Another Look at the Role of Borrower Characteristics in Predicting Mortgage Prepayments

Michael LaCour-Little*

Abstract

Mortgage prepayments are made for many different reasons, and research has been hampered by the inability to distinguish among those reasons for prepayment. In this article, I report results of a loan-level model where full information is available: namely, when the borrower refinances a mortgage with the same lender. With access to detailed loan-level data, the role of a number of loan and borrower characteristics can be examined.

After eliminating borrower mobility and liquidity demand factors, inclusion of loan and borrower characteristics unambiguously increases model explanatory power. But changes in predicted prepayment probabilities are most pronounced when the prepayment option is at-the-money. When the option is deeply in- or out-of-the money, borrower characteristics have little influence. This article presents the first major empirical investigation of "pure" refinancing behavior, while explicitly incorporating current actual borrower credit score and estimated current loan-to-value ratio as potentially limiting constraints.

Keywords: mortgages; prepayment; borrower; information

Introduction

Prepayment estimation is essential in forecasting expected mortgage cash flow patterns. Accordingly, mortgage and mortgage-backed security prices are highly dependent on prepayment assumptions. Portfolio lenders, firms engaged in mortgage servicing, and investors all may incur large losses when prepayment rates exceed expectations. Yet mortgage prepayments arise from varying borrower motivations: home sale caused by relocation, refinancing to take cash for other expenditures, switching from one mortgage product type to another, and, most importantly, to obtain a lower note rate when the market level of interest rates declines. Yet most academic and Wall Street approaches to prepayment modeling ignore these distinctions. This analysis overcomes many of those difficulties and provides a timely picture of pure refinancing behavior under recent market conditions.

---

*Michael LaCour-Little is Director of Financial Research at Citicorp Mortgage, Inc., and Adjunct Professor of Real Estate Finance at Washington University in St. Louis.

The views expressed are the author's alone and not necessarily those of Citigroup or any of its affiliates. Earlier versions of this article were presented at the George Washington University, the mid-year meeting of the American Real Estate and Urban Economics Association in May 1997, and the mid-year meeting of the Eastern Finance Association in April 1998. I appreciate the comments of Mark Eppli, Tony Yezer, Jim Wallace, Jocelyn Evans, David Ling, Richard Green, Kerry Vandell, Richard Buttimer, Amy Bogdon, and the two anonymous referees.

The Theory of Mortgage Prepayment

The mortgage market has evolved considerably over the past 20 years, with the development of a wide array of alternative instruments and the expansion of the secondary market, including a variety of mortgage-based derivatives. Researchers have recognized that contingent claims methodologies can provide important insights into market workings. A mortgage loan may be viewed as a fixed-income instrument combined with American put and call options held by the borrower and written by the lender. The right to prepay the mortgage at any time is a call option at par; the ability to default on the mortgage at any time is a put option, in which the mortgage is sold to the lender for the market value of the property. Prepayment options are more likely to be exercised when interest rates fall; default options, when house prices fall.

Following Archer, Ling, and McGill (1996), I represent the total probability of mortgage termination, $\lambda_{T\,t}$, for all reasons, as:

$$\lambda_{T\,t} = \lambda_{D\,t} + (1-\lambda_{D\,t})\,[\,\lambda_{M\,t} + (1-\lambda_{M\,t})\,\lambda_{P.NMt}\,] \tag{1}$$

where $\lambda_{D\,t}$ is the probability of terminating by default at time t, $\lambda_{M\,t}$ is the probability of terminating by moving at time t, and $\lambda_{P.NMt}$ is the probability of terminating by refinancing in place (prepaying and not moving) at time t. Notice that the last term, $\lambda_{P.NMt}$, is conditional upon not moving. Equation 1 yields insights into the previously discussed problems with empirical mortgage prepayment research. Option theory can provide explanations for both $\lambda_{D\,t}$ and $\lambda_{P.NMt}$, but is silent with respect to $\lambda_{M\,t}$.[1] But what is observed using mortgage-pool data, and sometimes even with loan-level data, is $\lambda_{T\,t}$. The implicit assumption then is that both $\lambda_{D\,t}$ and $\lambda_{M\,t}$ equal zero, which is surely not the case with heterogeneous borrowers.[2]

