Improving Parametric Mortgage Prepayment Models with Non-parametric Kernel Regression

Authors   Michael LaCour-Little, Michael Marschoun
               and Clark L. Maxam

Abstract   Developing a good prepayment model is a central task in the
                valuation of mortgages and mortgage-backed securities but
                conventional parametric models often have bad out-of-sample
                predictive ability. A likely explanation is the highly non-linear
                nature of the prepayment function. Non-parametric techniques
                are much better at detecting non-linearity and multivariate
                interaction. This article discusses how non-parametric kernel
                regression may be applied to loan level event histories to produce
                a better parametric model. By utilizing a parsimonious
                specification, a model can be produced that practitioners can use
                in valuation routines based on Monte Carlo interest rate
                simulation.

Introduction

The importance of mortgage loan prepayments has been the topic of much academic and practitioner research. Many highly publicized losses by investors in mortgage-related derivative securities have been attributed to unexpected changes in prepayments. Understanding prepayment risk is important in assessing the risk, capital, solvency and insurance of financial institutions that originate and hold fixed-rate mortgages. Prepayment risk also affects the government and government-sponsored enterprises (GSEs) such as Fannie Mae, Ginnie Mae and Freddie Mac, which guarantee securities backed by mortgage loans.[1] In addition, much of the S&L crisis during the 1980s can be attributed to the poor management of interest rate risk (duration mismatch) in mortgage portfolios brought on by rapid, unforecasted and unhedged prepayment changes. At present, approximately $6.0 trillion of home mortgage debt is outstanding. With $4.0 trillion now securitized in the form of mortgage-backed securities (MBS) guaranteed by the GSEs, holdings of these securities have spread beyond traditional banking institutions to all types of financial institutions, money funds and investment groups. These securities have, in turn, been re-packaged into derivatives such as CMO's, IO's, PO's and inverse floaters, all of which can exhibit even greater

prepayment sensitivity. As a result, the analysis and forecasting of mortgage loan prepayments has become increasingly crucial to a growing group of investors as well as regulatory bodies.

This article discusses how non-parametric techniques can be used to improve parametric modeling resulting in better out of sample estimation. Maxam and LaCour-Little (2001) previously applied these techniques to mortgage pool data with promising results; however, purely non-parametric approaches are of little practical value given the necessity of off-the-support predictions in the Monte Carlo interest rate simulations used to actually value mortgages. This article extends the kernel regression estimation approach in Maxam and LaCour-Little to individual loan event histories. Prepayment probability is estimated as a function of the ''moneyness'' of the prepayment option, the age of the mortgage and the previous path of interest rates. The patterns produced by the kernel then guide development of a parametric alternative, which is shown to be a ''second best'' solution in terms of model fit.

A mortgage is frequently modeled as consisting of two components: a straight bond, which fluctuates with interest rates in the usual manner and an option component reflecting the borrower's right to prepay the mortgage and refinance at any time.[2] Thus, a mortgage or MBS investor is implicitly writing a call on the underlying fixed-rate bond. The household decision to prepay is based, of course, on a variety of factors, some directly related to interest rates while others reflect broader demographic factors. In contrast to option theory predictions, it is well documented that mortgage prepayment option exercise appears to be sub-optimal (Green and LaCour-Little, 1999). Mortgages are prepaid even when prevailing mortgage rates are higher than loan rates (when the option is out-of-the-money) and mortgages are not prepaid even when the loan rate exceeds the prevailing mortgage rate (when the option is deep in-the-money). This apparent irrationality on the part of borrowers is part of the problem in predicting prepayments.

Among the earliest to examine the topic of prepayment were Dunn and McConnell (1981), Brennan and Schwartz (1985), Green and Shoven (1986), and Quigley (1988), all of whom identified the role of interest rates as well as borrower mobility on rates of mortgage prepayment. In two often-cited articles, Schwartz and Torous (1989, 1993) use variations on the proportional hazard approach together with a Poisson regression to integrate prepayment into an overall valuation framework. Academic interest in the topic accelerated during the early 1990s with theoretical papers by Brueckner (1992, 1994), Follain, Scott and Yang (1992), Kau, Keenan, Muller and Epperson (1992) and Stanton (1995). These articles dealt with optimal exercise of the borrower's call option given stochastic interest rates and the implications for mortgage contract design and pricing, both of mortgages and mortgage-backed securities. Concurrently, Wall Street firms were developing proprietary prepayment models for use in valuation routines that supported their trading strategies (Richard and Roll, 1989; Patruno, 1994; Hayre and Rajan, 1995; Hayre, Chaudhary and Young, 2000). The refinancing wave of 1993 followed by the sharp increase in rates during 1994 produced large losses

for many market participants, reinforcing the business imperative to develop better models.[3]

During the latter half of the 1990s, many researchers turned their attention to institutional constraints that might reduce prepayments, even when call options appeared to be deep in the money (Peristiani, et al., 1996; Archer, Ling and McGill, 1996, 1997; Caplin, Freeman and Tracy, 1997; and Green and LaCour-Little, 1999). These studies identified declines in collateral value, credit status and other macroeconomic forces, such as unemployment, as significant factors inhibiting prepayment. Research in the late 1990s and into the new century re-focused on the role of borrower mobility (Clapp, Harding and LaCour-Little, 2000; and Pavlov, 2001), and a more complex specification of mortgage termination risk using a competing risk framework (Deng, 1997; Deng, Quigley and Van Order, 2000; and Ambrose and LaCour-Little, 2001).

Early research was generally based on mortgage pool data. But with pool data, a prepayment on a guaranteed mortgage may actually have been triggered by default on the underlying loan. Mortgage default requires the guarantor[4] (FNMA, GNMA, FHLMC) to pay off the balance on the loan to the investor. Thus, ''prepayments'' received by investors may have actually been defaults. Although it is early return of principal for whatever reason that MBS investors care about, models built with pool level models cannot distinguish the underlying cause.

In contrast, the individual loan level event histories in the data in this study permit an accurate distinction between prepayments and defaults. Therefore, the prepayment function is not confounded by the issue of commingling defaults. In addition, this study differs from previous work since non-parametric techniques are used as a basis for more powerful parametric specifications. Specifically, non-parametric kernel regression is employed to empirically estimate the empirical shape of the prepayment function. This information is used to enhance the parametric specification and to show that the new functional form improves out-of-sample model fit. The non-parametric technique can capture acknowledged properties of the prepayment function, such as premium burnout, seasoning and the refinancing incentive, without specifying a functional form or underlying distribution for any of the covariates.

The rest of the article is organized as follows. In the next section, the prepayment problem and related factors are described in more detail. The following sections describe the data and methodology and presents results for the non-parametric, parametric and hybrid models. The final section is the conclusion. The Appendix contains a detailed description of the non-parametric technique. Equations from the Appendix (prefixed with an ''A'') occasionally appear in the text.

The Prepayment Problem

Various theoretical models have modeled prepayments in an optimization framework using specific functional and distributional assumptions. The effort

here is more practical and applied. Thus, the following question: Given an actual data set that a practitioner might possess, how can non-parametric methods be employed to produce a model that could actually be used in valuation routines employing Monte Carlo simulation?

Much prepayment research has been done in the private sector rather than for purely academic purposes. Wall Street investment banks, major lenders, the GSEs and others have developed numerous econometric models that attempt to identify the appropriate factors and predict prepayments accurately. The majority of these are proprietary and described only in very general terms in research papers offered to existing and prospective clients.[5] Since the majority of prepayment models use mortgage pool data, they may likely suffer from serious estimation error. As evidence of the weakness of many of these models, note the large losses incurred by some Wall Street firms and other MBS investors over the last few years and the large research teams assigned to prepayment forecasting. Nonetheless, several common factors have emerged, although particular specification and proxies are not always described. These factors are the refinancing incentive, seasoning or age, seasonality and premium burnout. Though each factor will be discussed individually, this work focuses on just three variables: the refinancing incentive, seasoning and premium burnout. There are three reasons for this admittedly parsimonious model specification. First, as noted, the limited set of factors selected are those common to virtually all commercially available prepayment models. Second, the kernel regression framework naturally limits the number of factors that can be employed, due to the so-called "curse of dimensionality." Third, practitioners could actually use a short specification model in valuation routines requiring Monte Carlo simulation of interest rates.

Refinancing Incentive

Mortgagors typically have the option to prepay their mortgages at any time. The decision to exercise this option involves comparison of current payments with payments under the current refinancing rate. Refinancing involves explicit costs such as points, closing fees and so on as well as implicit costs such as ability to qualify, possible price depreciation affecting the amount that can be borrowed and other factors. A mortgage is simply a self-amortizing ordinary annuity. Thus, the decision to refinance amounts to retiring the mortgage with value M and taking out a new annuity with principal, P, plus paying the explicit and implicit costs that may be out of pocket expenses or added to the new principal or both. Furthermore, early exercise of this option entails lost time value. Thus, the prepayment option has all of the characteristics of an American option on an annuity with an exercise price equal to the outstanding principal balance plus refinancing costs and an exercise date equal to the maturity of the mortgage.[6] Thus, one measure of the refinancing incentive would be to take the ratio of the current mortgage (annuity) value to the outstanding principal balance, A/P.

