Journal of the American Real Estate and Urban Economics Association
1993, V21, 4: pp. 431–449

Mortgage Prepayment and Default Decisions: A Poisson Regression Approach

Eduardo S. Schwartz* and Walter N. Torous*

This paper uses an extensive and geographically dispersed sample of single-family fixed rate mortgages to assess the prepayment and default behavior of individual homeowners. We make use of Poisson regression to efficiently estimate the parameters of a proportional hazards model for prepayment and default decisions. Poisson regression for grouped survival data has several advantages over partial likelihood methods. First, when dealing with time-dependent covariates, it is considerably more efficient in terms of computations. Second, it is possible to estimate full-hazard models which include, for example, functions of time as well as multiple time scales (i.e., age of the loan and calendar time), in a much more straightforward manner than partial likelihood methods for ungrouped data. Third, Poisson regression can be used to estimate non-proportional hazards models such as additive excess risk specifications. Taken together, our data and estimation methodology allow us to obtain a better understanding of the economic factors underlying prepayment and default decisions.

Investigating residential mortgage prepayments and defaults is the subject of much recent research. See, for example, Foster and Van Order (1984, 1985), Vandell and Thibodeau (1985) and Schwartz and Torous (1989, 1992). This research interest stems from the importance of prepayments and defaults to the solvency of mortgage lending institutions and their significance in establishing the capital requirements of secondary market agencies such as Freddie Mac.

In addition, the valuation and hedging of mortgage securities, such as Collateralized Mortgage Obligations (CMOs), Interest Only (IOs) and Principal Only (POs) Stripped Mortgage-Backed Securities, critically depend upon the modeling of the prepayment and default behavior of the underlying borrowers. Even in the case of guaranteed mortgages, default has valuation implications since it also results in the termination of the loan, but for different economic reasons than prepayment.

*University of California, Los Angeles, California 90024

This paper uses an extensive and geographically dispersed sample of single-family fixed rate mortgages to assess the prepayment and default behavior of individual homeowners. We make use of Poisson regression to efficiently estimate the parameters of a proportional hazards model for prepayment and default decisions. Poisson regression for grouped survival data has several advantages over partial likelihood methods. First, when dealing with time-dependent covariates, it is considerably more efficient in terms of computations. Second, it is possible to estimate full-hazard models which include, for example, functions of time as well as multiple time scales (i.e., age of the loan and calendar time), in a much more straightforward manner than partial likelihood methods for ungrouped data. Third, Poisson regression can be used to estimate non-proportional hazards models such as additive excess risk specifications. Taken together, our data and estimation methodology allow us to obtain a better understanding of the economic factors underlying prepayment and default decisions.

Estimation Methodology

This section deals with the estimation of prepayment and default functions.[1] We introduce Poisson regression as a means of estimating the effects of time-varying explanatory variables on prepayment and default rates in light of the censored nature of our data.

A prepayment function gives the probability of prepaying a mortgage during a particular time period, conditional on the mortgage not having been previously prepaid. By expressing this conditional probability as a function of various explanatory variables or covariates, we may assess the statistical significance of these variables in influencing prepayment decisions.

Let the continuous random variable T represent the time until prepayment. The prepayment function π is defined by

$$\pi = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t} = \frac{f}{F}$$

[1] We cast this discussion in terms of prepayments. The same concepts, however, apply to defaults.

where F represents the survivor function

$$F = P(T \geq t)$$

while f is the probability density function of T

$$f = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} = -\frac{dF}{dt}.$$

As we discuss later, these functions may depend on various explanatory variables and corresponding parameters.

There are two approaches to evaluating a prepayment function. The first is based on the survival time characteristics of individual mortgages. In particular, this approach utilizes ungrouped data and follows a specific mortgage through time, determining when and under what economic circumstances it is prepaid or whether it has not been prepaid during the sample period (censored). This is the strategy taken, for example, by Schwartz and Torous (1989). The second approach determines for each time period the proportion of mortgages prepaid conditional on the total number of mortgages at risk at the period's beginning. Poisson regression, used in this paper, implements this latter approach and takes advantage of the grouped nature of our data. For more details on Poisson regression see, for example, Breslow et al. (1983).[2]

A Poisson regression treats the number of mortgages prepaid per unit of time as a Poisson random variable with an intensity hypothesized to depend on posited explanatory variables. Recall that a random variable X is Poisson distributed if its probability function is given by

$$P(X = x) \equiv f(x) = \frac{\pi^x e^{-\pi}}{x!} \quad x = 0, 1, 2, \ldots$$

where π denotes the intensity of the Poisson process.

While Poisson processes have been used in a variety of applications, what we emphasize is their applicability in modeling mortgage prepayments, as well as defaults. To that end, we interpret π as the prepayment inci-

---

[2] We thank Dale Preston for clarifying for us the use of Poisson regression in estimating proportional hazard models with time-varying covariates.

dence rate or, in other words, the probability per unit of time that a mortgage will be prepaid conditional on not having been previously prepaid.