Consider a mortgage with book value $BV(r_c,t)$, where r_c is the contract rate of interest. The value of $BV(r_c,t)$ declines over time, according to its amortization schedule, with a loan term of n. At any point in time, t, household wealth is given by

$$W_t = FA_{\,t} + (H_{\,t} - MV_t) \tag{2}$$

where $FA_{\,t}$ are other financial assets at time t, $H_{\,t}$ is house value at time t, and MV_t is the market value of the mortgage at time t. Of course, MV_t is a function of contractual loan payments, remaining loan term, n-t, and the market rate of interest, r_m.

Household wealth maximization implies a pure refinancing strategy, namely, minimizing the value of MV_t. This is accomplished by following the rule:

$$\text{Prepay when } (MV(r_m,t) \,/\, BV(r_c,t)) > 1 \tag{3}$$

[1] Although the probability of moving may depend indirectly on the value of the prepayment option; that is, when the option is deeply out-of-the-money, borrowers are "locked in" to their current house and will be less likely to move (Green and Shoven 1986; Quigley 1987). Presumably, this phenomenon would have been more evident in the early 1980s than in recent years.

[2] For simplicity of exposition and consistency with Archer, Ling, and McGill's discussion, I ignore cash-out refinancing, which alters household capital structure concurrently with changing the market value of the mortgage.

In other words, the borrower should prepay the loan as soon as the option goes into-the-money, that is, as soon as the market rate of interest declines below the contract rate on $BV(r_c,t)$. In the presence of transaction costs, TC, however, the rule becomes:

$$\text{Prepay when } (MV(r_m,t) / (BV(r_c,t) + TC)) > 1 \quad (4)$$

If borrowers are heterogeneous with respect to TC, one would expect to see the rate of prepayment first increase gradually, as market rates decline below contract rates, then increase sharply, as the quantity $MV(r_m,t) - BV(r_c,t)$ comes to dominate the cost of refinancing, TC; equivalently, as the ratio $MV(r_m,t) / BV(r_c,t)$ increases above $1 + TC$. Accordingly, the prepayment function is expected to be highly nonlinear.

More sophisticated option pricing approaches argue that the cost of exercising the prepayment option today involves an additional cost as well: the loss of the right to exercise the call option in the future. This argument is mitigated by the fact that upon refinancing with a new mortgage, the borrower obtains another immediately exercisable prepayment option. One might argue that the cost of exercising now is the value of the currently held option less the value of the option to be obtained. For simplicity, I assume that the values of these two options are essentially the same, so the pure refinancing strategy is an appropriate household wealth-maximization strategy, constrained only by transaction costs. The latter are broadly defined to include not only explicit fees and costs associated with obtaining a new loan but search costs, barriers caused by credit problems, and so forth.

Following Richard and Roll (1989), I use the ratio of the contract rate, r_c / r_m as a convenient proxy for $MV(r_m,t) / BV(r_c,t)$, particularly since loans in our sample are quite new (all are 30-year mortgages, with none older than 6 years at the beginning of the study period).

Literature Review

A number of approaches to prepayment modeling have been taken. Archer and Ling (1993) broadly categorize work into two types: those focusing on call-motivated, or endogenous, prepayment versus those using a more purely empirical approach to study non-call-motivated, or exogenous, prepayments. The former, more theoretical work, focuses on valuation and optimal exercise of the embedded call option to prepay as rates fluctuate over time (Deng 1997; Follain, Scott, and Yang 1992; Kau and Kim 1992; Kau et al.1992; and McConnell and Singh 1994). Since this dimension of the literature is fairly well developed, I do not focus on it here.

Empirical research has generally employed publicly available agency mortgage-pool data, in which information on the underlying loans is limited to pool type, issuer, time, and weighted average note rate. Peters, Pinkus, and Askin (1984), Richard and Roll (1989), Schwartz and Torous (1989, 1992), Foster and Van Order (1990), and Quigley and Van Order (1990, 1995) address single-family residential mortgage prepayments in this fashion. Such research cannot distinguish defaults from prepayments nor can it distinguish prepayments that occur because the borrower relocated, refinanced to take cash out of the property for other expenditures, or refinanced consumer debt.[3] Prepayments are simply assumed to be

[3] When using mortgage-pool data, defaults are somewhat euphemistically referred to as "involuntary" prepayments.

mainly determined by interest rate movements. Not surprisingly, models using such data cannot explain observed phenomena very well.