This is essentially a measure of the moneyness of the option. Since collecting and computing this type of data is cumbersome to say the least, many researchers have used a simple metric, C-R, (Hakim, 1994) where C is the coupon rate on the mortgage and R is the gross refinancing rate reflecting the cost of refinancing. However, Richard and Roll (1989) argue that this measure is a poor approximation to A/P. They propose instead taking the ratio of the coupon rate to the refinancing rate, C/R. They document a highly non-linear relationship between this variable and conditional prepayment rates using a proprietary database of mortgage pools. In general, prepayment remains at frictional levels at C/R levels below one and accelerates rapidly as C/R exceeds one. This study employs natural log(C/R) as a measure of the prepayment incentive.

$$INCENT = \text{natural log (C/R).} \qquad\qquad (1)$$

Using the metric natural log(C/R) preserves the advantages the measure C/R offers over the measure C-R, but provides some computational advantages. In particular, using the measure C/R in a kernel with constant bandwidth implies that out-of-the money loans (C/R less than one) are smoothed more than in-the-money loans (C/R above one). Using the metric natural log(C/R) provides a simple way to avoid this unwanted asymmetric smoothing effect.[7]

Seasoning

It is also well documented that mortgage prepayments occur at different rates depending on the age of mortgage loans. The industry standard prepayment rate created by the Public Securities Association[8] (PSA) exhibits this pattern. The PSA "prepayment curve" assumes a constantly increasing prepayment probability per month for the first year, which then increases every month by a new constant increment until month thirty, and remains steady thereafter. However, this model ignores the effect of specific coupon rates. Premium GNMA pools, for example, typically season much more quickly than current coupon pools, which more closely follow the PSA prepayment curve. Some studies that use mortgage pool data, such as Schwartz and Torous (1993), simply use the age of the mortgage pool as a measure of the age of individual mortgage loans. As a result, these studies ignore the effect of seasoning of loans occurring prior to pool formation. This shortcoming is mitigated by the fact that the MBS-issuing agencies follow strict guidelines regarding the inclusion of already seasoned (*i.e.*, aged) mortgages in new pools. For example, GNMA pools are restricted to loans originated within one year of pool formation. Since actual loan age at time of prepayment can be observed, this study does not have any of the measurement problems endemic to pool data.

Premium Burnout

Borrowers face an array of costs, both explicit and implicit, when making the refinancing decision and these costs are likely to vary across borrowers. Thus, there is likely to be heterogeneity across households and different mortgagors may not repay given the same sets of parameters.[9] Heterogeneity produces a phenomenon commonly known as premium burnout.

To better understand premium burnout, consider the following scenario. When interest rates decrease to a certain point, the most eager and lowest cost borrowers in a mortgage pool find it favorable to prepay and therefore, refinance. The remaining borrowers are subject to higher refinancing costs, face difficulties obtaining credit, or are simply uninformed and thus have a lower likelihood to prepay. Thus, a different "critical level" of prepayment exists for different borrowers. As more and more favorable critical levels are reached more borrowers prepay. As a consequence, in an environment with fluctuating interest rates, the numbers of borrowers who have already prepaid grows each time a certain prepayment incentive is revisited and only the borrowers relatively less likely to prepay are left. As a result, the prepayment probability of a borrower who has been previously exposed to a certain prepayment incentive is expected to be lower than that of a borrower who is exposed to the same prepayment incentive for the first time. Moreover, there is evidence that households face different costs across time since prepayments are witnessed when a given critical level is reached more than once. Applied to mortgage pools this phenomenon produces premium burnout, which is the tendency for prepayments from premium pools to slow over time as more and more borrowers prepay and are removed from the pool. Several different measures of burnout have been used in the literature (see Hall, 2000) for a review). Adopting Schwartz and Torous (1993), this study measures burnout of loan *i* at time *t* as:

$$BURNOUT_{i,t} = \sum_{\tau=\tau_{Orig}}^{t} \max\{\log(c_i/r_\tau),0\}, \qquad (2)$$

where *c* is the mortgage coupon rate of loan *i*, *r* is the prevailing market rate for mortgage loan at time τ and τ_{Orig} is the time of origination. This measure accumulates the moneyness of the prepayment option (*INCENT* in Equation 1) over time. Moreover, during periods of high interest rate volatility this measure will assign large values to *BURNOUT*. Many other studies measuring burnout are hampered by pool data. A simple and readily obtainable measure of burnout used in some studies is the pool factor, defined as the ratio of the remaining balance to the original pool balance. This formulation suffers from severe measurement error since it includes scheduled amortization as well as prepayment. Thus, it is possible for a premium pool that has experienced significant repayment to have

the same pool factor as a discount pool that is well seasoned. Since data used in this study is at the loan level, it is not affected by any of these measurement issues.

Seasonality

The seasonal pattern of mortgage prepayments is also known to be important. Since household relocation follows a seasonal pattern of peaks in summer months and troughs in winter months, these patterns appear in prepayment rates as well. Many multiple regression models simply introduce dummy variables or lagged dummy variables to account for seasonality (Schwartz and Torous, 1993; and Hakim, 1994). Such an approach requires too many variables for kernel regression. Accordingly, the issue of seasonality is excluded here.

Other Factors: Interest Rates, Volatility, Geographic Region, Loan-to-Value Ratios and Credit Scores

While the refinancing decision depends on immediate interest rate savings, the level of rates will influence the refinancing decision, too. Although the measure of the refinancing incentive, log(C/R), captures the level implicitly, it does not account for borrower expectations about future rates. Many prepayment studies include either an interest rate level variable (usually the 10-year treasury rate) and/ or a term structure variable to capture the slope of the yield curve. In addition to the level of rates, option theory implies that volatility should also be important. Other things equal, an increase in volatility should increase the value of the prepayment option and thus reduce the likelihood that this option is exercised. In this study, the variable *BURNOUT* implicitly captures rate volatility.

The decision to refinance involves implicit as well as explicit costs and these costs may vary across borrowers. To account for heterogeneity several loan specific, geographic and macroeconomic variables have been suggested in the literature. Schwartz and Torous (1993) use a sixteen-factor proportional hazard model that accounts for economic region, loan-to-value ratio, regional housing return and volatility as well as the four variables mentioned above and lags of the preceding variables. Other studies claim to document relationships between prepayment and industrial production, lagged housing sales, other overall economic variables such as GNP, employment and CPI, and loan-specific variables such as loan-to-value ratios and credit scores (Peristiani et al., 1997). These factors are omitted in this study in order to focus on the narrower problem of producing a practically useful parametric model that utilizes non-parametric techniques.

Data

The data consists of monthly observations on 30-year conforming fixed rate mortgages (FRM) from the loan servicing records of Citicorp Mortgage, Inc.

during the years 1992–1997. The origination dates of the mortgage loans range throughout the entire period, and delinquent loans are excluded from the dataset. Available data is limited to loan amount, origination date, coupon rate and prepayment date. Exhibit 1 shows descriptive statistics. The data consists of about 60,000 loans, of which about 12,000 prepaid during the study period.

Exhibit 1 | Descriptive Statistics

Number of loans:	59,226
Fraction prepaid:	20.82%
Observation period:	1992:1–1997:12
Note rate:	7.98%
Original loan size:	$109,000

Note: These are 30-year fixed-rate mortgage loans.

Exhibit 2 shows the path of 30-year interest rates (based on the Freddie Mac primary mortgage market survey) and monthly conditional prepayment rates. Interest rates reached their minimum in late 1993, triggering a surge in prepayments. Interest rates fell again to relatively low levels at the beginning of 1996 and at the end of 1997 after increasing during the 1994–1995 period. Exhibit 2 also shows the sample prepayment probability as a function of the date. Prepayment probabilities are clearly dominated by the refinancing boom in late 1993 and early 1994, which coincide with the concurrent drop in interest rates. Note that none of the loans in the sample were older than two years when the 1993 refinancing boom occurred.

An event history file was created from the original data set. An individual record was created for each month of loan life, which contained a binary variable that indicated whether the loan prepaid in that particular month or not. For every loan-month record, the loan age (time in months since loan origination), a measure of refinancing incentive (the natural logarithm of the coupon rate divided by the market rate) and a measure of burnout (defined as the cumulative sum of non-negative refinancing incentive) were computed. This produced roughly 2,000,000 observations.

Exhibit 3 shows descriptive statistics for the event history data set. The average monthly prepayment probability is 0.63%, a number that is similar to values given in other studies.[10] The average spread is close to zero, however, a standard deviation of 0.11 shows that the sample contains substantial variety in this variable. For illustration, an observation in which the note rate is 10% above market (*e.g.*, a 7.7% rate in a 7% market) would yield a value of *INCENT* of approximately 0.1. The average *BURNOUT* is 1.2, which might represent, for example, a loan with a rate 10% above market for a period of twelve months. The