To emphasize the dependency of this prepayment rate on various explanatory variables, we assume

$$\pi = \pi(\mathbf{v}, \boldsymbol{\theta})$$

where $\mathbf{v}$ = a vector of explanatory variables and $\boldsymbol{\theta}$ = a vector of parameters to be estimated.

Prepayment data is typically available on a monthly basis. If n_t is the number of mortgages outstanding at the beginning of month t, the expected number of prepayments during that month, μ_t, is then given by

$$\mu_t = n_t \pi(\mathbf{v}, \boldsymbol{\theta}).$$

If c_t mortgages actually prepay during month t, then according to the Poisson specification, we have

$$P(c_t) = \frac{\mu_t^{c_t} \exp(-\mu_t)}{c_t!} = \frac{[n_t \pi(\mathbf{v}, \boldsymbol{\theta})]^{c_t} \exp[-n_t \pi(\mathbf{v}, \boldsymbol{\theta})]}{c_t!}.$$

Assume we have sample prepayment data $\{c_1, \ldots, c_N\}$. The corresponding log-likelihood function is the logarithm of the product of the marginal probabilities

$$\ln L(\boldsymbol{\theta} | c_1 \ldots c_N) = \sum_{t=1}^{N} \ln P(c_t) \propto \sum_{t=1}^{N} \{c_t \ln [n_t \pi(\mathbf{v}, \boldsymbol{\theta})] - n_t \pi(\mathbf{v}, \boldsymbol{\theta})\}.$$

The maximum likelihood estimator is that value of $\boldsymbol{\theta}$ which maximizes this log-likelihood function.

As noted earlier, the same concepts developed for prepayments also apply to defaults, though the explanatory variables may differ. Moreover, given that Poisson regression uses the number of mortgages at risk, that is subject to either prepayment or default, we can separately estimate prepayment and default functions while explicitly taking into account the fact that during this period, a mortgage can prepay or default, but not both. In other words, in this paper we deal with the competing nature of these risks by fitting cause-specific hazard models [see Kalbfleisch and Prentice (1980), pages 163–188).

Following Green and Shoven (1986) and Schwartz and Torous (1992), we model both the prepayment and default functions by proportional hazards models of the form

$$\pi[\tau; \mathbf{v}(t), \alpha_\pi, \beta_\pi] = \pi_0(\tau, \alpha_\pi) \exp [\beta_\pi \mathbf{v}(t)]$$
$$\delta[\tau; \mathbf{v}(t), \alpha_\delta, \beta_\delta] = \delta_0(\tau, \alpha_\delta) \exp [\beta_\delta \mathbf{v}(t)]$$

where the baseline hazard functions π_0 and δ_0 measure the probabilities of prepayment and default, respectively, under homogeneous conditions, $\mathbf{v} = 0$, and depend explicitly upon the age of the mortgage, τ. As we note later, we also model the log baseline hazard functions as polynomials in τ.

Data

The data for this study were provided by Freddie Mac. We have information on a large sample of 1–4 family, 30-year fixed rate loans over the sample period 1975 through and including 1990.

For each loan, we were provided its origination quarter, coupon rate, loan to value ratio (LTV) class at origination[3] and information on which of Freddie Mac's five regions of the country the loan was originated in (North Central, Northeast, Southeast, Southwest and West). For each region and LTV class, we calculated the number of active loans, that is the number of loans originated in a particular origination year which are active at the quarter's beginning. We then compared this number of active loans to the number of loans which prepaid or defaulted during the quarter. In total, we have data on 39,032 prepayments and 8,559 defaults.

In addition, Freddie Mac provided housing index returns and volatility data by region of the country. For further details on this data and the

[3] Data on LTV ratios was provided by categories. We converted this data into a continuous variable by taking the midpoint of the category as the level of the variable.

procedures used, see Abraham and Schauman (1991). We use this data to obtain explanatory variables for our default and prepayment functions. Refinancing rates for each quarter in the sample period were obtained from the Federal Reserve Bulletin.

Empirical Results

In this section, we define the explanatory variables used in our prepayment and default functions. We then detail and estimate the corresponding proportional hazards models. This section concludes with a discussion of the economic significance of our results.

Explanatory Variables

Our empirical analysis includes a variety of explanatory variables. These variables capture the various financial, socioeconomic and regional factors affecting homeowners' prepayment and default decisions.

We consider the following categorical variables: the region of the country in which the mortgage was originated, the quarter of the year in which the loan was terminated and the age of the mortgage in years at the time of prepayment or default. These variables allow us to determine whether prepayment and default activity vary by region, season of the year and mortgage age.

The next variable considered is the mortgage's LTV ratio at origination. We expect that prepayments and, especially, defaults are related to the homeowner's initial equity in the property.

We also consider the corresponding cumulative, from origination, regional housing index return. At least in the case of defaults, we expect fewer defaults associated with higher cumulative returns. However, since the cumulative housing index only measures the average house appreciation (or depreciation) in a region, it neglects important information about extreme price movements which may trigger default (or prepayment). To capture this, we also include the volatility of the housing index return by region.