Because of limitations in data, fewer researchers have studied prepayment at the loan level, although the direction of research over the last decade is clearly in that direction. In an influential early work, Green and Shoven (1986, 43) emphasize the role of borrower characteristics, arguing that it is "important to recognize that the primary determinants of the decision to sell a house are not related to interest rate fluctuations. They are largely concerned with the personal circumstances of the owner." Thus, Green and Shoven recognize that observed prepayments at aggregate levels include the effect of household mobility decisions and acknowledge that such decisions are largely independent of the interest rate environment. In contrast, Quigley (1987) finds household mobility reduced by the lock-in effect of favorable mortgage rates.[4]

Since option-theoretic work tends to overpredict actual empirical prepayments, much recent work has focused on transaction costs and other barriers to refinancing. Peristiani et al. (1997) explicitly examine the role of borrower credit characteristics in mortgage refinancing, using data from a sample of major metropolitan areas supplemented by TRW Information Services credit history variables. They find that measures of weak borrower credit are consistently associated with a lower probability of refinancing. They also find that lower levels of contemporaneous borrower equity (higher loan-to-value [LTV]) are similarly associated with lower probabilities of refinancing. Using the same data set, Bennett, Peach, and Peristiani (1998) argue that technological change and a shift from traditional portfolio lenders to mortgage bankers have caused transaction costs to fall over time, effectively increasing the propensity of borrowers to prepay.

Archer, Ling, and McGill (1996) and Archer and Ling (1997) extend the analysis of borrower characteristics still further. Both employ American Housing Survey data from the 1980s; the former excludes mobility-related prepayments, while the latter includes them. Both examine the role of borrower demographics, controlling for the extent to which the refinancing (call) option is in-the-money. Archer and Ling report that inclusion of borrower characteristics improves model predictive capability, mainly by indicating the extent to which the borrower may be income- or credit-constrained from refinancing. Archer, Ling, and McGill extend this analysis by examining the mobility decision that is interrelated with the refinancing decision. After controlling for the refinancing incentive (option value), they report that a range of demographic factors, including income, two-wage earner household, age, education, family size, and race have statistically significant effects on mortgage termination probability.

Green and LaCour-Little (1999) and Crawford and Wu (1998) explore the seemingly irrational under-exercise of the mortgage prepayment option. Green and LaCour-Little use a large data set of relatively high-rate mortgages originated during the 1980s that failed to prepay during the refinance boom of 1993. They find that estimates of contemporaneous borrower equity and proxies for local economic conditions, such as the rate of unemployment, can account for some, but not all, of the under-exercise of the prepayment option. In addition to using equity and credit constraints, Crawford and Wu (1998) attempt to control for ex-

[4] Quigley used data from the 1970s and early 1980s when mortgage rate volatility was much greater than it has been during recent times.

pected mobility by introducing an estimate of the probability of moving. Although these two most recent papers introduce important innovations, they still are unable to completely separate mobility from rate-driven prepayments and cannot control for cash-out refinancing at all.

In summary, data limitations have precluded separation of mortgage prepayments into those driven by borrower mobility versus those resulting from interest-rate driven refinancing objectives. Moreover, even when mobility reasons for prepayment can be excluded, or controlled for, refinancing to take cash out of a property to fund home improvements or other consumer expenditures remains.[5] This article remedies the situation by use of a data set that includes only borrowers who are refinancing rate and term (to alter their monthly payment schedule) with the same lender.