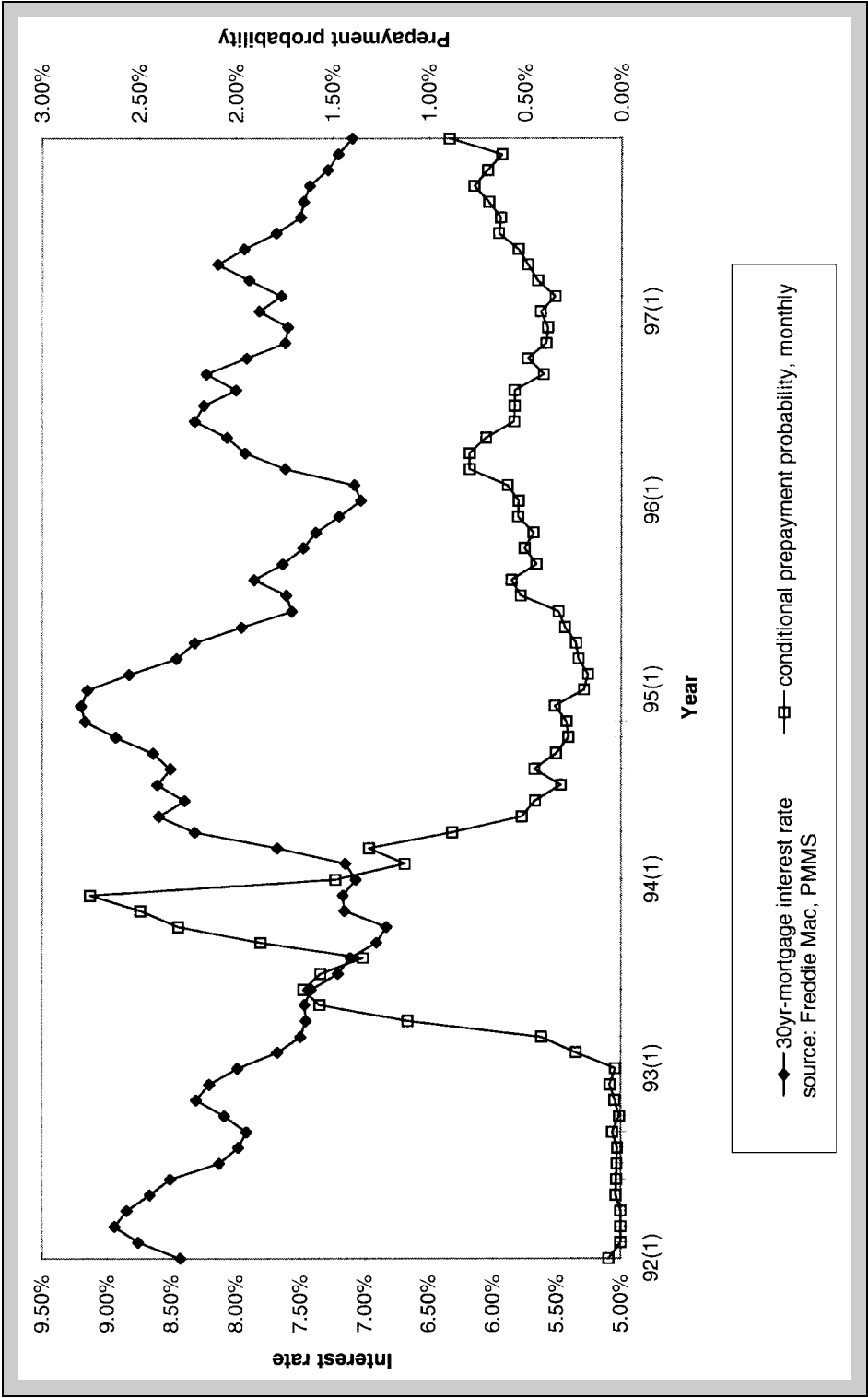**Exhibit 2** | Interest Rate Movements and Prepayment Probability

Exhibit 3 | Descriptive Statistics, Loan-Month Observations, Prepayment Events and Explanatory Variables

	Mean	Std. Dev.	Min.	Max.
INCENT[a]	0.013	0.11	−0.71	0.42
AGE (in months)	22.9	16.7	0	71
BURNOUT[b]	1.20	1.77	0.00	18.54

Notes: Total number of loan-months = 1,968,274; unconditional prepayment probability = 0.63%.

[a] *Incent*: Defined as natural log(mortgage rate / market rate).

[b] *Burnout*: Defined as sum of previous exposure of loan to positive values of *INCENT*.

range of burnout is heavily skewed, however, with the mode of the distribution at zero and a maximum of 18.5. Loan age ranges from zero to the full length of the observation period of seventy-one months, however, with increasing age the number of observations declines sharply. This decline in the numbers of observations with age occurs naturally given the data. Every loan generates an observation of a loan at age one, but only loans that were originated in January 1992 and were still alive in December 1997 can yield an observation on a loan at age seventy-one months.

Computation and Empirical Results

Baseline Parametric Specification

We begin with a baseline parametric model. The well-known logit model is a natural choice given the binary outcome data structure.[11]

$$\text{Prob } (P_{it} = 1|\mathbf{X}_{it}) = \exp (\mathbf{B}_{it}\mathbf{X}_{it})/(1 + \exp(\mathbf{B}_{it}\mathbf{X}_{it})), \qquad (3)$$

where $\mathbf{X} = X_{1t}, X_{2t}, \ldots X_{Kt}$, representing k explanatory variables indexed by month t, and $\mathbf{B}_i$ is the estimated effect of characteristic $\mathbf{X}_i$ on the probability that $P = 1$; here, $P = 1$ if the loan prepaid in month t, $P = 0$, if not.

Several specifications were evaluated and the one shown in Exhibit 4 gave the best model fit.[12] The specification, which employs quadratic terms in both spread and age, and a cubic term in spread, shows the expected signs and yields highly significant coefficient values. The signs of the coefficients suggest the familiar S-shaped response function. The coefficients on age and age squared suggest that the prepayments initially increase with age then later decrease. Exhibit 5 shows

Exhibit 4 | Baseline Logit Specification

Variable	Baseline
Constant	−6.78
	(0.05)
INCENT	6.40
	(0.18)
INCENT2	17.96
	(0.73)
INCENT3	−21.07
	(3.00)
AGE	0.093
	(0.003)
AGE2	−0.0012
	(0.0001)
BURNOUT	−0.058
	(0.009)

Note: Standard errors in parentheses.

**Exhibit 5** | Conditional Prepayment Baseline Logit Model Mid-Level Burnout
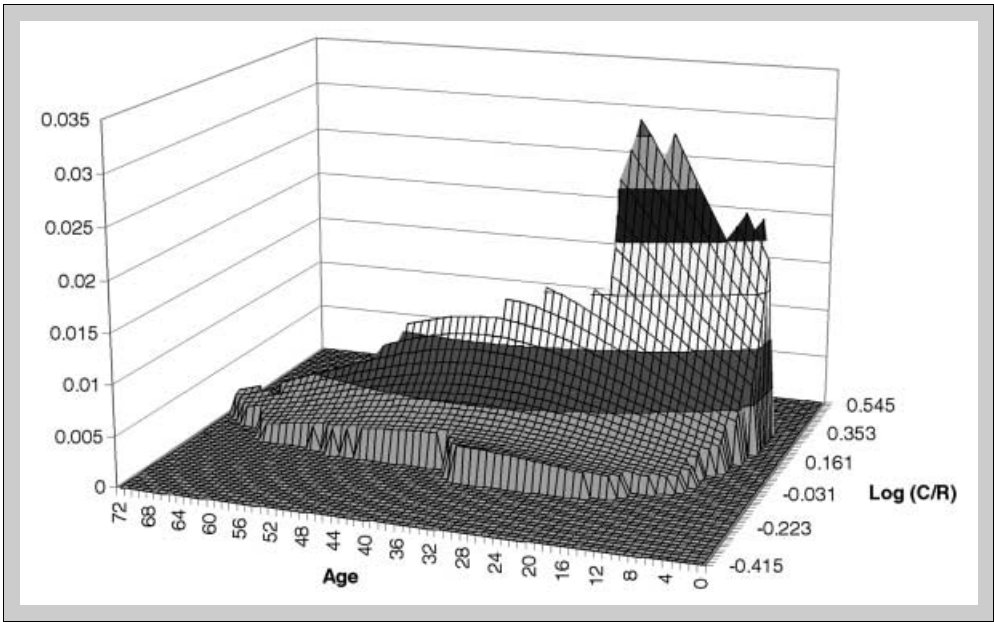
predicted prepayment probabilities for the logit model holding *BURNOUT* constant at one and letting *AGE* and *INCENT* vary.

Non-Parametric Kernel Regression

Most applications require an equation (A-5)[13] to be evaluated for a large number of values because the results are not parameter estimates of a function, as in conventional regression, but the numerical prepayment probability for one particular value of the explanatory variables. Each evaluation of an equation (A-5) requires the determination of the distances from the point for which the conditional prepayment probability is calculated to all 2,000,000 loan-month observations. These distances are then used to calculate the weights for all observations. The requirement to repeatedly calculate distances and weights for the entire data set explains why nonparametric regression is very computationally expensive. To reduce computation time, the continuous variables were transformed into discrete categorical levels. *INCENT* is transformed to take on fifty discrete values and *BURNOUT* can take on twenty-eight values. *AGE*, of course, is already discrete and can take on seventy-two values. Among the approximately 2 million total loan-months, there are only about 11,000 unique combinations of variables. Clustering according to these 11,000 combinations has the advantage that distances and weights may be limited to each combination only, rather than for each individual observation.