Refinancing opportunities are measured by the ratio of the prevailing mortgage rate to the loan's coupon rate. The smaller this ratio, the greater the financial incentive to prepay.

To further investigate the effects of refinancing opportunities on prepayments, we also consider four additional interest rate related explanatory variables. The first is the slope of the term structure, approximated by the difference between the current refinancing rate and the 90 day Treasury Bill rate. This variable provides information about borrowers' expectations of future interest rates which may affect prepayment and default decisions. Another variable indicating the possible future course of interest rates is the difference between this quarter's and last quarter's refinancing rate. We also include the volatility of refinancing rates approximated by the average of the last three months' squared changes in refinancing rates; the more variable the refinancing rates, the more valuable the option to prepay. Finally, we include a "burnout" variable to capture heterogeneity in borrowers. Intuitively, other things equal, the first time interest rates decrease to where it becomes financially advantageous to prepay, the eager to prepay borrowers (those who are either more informed or are subject to lower transaction costs) refinance. However, the remaining borrowers (those who are either less informed or are subject to higher transaction costs) will have a lower propensity to prepay. In particular, we define a burnout variable as the area measured by the difference between the coupon rate on the mortgage minus a fixed refinancing cost and the current refinancing rate, when this difference is positive:

$$burnout_t = \sum_{\tau=0}^{t} \max\,(m - k - r_\tau, 0)$$

where m = mortgage's coupon rate and k = refinancing cost. The larger $burnout_t$, all other things being equal, the slower prepayments.

Specification of Proportional Hazards Models

We estimate a prepayment model given by

$$\pi(\tau, \mathbf{v}(t), \beta_\pi) = \pi_0(\tau) \exp\left[\sum_{i=1}^{I} \beta_{i\pi}\mathbf{v}(t)\right] \tag{1}$$

where π_0 is the baseline hazard function. This categorical variable measures the effects of age on prepayments; since the sample includes no mortgages older in age than 16 years, we have 16 baseline parameters to estimate. The next four explanatory variables are categorical variables representing four of the five Freddie Mac regions of the country. The next three are categorical variables for the second through fourth quarters of the year, followed by the LTV ratio, the cumulative regional housing index return, the volatility of this return, the ratio of refinancing to coupon rates and the four additional interest related variables.[4]

Similarly, the particular default function that we estimate is given by

$$\delta[\tau, \mathbf{v}(t), \beta_\delta] = \delta_0(\tau) \exp\left[\sum_{i=1}^{I} \beta_{i\delta}\mathbf{v}_i(t)\right]. \tag{2}$$

We also estimate both prepayment and default functions using a polynomial specification for the log baseline hazard function as opposed to a categorical specification:

$$\pi[\tau, \mathbf{v}(t), \alpha_\pi, \beta_\pi] = \exp(\alpha_{1\pi}\tau + \alpha_{2\pi}\tau^2 + \alpha_{3\pi}\tau^3) \exp\left[\sum_{i=1}^{I} \beta_{i\pi}\, \mathbf{v}_i(t)\right] \tag{3}$$

$$\delta[\tau, \mathbf{v}(t), \alpha_\delta, \beta_\delta] = \exp(\alpha_{1\delta}\tau + \alpha_{2\delta}\tau^2 + \alpha_{3\delta}\tau^3) \exp\left[\sum_{i=1}^{I} \beta_{i\delta}\, \mathbf{v}_i(t)\right]. \tag{4}$$

In these specifications, τ represents the number of quarters since the loan's origination (age of the loan).

Results

Tables 1 and 2 report the Poisson regression results for the prepayment and default models, respectively.[5] In Table 1, Model 1 refers to equation (1), where the baseline probabilities are modeled as categorical variables, while Model 2 refers to equation (3), where the log baseline hazard function is given by a cubic polynomial in age. Similarly, in Table 2, Model 1 refers to equation (2), while Model 2 refers to equation (4).

---

[4] Note that there is one parameter associated with each mortgage age, but the number of parameters associated with each of the regions of the country as well as the season of the year is one less than the actual number of regions or seasons of the year. This follows from the fact that the first categorical variable (mortgage age) is specified to take into account the constant term in the prepayment function, implying that for the remaining categorical variables we need one parameter less than the number of categories (since having the four regional categorical variables and the three seasonal categorical variables all equal to zero implies that we are in the first region and the first quarter).

[5] The models were estimated using the program EPICURE: Risk Regression and Data Analysis Software developed by HiroSoft International Corporation.