While this data set is unique, it is subject to certain limitations and biases. First, since it contains loans originated during the period 1992 to 1996 that were or were not refinanced during the 1997–98 period, all results are conditional on the loans having survived until December 31, 1996. Since older loans—say, those originated during 1992—would have had multiple opportunities for refinancing during the period 1993 to 1996, the performance of these loans during 1997 to 1998 may not be representative of the entire 1992 loan cohort. A second potential bias is introduced in that the loans must have been refinanced with the same lender.[6] In the mortgage industry, 10 to 20 percent customer retention is the norm, so loans that are refinanced with the same lender may not be representative of all loans refinanced. Finally, I am unable to compare the lender's pricing with other prices available in the market over the study period, 1997 to 1998. If, for example, the lender offered especially attractive rates in December 1997, that pricing advantage might have triggered many loans to refinance at that time, while at other times less attractive rates may have inhibited refinancing or caused borrowers wishing to refinance to choose another lender. Unfortunately, there is no obvious means to control for any of these various biases in the data.

Methodology

Prepayment research using loan-level data is typically based on techniques of survival analyses, which originated in biological studies of mortality and have also found frequent application in industrial engineering failure-time studies. Loans "die" prior to scheduled maturity from either default or prepayment. Kalbfleisch and Prentice (1980) and Cox and Oakes (1985) provide classic statistical treatments of the topic; Allison (1995) may be consulted for a wide range of practical examples drawn from both medicine and sociology; Kiefer (1988) provides a review of the economics literature on duration modeling. An alternative approach, where survival time to prepayment is less at issue, is to estimate binary choice models during a particular study period, as in Archer and Ling (1993), Archer, Ling, and McGill (1996,

[5] In the larger data set, from which the rate-term refinancing loans were drawn, about 25 percent of all loans involved cash out to fund home improvements, second home purchases, or other consumer expenditures. Exploring these household choices is a promising area for future research.

[6] In the mortgage industry, this is called an "on-us refi." Lenders sometimes accord such loan applications special deference, for example, by reducing documentation requirements or reducing fees. If the mortgage market is competitive and rates across lenders are essentially identical, then "on-us refi" loans may have lower transaction costs.

1997),[7] and Green and LaCour-Little (1998). Following this approach, it is possible to specify the following model describing the prepayment probability:

$$P(\text{Prepay}=1) = \Phi(r_c / r_m, \mathbf{C}, \mathbf{M}, \mathbf{B}, \mathbf{T}) \tag{5}$$

where r_c / r_m is the ratio of mortgage coupon rate to the market mortgage interest rate, **C** represents institutional constraints on credit history that may limit refinancing, **M** represents mobility, **B** represents borrower characteristics, **T** represents the transaction costs of refinancing, and $\Phi(.)$ is the probit or logit regression function. In the data set, since the borrower is not moving, **M** is not relevant. Since house price volatility may influence the value of the embedded default option in the mortgage and, therefore, the value of the prepayment option, a proxy for housing price volatility was also tested in the empirical work.

As noted previously, I follow the approach of Richard and Roll (1989) in modeling the borrower's refinancing incentive as the ratio of coupon to market rate.[8] Naturally, as the quantity r_c / r_m gets larger, the option to refinance gets deeper and deeper in-the-money and is expected to dominate other characteristics (proxies for heterogeneous transaction costs), except, perhaps, institutional constraints that prevent refinancing. In the empirical analysis presented, the benchmark Freddie Mac 30-year Primary Mortgage Market Survey (PMMS) rate is used as the measure of r_m. Since the dollar benefit from refinancing depends both on the relative decline in note rates (r_c / r_m) and the amount borrowed, the original loan amount is also included in the extended model specification.

Data

A major loan-servicing firm that prefers anonymity provided the data. The first step in data collection was to identify all loans that prepaid during the five-quarter period from January 1, 1997, to March 31, 1998. During this period, the lender began the process of capturing current credit scores for all loans in the servicing portfolio. The score captured, consistent with secondary market guidelines, is the minimum of the borrower and co-borrower's FICO score,[9] a measure that ranges in value roughly from 400–800, with higher values indicating better credit, that is, fewer reported credit problems in the credit bureau data. Paid-off loans were then matched to all new loan originations during this time period, first by primary borrower social security number and then by date. These new loans were then filtered using the "rate-term" flag, which indicates that the purpose of the (new) loan was to refinance the existing loan's note rate and term only; that is, the borrower was not taking cash out of the property or using the funds to purchase another residence (which was the case about 25 percent of the time). After matching new loans to paid loans and filtering them for purpose of loan, data for time-of-origination characteristics were captured for them, and the com-

[7] Both papers use American Housing Survey data to examine mortgage terminations during particular refinancing "windows" during the 1980s. While the technique used in this article is similar, the data are much more current.