Equation (A-5) is employed to obtain the conditional estimate of the prepayment probability using the baseline bandwidth given in equation (A-3).[14] Exhibits 6 through 12 show the results evaluated for a dense grid of variable values. The pseudo-surfaces in the exhibits show the conditional prepayment probabilities letting two of the explanatory variables vary and holding the third constant. Exhibit 6 shows a cut through the three-dimensional structure of prepayment probabilities if age and spread (*INCENT*) are allowed to vary and burnout is held constant at zero. For areas of the grid where the density of underlying loan-month observations approaches zero the estimated prepayment probability is set to zero in order to avoid highly biased estimations. The kernel regression provides prepayment estimates for areas of the grid without observations, but those estimates have a large standard error and are not meaningful. The exhibits only show estimates of prepayment probabilities for combinations of variable values close to those that have been observed in the past. Extrapolation of prepayment probabilities to combinations never observed in the past should be undertaken with caution.[15]

Exhibit 6 shows the prepayment surface holding burnout constant at zero. It illustrates the highly nonlinear behavior of prepayment probabilities: below a spread of 0.1, prepayment probability is essentially flat, but increases quickly for values of *INCENT* above that threshold. The exhibit suggests that loans with zero burnout need to be exposed to positive spread for some time before prepayment probabilities increase. The exhibit also shows spikes in prepayment probability at

**Exhibit 6** | Conditional Prepayment Zero Burnout
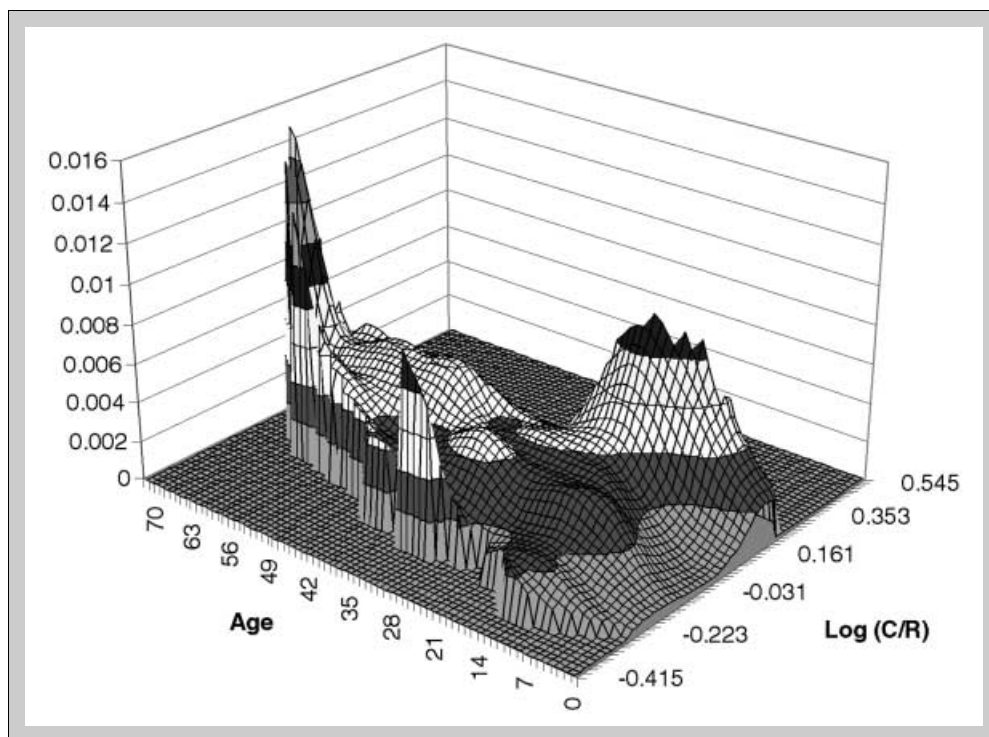
the edge of the pseudo surface, at low values of *INCENT*. These are examples where local noise in the observations may lead to questionable estimates. They could be avoided by more extensive smoothing (larger bandwidths), however, this would reduce the ability of the kernel regression to detect local variations in expected prepayment probabilities.

Exhibits 7 to 9 show the prepayment probabilities for increasing values of burnout. The effect of age on prepayment probability varies with burnout. In particular, at larger values of burnout, prepayment probabilities decrease with increasing age. Exhibits 10 to 12 show prepayment probabilities with varying *INCENT* and *BURNOUT* and holding *AGE* constant. Interestingly, the exhibit show little evidence that the prepayment probability declines as *BURNOUT* increases, after controlling for loan age, contrary to the conventional wisdom.

## Enhanced Logit Specification

The baseline model relies on two main variables, the refinance incentive (*INCENT*) and seasoning. The specification takes these variables and subjects them to some simple and intuitive transformations that reflect conventional wisdom about

Exhibit 7 | Conditional Prepayment Low Burnout
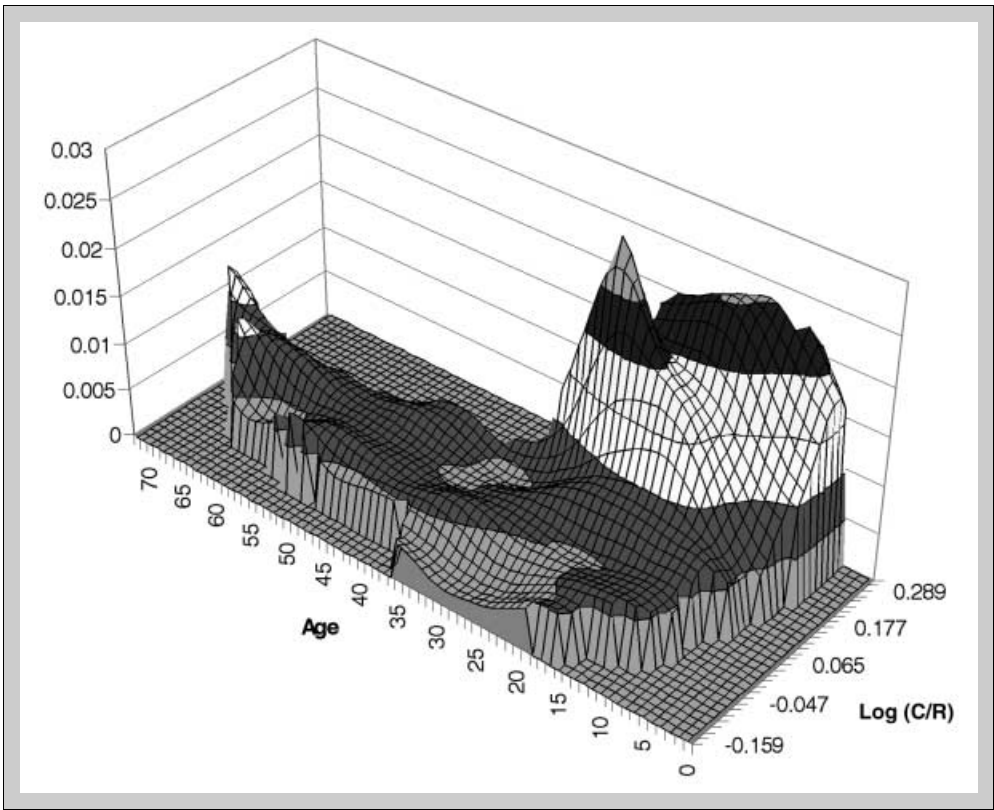
mortgage prepayments. The refinance incentive (*INCENT*) enters in its original form, in addition, its square and cube are used in the baseline specification. This transformation allows for an ''S'' curve in the prepayment response (*i.e.*, a sharp inflection followed by a gradual decrease in the rate of growth). The notion is that mortgages that are ''just in-the-money'' have the greatest interest rate sensitivity. Mortgages that are ''out of-the-money'' (coupon below market rate) and those that are ''deep in-the-money'' are less interest rate sensitive.

As shown in Exhibit 4, the signs on the powers of *INCENT* (positive on the linear and the square term, negative on the cube) are compatible with an S-curve. Similarly, prepayment speeds generally increase with seasoning, but then level off. This phenomenon is reflected, for instance, in the PSA ''ramp.'' To give the specification flexibility to accommodate this seasoning behavior, a square term is included next to the original seasoning variable (denoted ''*AGE*'') in the logit. The positive sign on *AGE* and the negative sign on its square are compatible with the existence of a ramp. The inclusion of the original burnout measure without any further transformations completes the original baseline specification.

Exhibit 8 | Conditional Prepayment Mid-Level Burnout
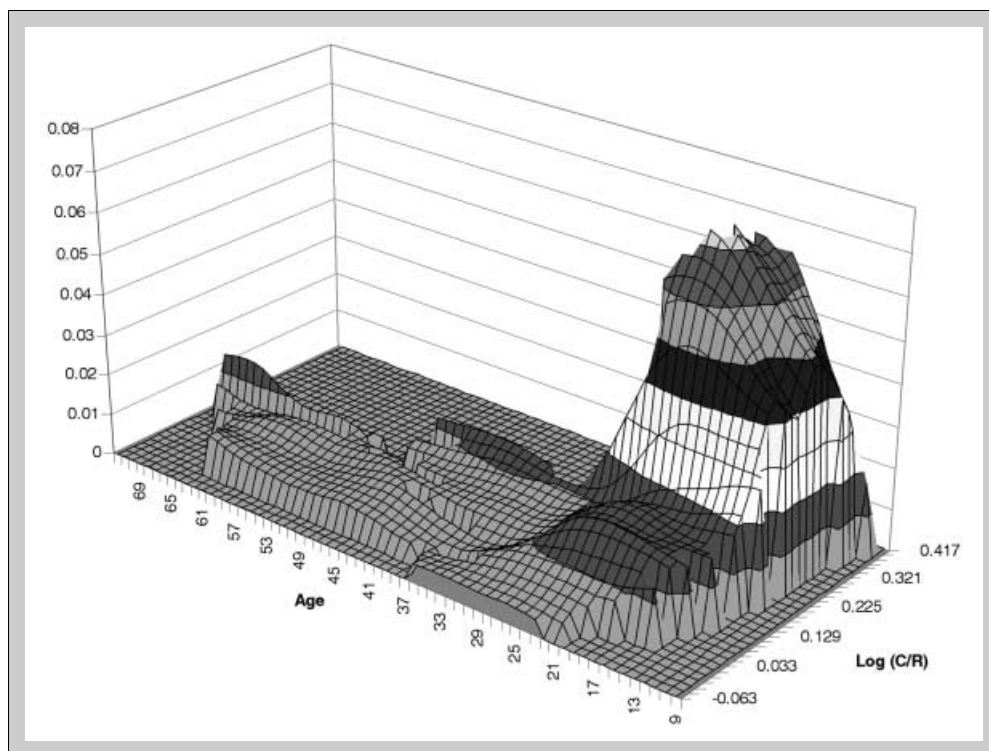
The task is now to improve on this model. The kernel charts (Exhibits 6–9) show the prepayment surface as *AGE* and *INCENT* are varied, holding *BURNOUT* constant at a low level. Comparing the kernel surface to the prepayment speeds implied by the baseline logit indicate two key differences between the actual prepayment behavior as measured by the kernel and the baseline logit. First, the kernel indicates that prepayment speeds increase at first but soon stabilize. In contrast, the effect of seasoning in the baseline logit specification is that of an arc: speeds increase at first and then decrease. A second discrepancy is the interaction of the refinance incentive and seasoning. The kernel reveals that the impact of the refinance incentive on prepayment speeds decreases as loans season. One explanation for this phenomenon is the decreasing remaining loan term and progressive loan amortization: given the same refinance incentive, a borrower with a more seasoned loan has a smaller gain from refinancing. Intuitively, the borrower has a shorter horizon over which to realize the savings from refinancing. This pattern is even more pronounced for mortgages with 15-year amortization, compared to the 30-year term loans used here. Because the decrease savings from refinancing with loan age occurs much faster in 15-year loans than in 30-year loans, this is a strong indication that this explanation holds. The weakening of the

