers as if

they were 10%-down loans, then the cost to the U.S. Treasury would be $4,356 for each $50,000 mortgage originated.

## 5. Conclusions

This paper presents a competing risks model of mortgage terminations and uses the model to analyze the cost of a currently proposed housing subsidy policy, namely a policy of zero downpayment mortgage loans for selected households. The empirical model estimates jointly the competing risks of mortgage default and prepayment in a multiple hazard framework with time-varying covariates.

The model indicates the sensitivity of default to the initial LTV ratio of the loan and the course of housing equity. The latter is a measure of the extent to which the default option is in the money.

The empirical model is used to analyze the subsidy provided, and the program costs, of zero downpayment mortgages. The cost analysis is quite conservative, in that it assumes that the recipients of zero downpayment mortgages differ from households making substantial down payments only in the time path of their housing equity.

The actual program costs depend upon the pricing of these mortgages to consumers. If they were priced in a manner appropriate to mortgages with 10% downpayments, then the additional estimated program costs are around 2–4% of funds made available–when housing prices increase steadily.

In an economy where house prices do not appreciate, the costs of the program would be much larger indeed. Our estimates suggest that additional program costs would be between $74,000 and $87,000 per million dollars of lending, for borrowers with incomes below 60% of the MSA median. If the expected losses from the program were not priced at all, the losses from default could exceed 10% of the funds available for loans.

## Acknowledgements

## Appendix: Descriptive statistics of the sample of mortgage loans

Table A1
Frequency distribution for age of loan at termination

| Termination age | Frequency | Percent | Cumulative frequency | Cumulative percent |
|---|---|---|---|---|
| 2 | 9 | 0.1 | 9 | 0.1 |
| 3 | 39 | 0.3 | 48 | 0.1 |
| 4 | 85 | 0.7 | 133 | 1.1 |
| 5 | 124 | 1.0 | 257 | 2.2 |
| 6 | 142 | 1.2 | 399 | 3.4 |
| 7 | 143 | 1.2 | 542 | 4.6 |
| 8 | 164 | 1.4 | 706 | 5.9 |
| 9 | 183 | 1.5 | 889 | 7.5 |
| 10 | 139 | 1.2 | 1,028 | 8.7 |
| 11 | 144 | 1.2 | 1,172 | 9.9 |
| 12 | 178 | 1.5 | 1,350 | 11.4 |
| 13 | 243 | 2.0 | 1,593 | 13.4 |
| 14 | 257 | 2.2 | 1,850 | 15.6 |
| 15 | 148 | 1.2 | 1,998 | 16.8 |
| 16 | 220 | 1.9 | 2,218 | 18.7 |
| 17 | 216 | 1.8 | 2,434 | 20.5 |
| 18 | 293 | 2.5 | 2,727 | 23.0 |
| 19 | 287 | 2.4 | 3,014 | 25.4 |
| 20 | 348 | 2.9 | 3,362 | 28.3 |
| 21 | 325 | 2.7 | 3,687 | 31.1 |
| 22 | 340 | 2.9 | 4,027 | 33.9 |
| 23 | 488 | 4.1 | 4,515 | 38.0 |
| 24 | 532 | 4.5 | 5,047 | 42.5 |
| 25 | 490 | 4.1 | 5,537 | 46.7 |
| 26 | 465 | 3.9 | 6,002 | 50.6 |
| 27 | 457 | 3.9 | 6,459 | 54.4 |
| 28 | 469 | 4.0 | 6,928 | 58.4 |
| 29 | 419 | 3.5 | 7,347 | 61.9 |
| 30 | 326 | 2.7 | 7,673 | 64.7 |
| 31 | 323 | 2.7 | 7,996 | 67.4 |
| 32 | 272 | 2.3 | 8,268 | 69.7 |
| 33 | 353 | 3.0 | 8,621 | 72.7 |
| 34 | 236 | 2.0 | 8,857 | 74.6 |
| 35 | 296 | 2.5 | 9,153 | 77.1 |
| 36 | 226 | 1.9 | 9,379 | 79.0 |
| 37 | 123 | 1.0 | 9,502 | 80.1 |
| 38 | 163 | 1.4 | 9,665 | 81.5 |
| 39 | 188 | 1.6 | 9,853 | 83.0 |
| 40 | 177 | 1.5 | 10,030 | 84.5 |
| 41 | 171 | 1.4 | 10,201 | 86.0 |
| 42 | 201 | 1.7 | 10,402 | 87.7 |
| 43 | 113 | 1.0 | 10,515 | 88.6 |
| 44 | 112 | 0.9 | 10,627 | 89.6 |
| 45 | 81 | 0.7 | 10,708 | 90.2 |