From Table 1, we first observe that prepayments differ significantly[6] across regions of the country, with mortgages in the West prepaying fastest. Prepayments also vary significantly with the quarter of the year, prepayments being, as expected, fastest in the spring (second quarter) and slowest in the winter (first quarter). We also note that loans with higher LTV ratios at origination prepay significantly faster, indicating that more indebted homeowners are more concerned about refinancing. Cumulative housing index returns and their volatility are significant in explaining prepayments. The negative coefficient on housing index volatility is consistent with fewer prepayments for homeowners who experience significant decreases in their property values, thereby requiring larger loans to refinance when prepaying.[7] As expected, the refinancing variable is strongly significant with a negative sign, indicating that prepayments increase when refinancing rates decrease.

Our additional interest rate variables also explain prepayments. The burnout variable is significantly negative in its effect on prepayments, indicating that the more burned out the mortgage, the less likely prepayment.[8] We also see that when refinancing rates are increasing, prepayments increase as borrowers attempt to lock in what they believe to be relatively low rates. The significantly positive coefficient on the slope of the term structure may be due to the borrowers refinancing out of fixed rate mortgages into ARMs when long rates exceed short rates. Finally, the volatility of interest rates significantly affects prepayments, but in an opposite direction to what would be expected. Higher interest rate volatility increases the value of the prepayment option and therefore should decrease, not increase, prepayments.

Also from Table 1, we see that all the baseline parameters are significant. In Figure 1, we plot both baseline prepayment functions, zeroing out all other explanatory variables.[9] Notice that they initially increase

---

[6] Given the large sample of prepayments and default observations, as an additional check on the statistical significance of our results, we also computed likelihood ratio test statistics for the explanatory variables. This alternative procedure did not change our statistical conclusions.

[7] However, we expect these last two variables to have more economic significance in explaining defaults than prepayments.

[8] We also used refinancing costs of $k = 0$, 0.02 and 0.03 with no effect on our empirical results.

[9] We plot scaled baseline probabilities subject to the constraint that both baseline probability functions are the same at the end of the first year.

Table 1 ■ Poisson regression results—prepayments.

	Model 1		Model 2	
Explanatory Variable	Coef	Std. Error	Coef	Std. Error
West Region			−2.096	0.008
Southwest Region	−0.285	0.003	−2.379	0.008
Southeast Region	−0.277	0.003	−2.372	0.008
Northeast Region	−0.185	0.003	−2.282	0.008
North Central Region	−0.144	0.002	−2.241	0.008
2nd Quarter	0.536	0.002	0.541	0.002
3rd Quarter	0.459	0.002	0.434	0.002
4th Quarter	0.311	0.002	0.268	0.004
LTV ratio at origination	0.103	0.004	0.108	0.004
Housing Return ($\times 10^{-3}$)	0.225	0.020	0.242	0.020
Housing Variance	−0.913	0.083	−0.964	0.083
Refinancing Rate/Coupon Rate	−2.917	0.005	−2.026	0.005
Slope of Term Structure	0.107	0.001	0.122	0.001
Interest Rate Volatility	0.239	0.005	0.230	0.005
Direction of Interest Rates	0.068	0.001	0.062	0.001
Burnout (k = 0.01) ($\times 10^{-2}$)	−0.899	0.011	−0.803	0.008
Baseline Age 1	0.176	0.001		
Baseline Age 2	0.265	0.002		
Baseline Age 3	0.280	0.002		
Baseline Age 4	0.260	0.002		
Baseline Age 5	0.294	0.003		
Baseline Age 6	0.428	0.004		
Baseline Age 7	0.438	0.004		
Baseline Age 8	0.417	0.004		
Baseline Age 9	0.373	0.003		
Baseline Age 10	0.364	0.004		
Baseline Age 11	0.342	0.003		
Baseline Age 12	0.346	0.004		
Baseline Age 13	0.345	0.004		
Baseline Age 14	0.344	0.004		
Baseline Age 15	0.369	0.005		
Baseline Age 16	0.380	0.011		
Age in Quarters ($\times 10^{-1}$)			0.767	0.006
Age in Quarters^2 ($\times 10^{-2}$)			−0.172	0.002
Age in Quarters^3 ($\times 10^{-4}$)			0.118	0.002

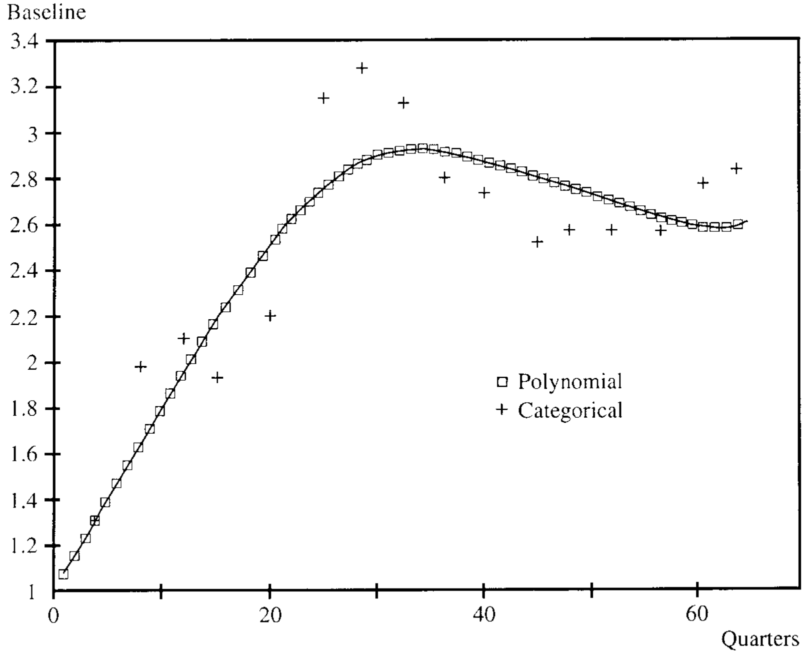
The dependent variable is the quarterly probability of prepayment. Asymptotic standard errors in parentheses.