Exhibit 9 | Conditional Prepayment High Burnout
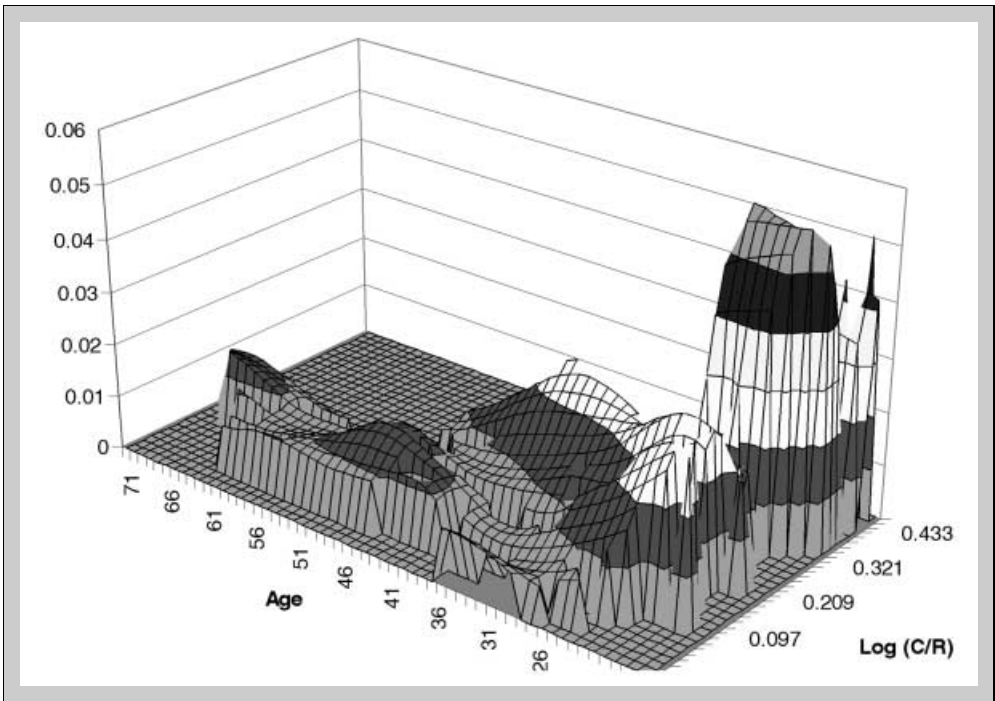
refinance incentive over time is an empirical fact that the baseline logit specification missed.

To overcome some of the shortcomings of the baseline logit specification in the enhanced logit, piecewise linear variables are used instead of polynomials. The original age variable is capped at sixty months and renamed *AGEC*. New variables, *AGE6* = max(*AGEC*-6,0) and *AGE12* = max(*AGEC*-12,0), are constructed. Estimating the logit specification with *AGEC*, *AGE6* and *AGE12* allows an estimation of the age profile as a continuous function with independent slopes for ages 0 to 6, 6 to 12 and 12 to 60. The slope for months 6 to 12 is equal to the sum of the coefficients on *AGEC* and *AGE6*, the slope for the months 12 to 60 is equal to the sum of *AGEC*, *AGE6* and *AGE12*. This avoids the arc in the age profile noted in the previous specification. To keep the specification symmetric, we define *INCENT04* = max(*INCENT*-0.04,0) and *INCENT22* = max(*INCENT*-0.22,0) to produce the S-curve. As a final innovation, we define an interaction between the refinance incentive and seasoning. Interactions are often difficult because they may have unintended consequences. In this application, we need to restrict the interaction such that it does not produce an inversion of the spread effect for highly seasoned loans, i.e. a case in which loans with greater

Exhibit 10 | Conditional Prepayment Age = 12
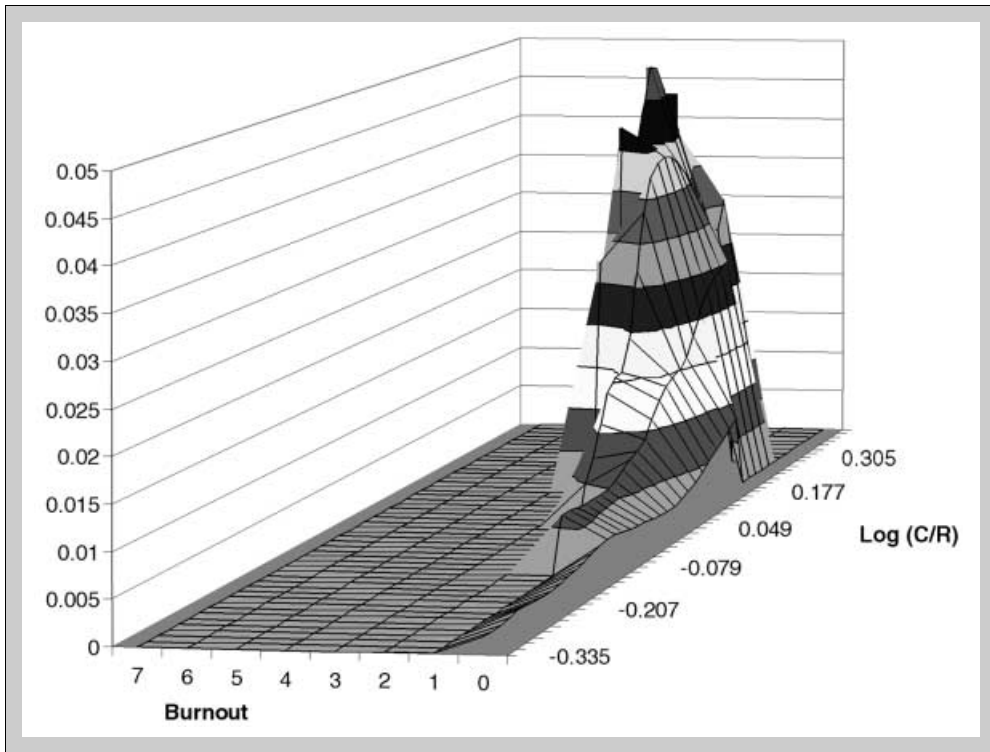
refinancing incentive prepay more slowly. Analysis of the kernel charts and the baseline logit charts suggested the following interaction: *INCENTAGE* = max(min(*AGEC*,24)*min(*INCENT*,0.1)-1,0). Finally, we add *BURNOUT* to the specification, which is unchanged.

Exhibit 13 shows the model statistics while Exhibit 14 shows that the new enhanced logit model implied prepayment surface much more closely matches the corresponding kernel chart (Exhibit 6). All variables are highly significant. Visually, the specification seems to work. The implied seasoning ramp is steep at first, but levels out thereafter, as suggested by the kernel. In the first six months, the probability of prepayment increases approximately by a factor of 1.44 every month (exp(0.367) = 1.44). But after twelve months, the seasoning effect becomes negligible. The very steep slope actually becomes slightly negative (0.3676 − 0.0798 − 0.2918 = −0.004), but the value is so small that it has little practical effect. The coefficients on the *INCENT* variable have the classic S-curve shape: for values of *INCENT* less than 0.04, the slope is 2.6. For values of *INCENT* between 0.04 and 0.22, the slope increases to 2.6 + 7.1 = 9.3, beyond values of