Table A1 (*Continued*)

| Termination age | Frequency | Percent | Cumulative frequency | Cumulative percent |
|---|---|---|---|---|
| 46 | 101 | 0.9 | 10,809 | 91.1 |
| 47 | 52 | 0.4 | 10,861 | 91.5 |
| 48 | 92 | 0.8 | 10,953 | 92.3 |
| 49 | 100 | 0.8 | 11,053 | 93.1 |
| 50 | 103 | 0.9 | 11,156 | 94.0 |
| 51 | 71 | 0.6 | 11,227 | 94.6 |
| 52 | 251 | 2.1 | 11,478 | 96.7 |
| 53 | 214 | 1.8 | 11,692 | 98.5 |
| 54 | 45 | 0.4 | 11,737 | 98.9 |
| 55 | 24 | 0.2 | 11,761 | 99.1 |
| 56 | 30 | 0.3 | 11,791 | 99.4 |
| 57 | 45 | 0.4 | 11,836 | 99.7 |
| 58 | 5 | 0.0 | 11,841 | 99.8 |
| 59 | 2 | 0.0 | 11,843 | 99.8 |
| 60 | 23 | 0.2 | 11,866 | 100.0 |

Table A2
Distribution of loan terminations

| Type of termination | Frequency | Percent | Cumulative frequency | Cumulative percent |
|---|---|---|---|---|
| A: Termination | | | | |
| Default | 539 | 4.5 | 539 | 4.5 |
| Prepayment | 10,427 | 87.9 | 10,966 | 92.4 |
| B: Censored | 900 | 7.6 | 11,866 | 100.0 |

Table A3
Distribution of estimated equity ratios at termination

| Equity ratio | Frequency | Percent | Cumulative frequency | Cumulative percent |
|---|---|---|---|---|
| Equity ratio ≤ 0.1 | 195 | 1.6 | 195 | 1.6 |
| 0.1 < Equity ratio ≤ 0.2 | 500 | 4.2 | 695 | 5.9 |
| 0.2 < Equity ratio ≤ 0.3 | 1,516 | 12.8 | 2,211 | 18.6 |
| 0.3 < Equity ratio ≤ 0.4 | 2,414 | 20.3 | 4,625 | 39.0 |
| 0.4 < Equity ratio ≤ 0.5 | 2,687 | 22.6 | 7,312 | 61.6 |
| Equity ratio > 0.5 | 4,554 | 38.4 | 11,866 | 100.0 |

Table A4
Distribution of borrower income at origin relative to MSA median

| Ratio of household income to MSA median | Frequency | Percent | Cumulative frequency | Cumulative percent |
|---|---|---|---|---|
| Ratio ≤ 60% | 258 | 2.2 | 258 | 2.2 |
| 60% < Ratio ≤ +100% | 2,332 | 19.7 | 2,590 | 21.8 |
| 100% < Ratio ≤ 150% | 4,194 | 35.3 | 6,784 | 57.2 |
| 120% < Ratio ≤ 200% | 2,303 | 19.4 | 9,087 | 76.6 |
| Ratio > 200% | 2,779 | 23.4 | 11,866 | 100.0 |

Table A5
Distribution of loan-to-value ratio at origin

| LTV | Frequency | Percent | Cumulative frequency | Cumulative percent |
|---|---|---|---|---|
| LTV ≤ 60 | 1,383 | 11.7 | 1,383 | 11.7 |
| 60 < LTV ≤ 80 | 5,293 | 44.6 | 6,676 | 56.3 |
| 80 < LTV < 90 | 572 | 4.8 | 7,248 | 61.1 |
| 90 ≤ LTV < 95 | 2,831 | 23.9 | 10,079 | 84.9 |
| 95 ≤ LTV ≤ 100 | 1,787 | 15.1 | 11,866 | 100.0 |

Table A6
Distribution of year of origin of mortgage

| Year of origin | Frequency | Percent | Cumulative frequency | Cumulative percent |
|---|---|---|---|---|
| 1976 | 639 | 5.4 | 639 | 5.4 |
| 1977 | 1,195 | 10.1 | 1,834 | 15.5 |
| 1978 | 2,267 | 19.1 | 4,101 | 34.6 |
| 1979 | 2,704 | 22.8 | 6,805 | 57.3 |
| 1980 | 2,912 | 24.5 | 9,717 | 81.9 |
| 1981 | 1,127 | 9.5 | 10,844 | 91.4 |
| 1982 | 682 | 5.7 | 11,526 | 97.1 |
| 1983 | 340 | 2.9 | 11,866 | 100.0 |

# References

Bailey, M.J., R.F. Muth, and H.O. Nourse, 1963, A regression method for real estate price index construction, Journal of the American Statistical Association 58, 933–942.