Table 2 ■ Poisson regression results—defaults

	Model 1		Model 2	
Explanatory Variable	Coef	Std. Error	Coef	Std. Error
West Region			−12.84	0.087
Southwest Region	0.678	0.023	−12.16	0.090
Southeast Region	−0.690	0.023	−13.52	0.088
Northeast Region	−1.002	0.033	−13.83	0.091
North Central Region	−0.107	0.018	−12.94	0.088
2nd Quarter	0.277	0.019	0.259	0.019
3rd Quarter	0.289	0.017	0.257	0.017
4th Quarter	0.293	0.018	0.242	0.017
LTV Ratio at Origination	6.762	0.079	6.747	0.079
Housing Return ($\times 10^{-2}$)	−0.130	0.013	−.118	0.001
Housing Variance	3.790	0.766	3.651	0.769
Refinancing Rate/Coupon Rate	−2.160	0.034	−2.179	0.033
Slope of Term Structure	0.171	0.005	0.166	0.005
Interest Rate Volatility	0.162	0.036	0.169	0.036
Direction of Interest Rates	0.105	0.010	0.111	0.010
Burnout (k = 0.01) ($\times 10^{-1}$)	0.120	0.004	0.113	0.004
Baseline Age 1 ($\times 10^{-5}$)	0.519	0.042		
Baseline Age 2 ($\times 10^{-5}$)	0.856	0.068		
Baseline Age 3 ($\times 10^{-5}$)	1.184	0.094		
Baseline Age 4 ($\times 10^{-5}$)	1.275	0.104		
Baseline Age 5 ($\times 10^{-5}$)	1.244	0.106		
Baseline Age 6 ($\times 10^{-5}$)	1.150	0.102		
Baseline Age 7 ($\times 10^{-5}$)	0.989	0.090		
Baseline Age 8 ($\times 10^{-5}$)	0.901	0.084		
Baseline Age 9 ($\times 10^{-5}$)	0.857	0.082		
Baseline Age 10 ($\times 10^{-5}$)	0.762	0.075		
Baseline Age 11 ($\times 10^{-5}$)	0.672	0.069		
Baseline Age 12 ($\times 10^{-5}$)	0.573	0.064		
Baseline Age 13 ($\times 10^{-5}$)	0.400	0.052		
Baseline Age 14 ($\times 10^{-5}$)	0.454	0.071		
Baseline Age 15 ($\times 10^{-5}$)	0.419	0.095		
Baseline Age 16 ($\times 10^{-5}$)	0.253	0.149		
Age in Quarters ($\times 10^{-1}$)			1.824	0.060
Age in Quarters^2 ($\times 10^{-2}$)			−.603	0.022
Age in Quarters^3 ($\times 10^{-4}$)			0.541	0.025

The dependent variable is the quarterly probability of default. Asymptotic standard errors in parentheses.

Figure 1 ■ Categorical and Polynomial Baseline Prepayment Functions



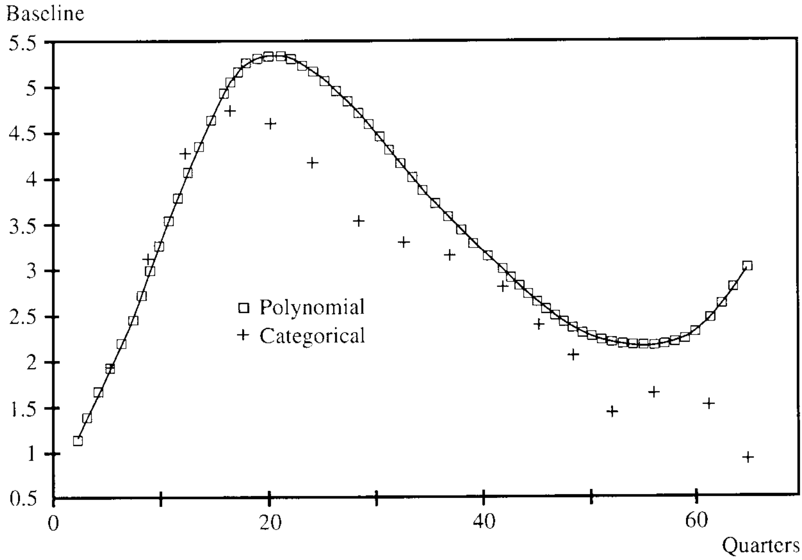
This figure shows the relationship between the baseline hazard functions and the age of the mortgage measured in quarters. The functions are scaled to be equal in the fourth quarter.