Exhibit 11 | Conditional Prepayment Age = 24
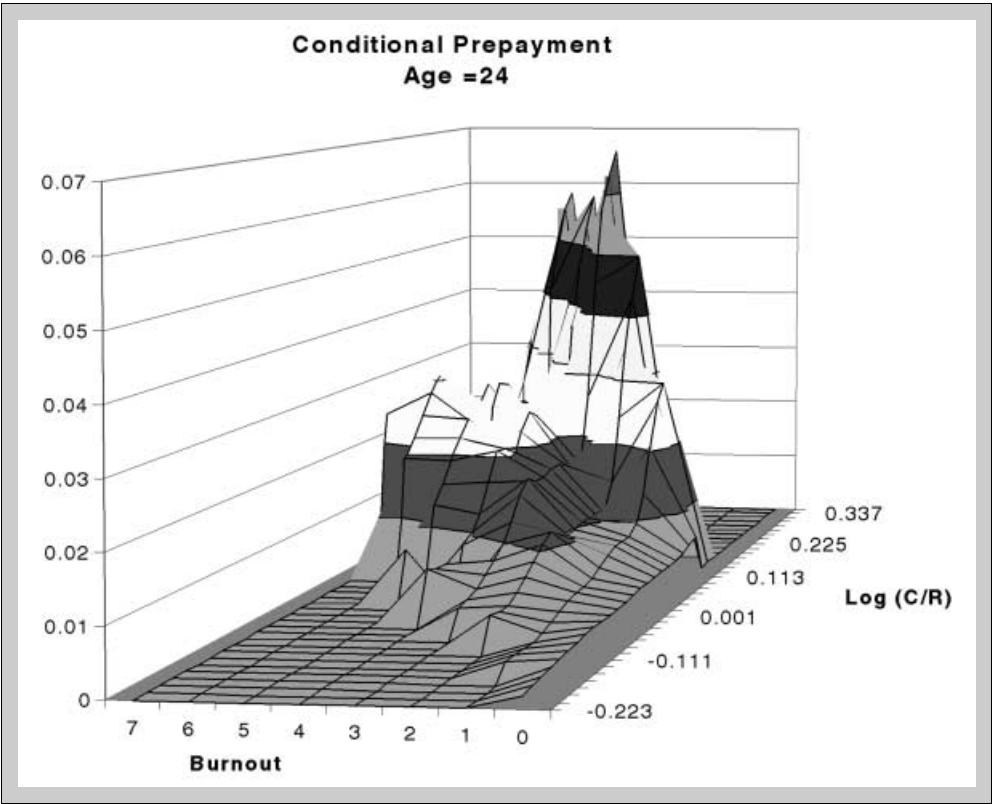
INCENT of 0.22 the slope decreases to $2.6 + 7.1 - 5.5 = 4.2$. The interaction of INCENT and AGE has the expected negative sign, implying those more seasoned loans are less rate sensitive. BURNOUT also has a negative sign, suggesting that loans that previously passed up refinancing opportunities are less likely to prepay; however, the strength of the effect is much smaller than one would infer from the baseline specification.

A more complex logit model was constructed thanks to the insights gained from the non-parametric regression. The appropriate interactions and age-dependent splines were selected based on examination of the kernel pseudo-surfaces. Do these changes produce a better model? The next section examines performance compared to the non-parametric and baseline parametric models.

Out-of-Sample Predictive Ability

Though the preceding discussion has emphasized the highly non-linear nature of the prepayment function, introducing a new technique is only warranted if it has
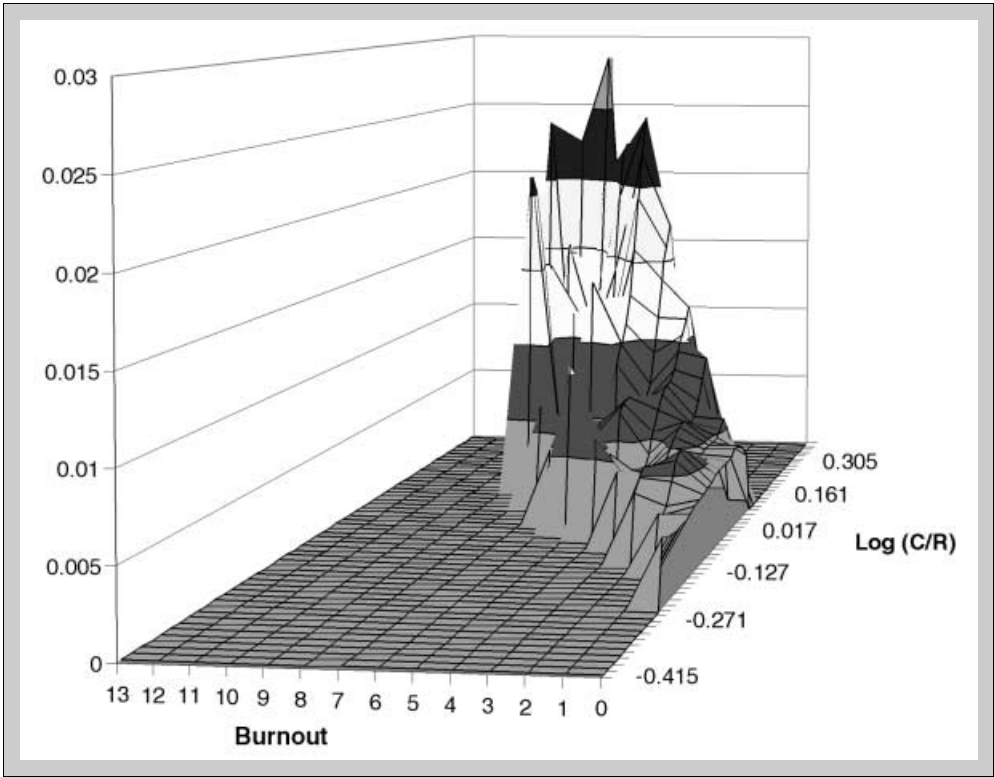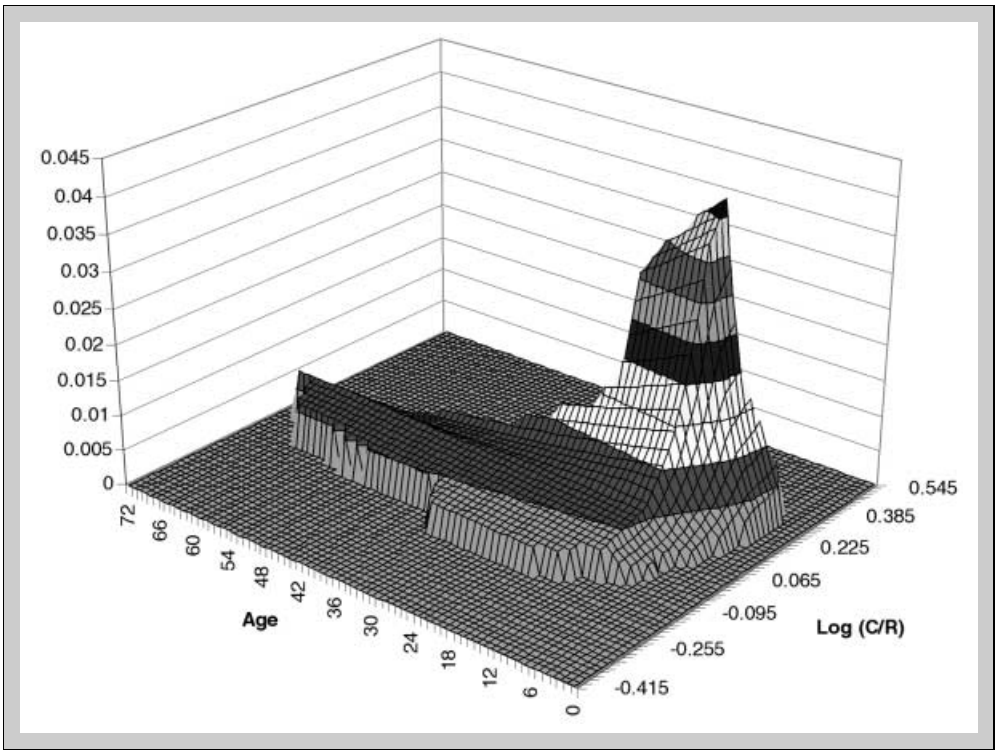
Exhibit 12 | Conditional Prepayment Age = 36

Exhibit 13 | Enhanced Logit Model

Parameters	Estimates	Standard Error
Constant	−9.116	0.179
INCENT	2.629	0.200
INCENT04	7.080	0.354
INCENT22	−5.485	0.500
AGEC	0.368	0.034
AGE6	−0.292	0.010
AGE12	−0.292	0.010
INCENTAGE	−0.185	0.029
BURNOUT	−0.010	0.006

Exhibit 14 | Conditional Prepayment Enhanced Logit Model Mid-Level Burnout

better predictive capabilities than existing alternatives. Traditional measures of in-sample fit such as $R2$ are not appropriate for a kernel regression estimator because kernel regression estimators can be made to fit any data arbitrarily well. A better test is superior out-of-sample predictive ability. An out-of-sample test for this purpose was conducted using the standard holdout technique. The original sample of mortgage loan observations was randomly divided into two subsamples of equal size and one was used to estimate the kernel regression as previously described. In addition, both the baseline logit and the enhanced logit were estimated using the same subsample. The predictive ability of the kernel regression model and the parametric alternatives were compared using the holdout sample.

Three distinct measures are used to compare predictive ability. The first is the Kolmogorov-Smyrnov (K-S) statistic, which is defined as the maximum separation of the cumulative density functions of the out-of-sample prepayment and non-prepayment events. This is a particularly popular metric in the credit-scoring arena. The K-S statistic of a model is calculated by assigning to each out-of-sample prepayment event its estimated prepayment probability. The empirical cumulative density function for these prepayment probabilities $Y_1, Y_2, \ldots, Y_n$ is defined as:

$$F_p(y) = (\text{number of } Y_i\text{'s less or equal } y)/n, \qquad (4)$$

for all real numbers y. Then, each out-of-sample non-prepayment event is assigned its estimated prepayment probability, and the empirical cumulative density function $F_{np}(y)$ is constructed analogously. The K-S Statistic is defined as:

$$\text{KS} = \max_y \{|F_{np}(y) - F_p(y)|\}. \qquad (5)$$

The K-S Statistic measures how well the estimator separates the two possible outcomes.[16] A disadvantage of the K-S Statistic is that it does not penalize inconsistent estimators. To prevent an inconsistent estimator from winning the race, two additional goodness-of-fit measures akin to the R^2 measure are used. One is Efron's R^2, which is defined as:

$$1 - \frac{\sum\limits_{i=1}^{n} (y_i - \hat{F}_i)^2}{\sum\limits_{i=1}^{n} (y_i - \bar{y})^2}, \qquad (6)$$

where y_i equals one in the case of prepayment and zero otherwise, $\hat{F}_i$ is the estimated prepayment probability and $\bar{y}$ is the in-sample average of y_i.[17] Efron's R^2 will be equal to zero if the squared error of the estimator is as large as the squared error of a naive estimator, which always predicts the sample mean. In contrast, Efron's R^2 will equal unity for a perfect estimator, which correctly estimates prepayment with probability one for all prepayments and prepayment with probability zero for all non-prepayments.