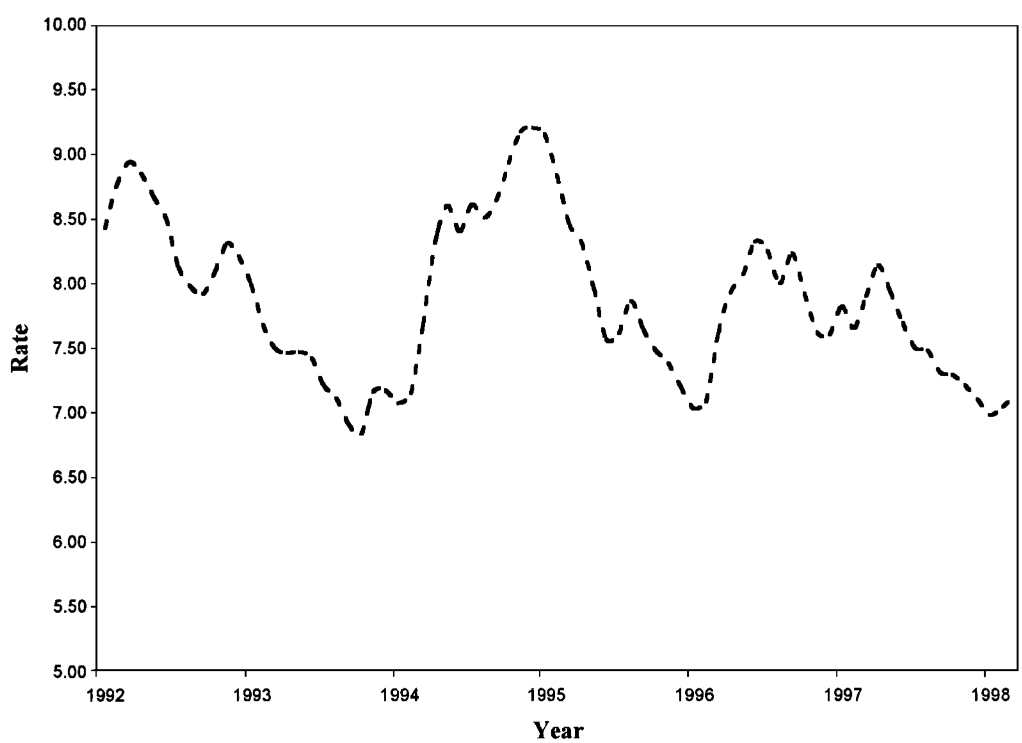
[8] Other popular measures used in the literature are the difference between coupon and market rate and the present value of the difference in payments over remaining loan term, or the ratio of the value of the mortgages at market rate versus contract rate. Richard and Roll (1989) show that for relatively new mortgages, the ratio of the contract rate to market rate is a reasonable proxy for the more cumbersome market value calculation.

[9] The FICO score is a product of Fair, Isaac, and Co., San Rafael, California, and has become a credit industry standard for evaluating borrower creditworthiness.

bined data set was merged by primary borrower social security number. Thus, in the final data set, an observation consists of a unique borrower social security number followed by data on the two loans obtained during the period 1992 to 1998, including the date of prepayment of the first loan, which coincides with the origination date of the second loan. In order to focus on borrower characteristics independently of loan type, the first loans (686 total) were restricted to 30-year fixed-rate mortgages of both conforming and nonconforming loan sizes. Of course, the borrower's second loan may be of any type. A sample of 686 nonprepaying (and nondefaulting) loans originated during 1992 to 1996 were drawn randomly from the loan servicing system to create the control group. These random draws were taken year by year, so that the distribution of loan age in the control group was approximately equal to that of the prepaying loans.

Prepayment behavior for both groups was observed during the five-quarter period from January 1, 1997, to March 31, 1998. This time period was characterized by rapidly declining interest rates, triggering a "mini refi boom" that has been compared to the refinance boom of 1993.[10] Figure 1 shows the movement of interest rates over the period 1992 to 1998.