with increasing age, peaking at approximately 30 quarters, and then decrease.[10]

The default results are reported in Table 2. Defaults also differ significantly across regions of the country, with mortgages in the Southwest having the highest probability of default during the sample period. The parameters corresponding to the seasonal variables are also significant, though the differences between defaults in different quarters of the year are not as pronounced as in the case of prepayments. As expected, the mortgage's LTV ratio at origination is highly significant in explaining

[10] The figure is terminated at 64 quarters since our data contains loans with a maximum age of 64 quarters.

Figure 2 ■ Categorical and Polynomial Baseline Default Functions

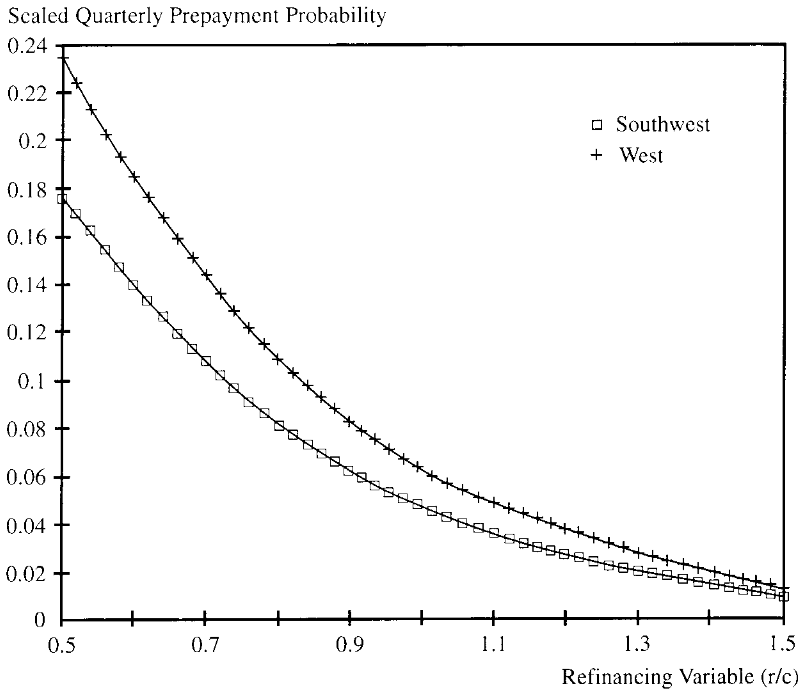


This figure shows the relationship between the baseline hazard functions and the age of the mortgage measured in quarters. The functions are scaled to be equal in the fourth quarter.

defaults, with higher LTV ratios implying higher probabilities of default. Also as predicted, cumulative housing index returns and their volatility significantly affect defaults, the first variable negatively (since lower property values imply higher probability of default) and the second variable positively (since it is primarily the lower end of this distribution which is subject to default). As in the case of prepayments, the refinancing variable is strongly significant with a negative sign, although in the case of defaults, the explanation is not so clear unless periods of low refinancing rates are associated with recessions and high unemployment. The additional interest rate variables are also significant, though the economic reasons why are also not clear.

Finally, all the default baseline parameters are significant in Table 2. Figure 2 plots these scaled baseline default functions. They increase steeply to attain a maximum at approximately 16 quarters and then decrease thereafter. Interestingly, the polynomial specification indicates that defaults eventually increase with increasing age, while the categorical spec-

Figure 3 ■ Prepayment Probability as a Function of the Refinancing Variable



This figure shows the relationship between the scaled quarterly probability of prepayment and the refinancing variable for both the Southwest and West regions.

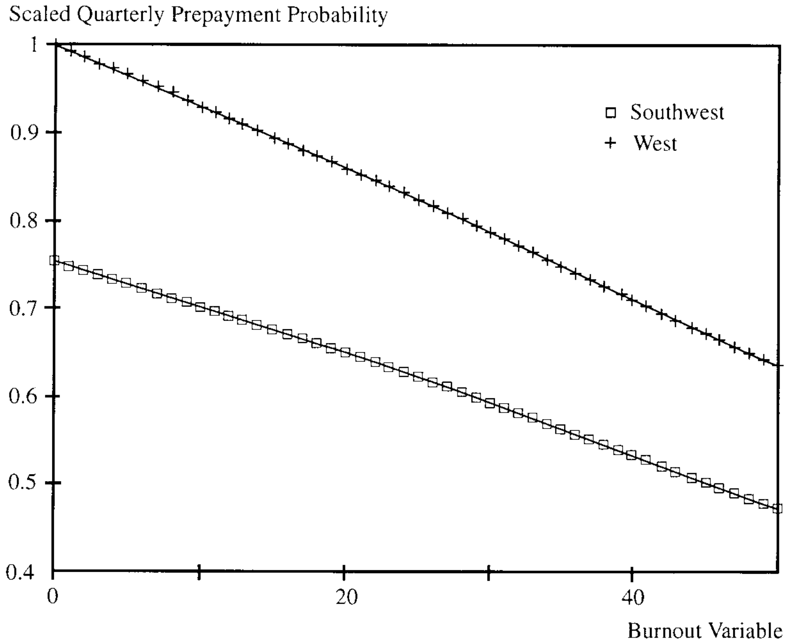
ification implies, more reasonably, that the probability of default continues to decrease with increasing age.