A third measure is McFadden's R^2, which is defined as:

$$1 - \frac{\ell_1}{\ell_0}, \qquad (7)$$

where ℓ is the log likelihood function $\sum_{i=1}^{n}[y_i \log \hat{F}_i + (1 - y_i) \log(1 - \hat{F}_i)]$, ℓ_1 is the out-of-sample log likelihood of the estimated model and ℓ_0 is the out-of-sample log likelihood for $\hat{F}_i = \bar{y}$. As in the previous example, McFadden's R^2 will equal

zero if the likelihood of the estimator does not exceed that of a naive estimator always predicting the sample mean, and will equal unity for a perfect estimator.

The final measure is concordance, the familiar ''c'' statistic output by most statistical packages when a binary outcome regression model is estimated. The measure is related to the K-S Statistic inasmuch as it measures the difference in ranking of the model's prediction and a naive estimate. To calculate this measure, each prepayment event in the sample is paired with each non-prepayment event in the sample. The concordance measure is the fraction of pairs for which the fitted prepayment probability of the prepayment event exceeds that of the non-prepayment event. A perfectly predictive model has a concordance measure of 100%.

The results of the out-of-sample comparisons of the predictive abilities of the three models are shown in Exhibit 15. The kernel regression model has the best goodness of fit statistics across all measures. The enhanced logit model is second in every measure and the baseline logit is last. If one interprets the R^2 values as linear measures, the improvement of the kernel regression estimate over the parametric models is large compared with the difference between the parametric models and a naive predictor, which always forecasts the sample mean. The enhanced logit model improves on the baseline by all measures. Note that these are out of sample results for an extremely large loan level categorical data set. Even with the tremendous advances in computational speed, large-scale

Exhibit 15 | Out-of-Sample Goodness-of-Fit Measures

	Kernel Regression	Baseline Logit	Enhanced Logit
K-S Statistic[a]	38.89%	33.98%	35.21%
Efron's R^{2}[b]	0.012	0.007	0.009
McFadden's R^{2}[c]	0.095	0.059	0.079
Concordance	0.772	0.729	0.752

Notes: Kernel regression estimate and parametric alternatives.

[a]K-S Statistic: defined as the maximum separation of the cumulative density functions of the out-of-sample prepayment and non-prepayment events.

[b]Efron's R^2: defined as $1 - \Sigma_{i=1}^{n} (y_i - \hat{F}_i)^2 / \Sigma_{i=1}^{n} (y_i - \bar{y})^2$, where y_i equals one in the case of prepayment and zero otherwise, $\hat{F}_i$ is the estimated prepayment probability, and where $\bar{y}$ is the in-sample average of y_i.

[c]McFadden's R^2: defined as $1 - \ell_1/\ell_0$, where ℓ is the log likelihood function $\Sigma_{i=1}^{n} [y_i \log \hat{F}_i + (1 - y_i) \log (1 - \hat{F}_i)]$, ℓ_1 is the out-of-sample log likelihood of the estimated model and ℓ_0 is the out-of-sample log likelihood for $\hat{F}_i = \bar{y}$.

[d]Concordance: $C_{i,t} = T^{-1} \Sigma_{t=1}^{T} [(X_{i,t} X_{j,t}) + (1 - x_{i,t})(1 - X_{j,t})]$, where t is sample size.

application of kernel regression estimation is still time consuming. The enhanced logit model, however, offers a significant improvement and is computationally efficient.

Conclusion

This article has presented non-parametric kernel regression as a technique for estimating prepayments using loan level data. The high degree of non-linearity in the relationships, and the large number of observations available in an event history data format, favor the use of non-parametric techniques that are free of both functional and distributional assumptions. As previously shown by Maxam and LaCour-Little (2001), the approach shows superior out-of-sample predictive capability when compared to conventional parametric regression. This suggests that identification of small-scale perturbations in the prepayment function can indeed increase predictive ability. Moreover, non-parametric tools can provide significant enhancement to parametric models of the sort practitioners might actually use in mortgage valuation.

As computer speeds continue to increase, similar techniques will become relatively more attractive, perhaps allowing a greater number of factors to be included in model specification. In addition to such technology-dependent extensions, we hope to incorporate the simple prepayment model developed here into a more complete Monte Carlo valuation framework.

Appendix

Kernel Density Estimators and Non-parametric Regression

The basic kernel estimator for estimating the joint density of a set of random variables is:

$$\hat{f}(x) = \frac{1}{t|H|} \sum_{i=1}^{t} K(H^{-1}(x - x_i)), \qquad \text{(A-1)}$$

of an unknown density $f(\mathbf{x})$ where $\mathbf{x}$ is a set of n-dimensional vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t$, t is the number of observations, $K(\cdot)$ is an appropriate kernel function and H is a bandwidth or smoothing parameter matrix. In words, the distance between an arbitrary row or vector, $\mathbf{x}$, is scaled by H and assigned a probability mass via the K. The scaled average of these masses is the estimate of the joint density at the given $\mathbf{x}$. Intuitively, one may think of a kernel estimate as a standardized distance between a point and every other data point, converted into a probability measure

based on distance. Points close to the data receive relatively more probability mass than those farther away. Adding each of these masses gives the kernel estimate. This is similar to constructing a histogram where H is the bin width of the histogram.

Research has shown that the choice of kernel function, $K(\cdot)$, is not nearly as important as the choice of the smoothing parameter or bandwidth, H (Hardle, 1990; and Scott 1992). The kernel density estimate is affected by the properties of the kernel, but due to the high frequency averaging process, various kernel functions are nearly equivalent.[18] The kernel should be a smooth, clearly uni-modal, symmetric function (Hardle, 1990; and Scott 1992). Many standard density functions fit these criteria. The Normal kernel is the most popular choice despite its high computational overhead when compared to simple kernels such as the Uniform kernel. Nonetheless, this study employs a multivariate Normal kernel as a point of reference due to its familiarity:

$$K(x) = (2\pi)^{-\frac{n}{2}} \exp\left(\frac{-x^T x}{2}\right). \qquad \text{(A-2)}$$

In contrast to kernel choice, optimal bandwidth is quite important to the estimator efficiency. Fortunately, theoretical research has identified optimal bandwidths for most popular choices of kernels. Scott (1992) shows that the optimal bandwidth for the independent multivariate normal is:

$$\hat{H} = \text{diag}(\Sigma^{-1/2} * t^{-1/(n+4)}), \qquad \text{(A-3)}$$

where Σ is the variance/covariance matrix of the covariates or vector $\mathbf{x}$, and n is the number or dimension of the factors.[19] Most standard kernel bandwidths can be obtained by simply multiplying Equation (2) by an appropriate constant.[20]

Since the kernel estimates the joint density of the variables under consideration, it is straightforward to obtain conditional and marginal densities using standard results. The theoretical regression estimate, $r(x)$, of a scalar Y given a vector $\mathbf{X}$ of explanatory factors is:

$$r(x) = E(Y|X = x) = \int yf(y|x)dy = \frac{\int yf(x, y)dy}{\int f(x, y)dy}, \qquad \text{(A-4)}$$

since $f(y|x) = \dfrac{f(x, y)}{f(x)}$ and $f(x) = \int f(x, y)dy$.

Substituting in the appropriate kernel estimators and simplifying yields the regression estimator:

$$\hat{m}(x) = E(Y|X = x) = \frac{\sum_{i=1}^{t} K(H^{-1}(x - x_i))Y_i}{\sum_{i=1}^{t} K(H^{-1}(x - x_i))}, \tag{A-5}$$

which is often called the Nadaraya-Watson estimator. H denotes the appropriate bandwidth parameter matrix for the covariate vector, $\mathbf{x}$.

Since the method relies exclusively on the data, its "denseness" or dimensionality is very important. To be effective large data sets should be employed with as many joint observations along the range of each of the variables as possible. Sparse observations can produce biased estimates (the dimensionality problem). When the data is sparse, the technique must extrapolate further and further distances in order to obtain an estimate. Thus, it is very important that the variables jointly cover as much of the feasible range as possible, but it is also important to realize that joint observations in some areas may be economically unlikely or impossible.[21]

Endnotes

[1] The proper status of the GSEs in the economy is a topic of ongoing controversy. Recent debate has focused on their level of risk exposure, their disclosure practices and their accountability to shareholders.

[2] Some small fraction of residential mortgage contracts, particularly in the sub-prime segment, contains prepayment penalties, typically priced at 25–75 basis points in note rate. Commercial mortgages, in contrast, customarily contain prepayment penalties.

[3] The most recent casualty involving the valuation of mortgages is Homeside Lending, Inc. This top ten U.S. mortgage lender was acquired by National Bank of Australia in 1997 for approximately $1.0 billion; in September 2001, the parent announced a write-off of $1.2 billion to cover unanticipated losses in the valuation of mortgages and mortgage servicing rights.

[4] Each of the agencies was originally created by Congress as government agencies in order to establish a national secondary market for government-backed mortgage securities. Since that time each has evolved to become a privately owned corporation. For example, Fannie Mae is currently the largest corporation in the U.S. when ranked by assets.

[5] For a general description of a typical Wall Street model estimated using mortgage pool data, see Hayre, Chaudhary and Young (2000). Maxam and LaCour-Little (2001) reports the functional form of the Citicorp model in use during 1992–1997, but not actual parameter values.