Black, F. and M.S. Scholes, 1972, The pricing of options and corporate liabilities, Journal of Political Economy 81, 673–659.

Brennan, M.J. and E.S. Schwartz, 1985, Determinants of GNMA mortgage prices, AREUEA Journal 13, 209–228.

Buser, S.A. and P.H. Hendershott, 1984, Pricing default free mortgages, Housing Finance Review 3, 405–429.

Case, K.E. and R.J. Shiller, 1987, Prices of single-family homes since 1970: New indexes for four cities, New England Economic Review, Sept./Oct., 45–56.

Cox, D.R. and D. Oakes, 1984, Analysis of survival data (Chapman and Hall, New York).

Cox, J.C., J.E. Ingersoll and S.A. Ross, 1985, A theory of the term structure of interest rates, Econometrica 53, 385–467.

Cunningham, D.F. and C.A. Capone, Jr., 1990, The relative termination experience of adjustable to fixed-rate mortgages, Journal of Finance 45, no. 5, 1687–1703.

Cunningham, D.F. and P.H. Hendershott, 1984, Pricing FHA mortgage default insurance, Housing Finance Review 3, no. 4, 373–392.

Deng, Y.H., J.M. Quigley and R. Van Order, 1995, Mortgage terminations, Center for Real Estate and Urban Economics, Working Paper 95-230, University of California, Berkeley.

Dunn, K.B. and J.J. McConnell, 1981, Valuation of mortgage-backed securities, Journal of Finance 36, 599–617.

Foster, C. and R. Van Order, 1984, An option-based model of mortgage default, Housing Finance Review 3, no. 4, 351–372.

Foster, C. and R. Van Order, 1985, FHA terminations: A prelude to rational mortgage pricing, Journal of AREUEA 13, no. 3, 273–291.

Green, J. and J.B. Shoven, 1986, The effect of interest rates on mortgage prepayments, Journal of Money, Credit and Banking 18, 41–50.

Hendershott, P.H. and R. Van Order, 1987, Pricing mortgages: An interpretation of the models and results, Journal of Financial Services Research 1, 77–111.

Kau, J. and D. Keenan, 1995, An overview of the option-theoretic pricing or mortgages, Journal of Housing Research 6, no. 2, 217–244.

Kau, J., D. Keenan, W. Muller III and J. Epperson, 1992, A generalized valuation model for fixed-rate residential mortgages, Journal of Money, Credit and Banking 24, no. 3, 279–299.

Kau, J., D. Keenan, W. Muller III and J. Epperson, 1995, The valuation at origination of fixed rate mortgages with default and prepayment, Journal of Real Estate Finance and Economics 11, no. 1, 5–36.

Lancaster, T., 1990, The econometric analysis of transition data (Cambridge University Press, Cambridge).

Quigley, J.M. and R. Van Order, 1990, Efficiency in the mortgage market: The borrower's perspective, AREUEA Journal 18, no. 3, 237–252.

Quigley, J.M. and R. Van Order, 1995, Some tests of contingent claims models of mortgage default, Journal of Real Estate Finance and Economics 1, no. 2, 99–117.

Quigley, J.M., R. Van Order and Y.H. Deng, 1994, The competing risks for mortgage termination by default and prepayment: A minimum distance estimator, paper presented at the NBER Summer Institute, Cambridge, MA.

Schwartz, E.S. and W.N. Torous, 1989, Prepayment and the valuation of mortgage-backed securities, Journal of Finance 44, 375–392.

Van Order, R. and A. Schnare, 1994, Finding common ground, Secondary Mortgage Markets 11, no. 1.

Van Order, R. and P. Zorn, 1993, Income, location and default: Some preliminary results, unpublished paper.

Zellner, A., 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, Journal of the American Statistical Association 57, 348–368.