Discussion

This subsection examines the economic significance of the most important variables explaining prepayments and defaults. For comparison purposes, we zeroed out all other explanatory variables apart from the particular variable under investigation.

Figure 3 illustrates the effects of the refinancing variable on prepayments for loans in the quickest (West) and slowest (Southwest) prepaying regions of the country, holding all other variables constant. Notice that prepayments accelerate when the refinancing rate is lower than the loan's

Figure 4 ■ Prepayment Probability as a Function of the Burnout Variable



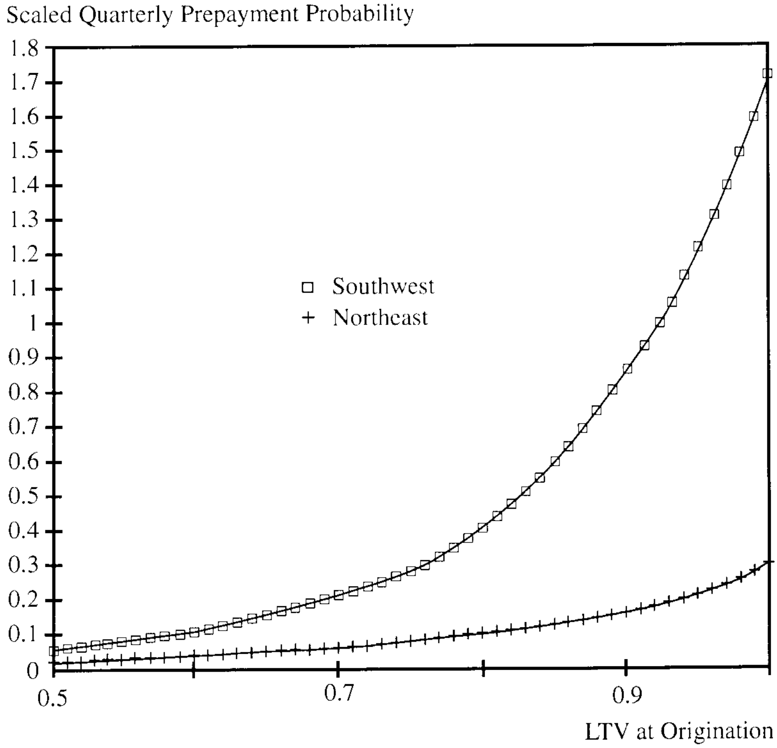
This figure shows the relationship between the scaled quarterly probability of prepayment and the burnout variable for both the Southwest and West regions.

coupon rate. Also, the regional differences in prepayment rates appear to be economically significant.

The effects of burnout for the quickest and slowest prepaying regions of the country are shown in Figure 4. From this figure, we see that burnout is an important variable explaining prepayment in both regions of the country, though this regional difference does not appear to vary with the level of burnout.

To investigate the significance of the LTV ratio at origination on defaults, Figure 5 presents default probabilities for the highest (Southwest) and the lowest (Northeast) defaulting regions of the country as a function of LTV, holding all other variables constant. Especially for the Southwest, increasing the LTV ratio at origination dramatically increases the probability of default. These regional differences suggest underlying socioeco-

Figure 5 ■ Default Probability as a Function of the LTV at Origination



This figure shows the relationship between the scaled quarterly probability of default and the LTV at origination for both the Southwest and Northeast regions.