[6] In the absence of transaction costs, the analogy with an option is not completely correct, because in exchange for giving up the old prepayment option a refinancing borrower obtains a new loan with its own prepayment option.

[7] The kernel regression model was estimated using all three different measures of spread. The estimated prepayment probabilities and the goodness-of-fit measures are largely insensitive to the choice of the measure of spread.

[8] The Public Securities Administration is now called The Bond Market Association.

[9] For work on the role of borrower characteristics in determining prepayments, see Archer, Ling and McGill (1996, 1997) and LaCour-Little (1999).

[10] For instance, Mattey and Wallace (2001) report a weighted average monthly mortality rate of about 1.0% for Freddie Mac pools issued from 1991 to 1994.

[11] A slight change to the logit distributional assumption produces a model that is equivalent to the proportional hazard model previously mentioned and often used by academic researchers in the field.

[12] A base model with linear variables in both age and spread was evaluated, along with a model that uses squared values of age and spread as variables in addition to those of the base model, and a third model that adds cubed spread as a variable. The specification with the best goodness-of-fit measures was selected.

[13] Equations discussed in this section may be found in the Appendix.

[14] The kernel bandwidth determined by Equation (2) is 0.02 for spread, 2.4 for age and 0.29 for burnout. Comparing the size of the kernel with the standard deviations of the explanatory variables shown in Exhibit 2 shows that the size of the kernel is small compared with the support of the variables.

[15] Kernel regression calculates a weighted mean of local observations. However, if no local observations are available, a gaussian kernel will use a weighted mean of the closest neighbors, which can be arbitrarily far away, as a conditional expectation.

[16] See Law and Kelton (1982). Exhibits 10 to 12 show the empirical c.d.f.s for the three models.

[17] For an overview of goodness-of-fit measures of dichotomous models see Estrella (1998) and Amemyia (1981).

[18] Though an abuse of terminology, a type of ''Central Limit'' obtains.

[19] Optimal bandwidth depends on a number of factors, especially the density of observations. A computationally intensive test whether it is appropriate to use Equation (2) to determine the bandwidth can be conducted with cross validation using the leave-one-out method. This method evaluates the sum of variance plus bias for a given bandwidth by calculating the mean squared error for all observations in the sample, where the conditional expectation y for a given observation is calculated using all observations but the observation for which the conditional expectation is calculated. Searching over a grid of bandwidths for minimum squared error yields the optimal bandwidth (see Hardle, 1990). This method of cross-validation was used to determine the optimal bandwidth for the kernel regression using the data set. Due to the size of the data set, the optimal bandwidth derived using cross validation turned out to be very close to the result derived using Equation (2). Because cross validation requires exorbitant computation time, and determining the bandwidth using Equation (2) was shown to yield acceptable results in a similar context, Equation (3) was used in this study.

[20] For example, Scott (1992) gives an equivalent smoothing factor of 1.74 for equating the optimal Normal kernel bandwidth to the optimal Uniform kernel bandwidth.

[21] For example, in the empirical work here the combination of high burnout and low mortgage age is extremely unlikely.

References

Ambrose, B. and M. LaCour-Little, Prepayment Risk in Adjustable Rate Mortgages Subject to Initial Year Discounts: Some New Evidence, *Real Estate Economics*, 2001, 29:2, 305–28.

Amemiya, T., Qualitative Response Model: A Survey, *Journal of Economic Literature*, 1981, 19:4, 1483–1536.

Archer, W., D. Ling and G. Gill, The Effect of Income and Collateral Constraints on Residential Mortgage Terminations, *Regional Science and Urban Economics*, 1996, 26, 235–61.

——., Demographic Versus Option-Driven Mortgage Terminations, *Journal of Housing Economics*, 1997, 6:2, 137–63.

Brennan, M. J. and E. S. Schwartz, Determinants of GNMA Mortgage Prices, *Journal of the American Real Estate and Urban Economics Association*, 1985, 13, 209–28.

Breuckner, J., Borrower Mobility, Self-Selection, and the Relative Prices of Fixed- and Adjustable-Rate Mortgages, *Journal of Financial Intermediation*, 1992, 2, 401–21.

——., Borrower Mobility, Adverse Selection, and Mortgage Points, *Journal of Financial Intermediation*, 1994, 3, 416–41.

Caplin, A., C. Freeman and J. Tracy, Collateral Damage: How Refinancing Constraints Exacerbate Regional Recessions, *Journal of Money, Credit, and Banking*, 1997, December, 496–516.

Clapp, J. M., J. P. Harding and M. LaCour-Little, Expected Mobility: Part of the Prepayment Puzzle, *Journal of Fixed Income*, 2000, June, 68–78.

Deng, Y., 1997. Mortgage Termination: An Empirical Hazard Model with a Stochastic Term Structure, *Journal of Real Estate Finance and Economics*, 1997, 14:3, 310–331.

Deng, Y., J. Quigley and R. Van Order, Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options, *Econometrica*, 2000, 68:2, 275–307.

Dunn, K. and J. McConnell, Valuation of GNMA Mortgage Backed Securities, *Journal of Finance*, 1981, 36, 599–616.

Estrella, A., A New Measure of Fit for Equations With Dichotomous Dependent Variables, *Journal of Business and Economic Statistics*, 1998, 16:2, 198–205.

Follain, J. R., J. O. Scott and T. L. T. Yang, Microfoundations of a Mortgage Prepayment Function, *Journal of Real Estate Finance and Economics*, 1992, 5, 197–217.

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

Green, R. K. and M. LaCour-Little, Some Truths About Ostriches: Who Never Refinances Their Mortgage and Why They Don't, *Journal of Housing Economics*, 1999, 8: 233–48.

Green, W., *Econometric Analysis*, Second Edition, MacMillan, 1993.

Hakim, S., Regional Diversity, Borrower Characteristics and Mortgage Prepayment, *Review of Financial Economics*, 1994, 1:2, 17–30.

Hall, A., Controlling for Burnout in Estimating Mortgage Prepayment Models, *Journal of Housing Economics*, 2000, 9:4, 215–32.

Hardle, W., 1990, *Applied Nonparametric Regression*, New York, NY: Cambridge University Press, 1990.

Hayre, L. and A. Rajan, *Anatomy of Prepayments: The Salamon Brothers Prepayment Model*, Salomon Brothers, 1995.

Hayre, L., S. Chaudhary and Robert A. Young, Anatomy of Prepayments, *Journal of Fixed Income*, 2000, 10:1, 19–49.

Kau, J., D. Keenan, W. Muller II and J. Epperson, Pricing Commercial Mortgages and Their Mortgage-Backed Securities, *Journal of Real Estate Finance and Economics*, 1990, 3, 333–56.

——., A Generalized Valuation Model for Fixed Rate Residential Mortgages, *Journal of Money, Credit and Banking*, 1992, 24:3, 279–99.

LaCour-Little, M., Another Look at the Role of Borrower Characteristics in Predicting Mortgage Prepayments, *Journal of Housing Research*, 1999, 10, 45–60.

Law, A. M. and W. David Kelton, *Simulation Modeling and Analysis*, New York, NY: McGraw-Hill, 1982.

Mattey, J. and N. Wallace, Housing Price Cycles and Prepayment Rates of U.S. Mortgage Pools, *Journal of Real Estate Finance and Economics*, 2001, 23:2, 161–84.

Maxam, C. and M. LaCour-Little, Applied Non-Parametric Regression Techniques: Estimating Prepayments on Fixed Rate Mortgage-Backed Securities, *Journal of Real Estate Finance and Economics*, 2001, 23:2, 139–61.

Office of Federal Housing Enterprise Oversight (OFHEO), Risk-Based Capital Regulation: Second Notice of Proposed Rulemaking, Federal Register, 1999, 64, 319.

Patruno, G. N., Mortgage Prepayments: A New Model for a New Era, *Journal of Fixed Income*, 1994, 4:3, 42–56.

Pavlov, A., Competing Risks of Mortgage Terminations: Who Refinances, Who Moves, and Who Defaults?, *Journal of Real Estate Finance and Economics*, 2001, 23:2, 185–212.

Peristiani, S., P. Bennett, G. Monsen, R. Peach and J. Raiff, Effects of Household Creditworthiness on Mortgage Refinancings, Federal Reserve Bank of New York Research Paper 9622, August 1996.

——., Effects of Household Creditworthiness on Mortgage Refinancings, *Journal of Fixed Income*, 1997, 7:3, 7–21.

Quigley, J. M., ''Interest Rate Variations, Mortgage Prepayments, and Household Mobility, *The Review of Economics and Statistic*s, 1988, 636–42.

Richard, S. and R. Roll, 1989, Prepayments on Fixed Rate Mortgage Backed Securities, *Journal of Portfolio Management*, 1989, 15:3, 73–83.

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

——., Mortgage Prepayment and Default Decisions: A Poisson Regression Approach, *Journal of the American Real Estate and Urban Economics Association*, 1993, 21:4, 431–49.

Scott, D., 1992, *Multivariate Density Estimation: Theory Practice and Visualization*, New York, NY: John Wiley and Sons, 1992.

Stanton, R., Rational Prepayment and the Valuation of Mortgage-Backed Securities, *The Review of Financial Studies*, 1995, 8, 677–708.

Michael LaCour-Little, Washington University in St. Louis and Wells Fargo Home Mortgage in Clayton, MO 63105 or michael.lacour-little@mortgage.wellsfargo.com.

Michael Marschoun, PMI Mortgage Insurance Co., San Francisco, CA 94111 or mmarschoun@pmigroup.com.

Clark L. Maxam, Montana State University, Bozeman, MT 59717 or clmaxam@montana.edu.