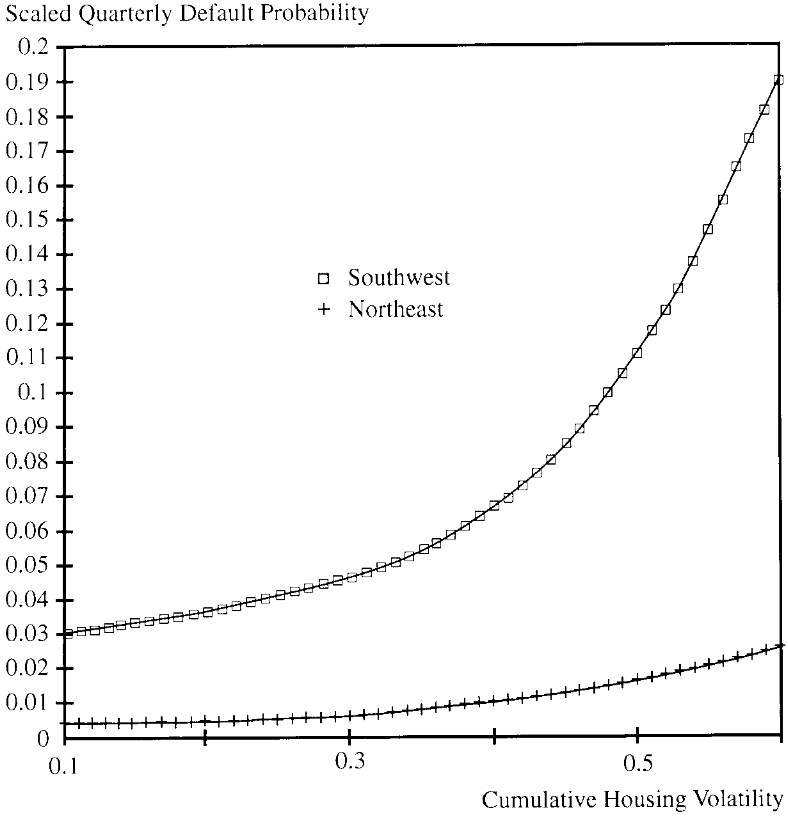
nomic variables, not included in our analysis, affect homeowners' default decisions.

Finally, Figure 6 shows default probabilities for these extreme regions, now as a function of cumulative housing volatility. As expected, higher volatilities result in a higher probability of default, especially in the Southwest region of the country.

Conclusions

This paper applies Poisson regression to estimate prepayment and default functions for fixed rate residential mortgage loans. This estimation meth-

Figure 6 ■ Default Probability as a Function of the Cumulative Housing Volatility



This figure shows the relationship between the scaled quarterly probability of default and the cumulative housing volatility for both the Southwest and Northeast regions.

odology has a number of advantages, including its ability to accommodate both the censored nature of our data as well as the time-dependent nature of the explanatory variables.

We make use of an extensive prepayment and default data set provided by Freddie Mac. These data allow us to categorize the region of the country where the loan was originated, the loan's LTV ratio at origination and other variables measured at the time when the loan was prepaid or defaulted. These latter variables include the mortgage's age, season of the year, the prevailing refinancing rate relative to the initial coupon rate, corresponding housing index returns and their volatility, as well as variables capturing the behavior of refinancing rates, including burnout.

Our empirical results indicate significant regional differences both in prepayment and default behavior. Refinancing opportunities play a significant role in explaining prepayments, while LTV ratios at origination and housing volatility significantly affect defaults. It is the volatility of housing index returns, as opposed to the index returns per se, which has the larger effect on defaults since index returns capture the average increase in house prices, while default depends on the lower tail of this distribution. Furthermore, the age of the mortgage plays a significant role in explaining both prepayments and defaults.

The next step in this research is to incorporate these empirically determined prepayment and default functions into a mortgage security valuation framework [see, for example, Schwartz and Torous (1992)]. This framework would not only recognize the importance of both prepayment and default to the valuation of these securities but also that different variables underlie homeowners' prepayment and default decisions.

Without implicating them, we thank Freddie Mac for providing the data for this study and for financial support in its initial stages. In particular, we thank Bob Van Order for his helpful comments. We also thank Dick Kazarian for his comments at the January 1993 AREUEA meetings in Anaheim, Dale Preston for his help in the use and understanding of the EPICURE software package used to fit our hazard functions and two anonymous referees as well as the editor, Kerry Vandell, for their comments and suggestions.

References

Abraham, J.M. and W.S. Schauman. 1991. New Evidence on House Prices from Freddie Mac Repeat Sales. *Journal of the American Real Estate and Urban Economics Association* 19(3): 333–52.

Breslow, N.E., J. Lubin, P. Marek and B. Langholz. 1983. Multiplicative Models and Cohort Studies. *Journal of the American Statistical Association* 78: 1–12.

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

———. 1985. FHA Terminations: A Prelude to Rational Mortgage Pricing. *Journal of the American Real Estate and Urban Economics Association* 13(3): 273–91.

Green, J. and J. Shoven. 1986. The Effects of Interest Rates on Mortgage Prepayments. *Journal of Money, Credit and Banking* 18(1): 41–59.

Kalbfleisch, J.D. and R.L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. John Wiley and Sons.

Schwartz, E.S. and W.N. Torous. 1989. Prepayment and the Valuation of Mortgage Backed Securities. *The Journal of Finance* 44(2): 375–92.

———. 1992. Prepayment, Default and the Valuation of Mortgage Pass Through Securities. *The Journal of Business* 65(2): 221–39.

Vandell, K.D. and T. Thibodeau. 1985. Estimation of Mortgage Defaults Using Disaggregate Loan History Data. *Journal of the American Real Estate and Urban Economics Association* 13(3): 292–316.