Figure 1. **Movement of Benchmark 30-Year Mortgage Rate, 1992 to 1998**



[10] The similarity is imperfect inasmuch as 1993 was a very different point in the business cycle; the yield curve was considerably steeper, and there were large numbers of high-rate mortgages taken out in the late 1980s that were ripe for refinancing.

As described above, the combined data set consists of 1,372 loans, half prepaid and half not. Average original loan size is $143,000, but nonconforming loans are included in the data, allowing for some much larger loans. The average note rate is 8.35 percent, with average points of 0.80 percent[11]; however, since loans were originated over the period from 1992 to 1996, there is considerable variation, including low coupons from 1993 and much higher coupons from 1994, when rates exceeded 9 percent in some months. Table 1 displays descriptive statistics. Table 2 gives the geographic distribution of loans in the data set. Virtually every state is represented in the data set, with the greatest concentrations from Cali-

Table 1. **Descriptive Statistics of Sample**

	If Prepaid		If Not Prepaid	
Variables	Mean	Standard Deviation	Mean	Standard Deviation
FICO score in 1997	740	45	720	66
Borrower age	41	11	43	12
Dummy for married	0.72	0.45	0.73	0.44
Points paid	0.69	0.94	0.89	1.09
Note rate on loan	8.62	0.55	8.09	0.66
Loan size ($100,000)	1.62	1.21	1.23	0.95
Income ($1,000)	92.4	85.8	78.4	111.7
LTV	0.75	0.16	0.71	0.19
Estimated current LTV	69.0	15.1	66.7	18.5
State level housing appreciation rate	0.097	0.088	0.079	0.076
Ratio of contract rate to market rate	1.17	0.08	1.07	0.09
Dummy for Asian borrower	0.045	0.21	0.035	0.18
Dummy for Hispanic borrower	0.07	0.26	0.10	0.29
Dummy for black borrower	0.056	0.23	0.11	0.32
Dummy for loan originated in 1993	0.06	0.23	0.06	0.23
Dummy for loan originated in 1994	0.26	0.44	0.26	0.44
Dummy for loan originated in 1995	0.21	0.41	0.21	0.41
Dummy for loan originated in 1996	0.16	0.36	0.16	0.36

[11] Data on points paid at time of origination are especially difficult to obtain because of the role of third parties, such as mortgage brokers, who often receive points from loan applicants prior to the loan getting onto the books of the lender. To deal with this problem, loans originated by third parties were excluded from the data.

Table 2. **Geographic Distribution of Sample by State**

State	Percent of All Loans
California	13.0
Florida	8.7
New York	33.5
Illinois	12.7
All other states	32.1

fornia (13 percent), Florida (9 percent), Illinois (13 percent), and New York (34 percent). Discussion of the available variables follows.

Loan characteristics include interest rate, points paid, and original loan size. Origination year and month are also included, as well as payoff year and month in the case of prepaying loans. As a measure of borrower creditworthiness at the time of potential refinancing, contemporaneous credit score is captured.[12] LTV is available in two forms, that at origination and the contemporaneous estimate.[13] Borrower characteristics at the time of loan origination include household income, primary borrower age, borrower race,[14] and marital status.[15] Collateral characteristics include geographic identifier of state.[16] The borrower's refinancing incentive is measured by the ratio of note rate to the 3-month lagged market rate calculated at the time of prepayment for loans prepaying; the 15-month average Freddie Mac PMMS rate[17] of 7.57 is used for loans not prepaying. As a proxy for housing price volatility, I computed the standard deviation of the Office of Federal Housing Enterprise Oversight (OFHEO) Housing Price Index (HPI), first over the period 1992–96 and then for the period

[12] The contemporaneous score is Fair Isaac, Inc's FICO score, as marketed by Equifax, Inc. under their brand name "Beacon." Higher values of the score indicate better credit, with 660 a frequently used cutoff level for "A" paper. The score used, following secondary market guidelines, is the minimum of the borrowers' scores in the case of multiple borrowers.

[13] Contemporaneous LTV is estimated using the Office of Federal Housing Enterprise Oversight's House Price Index by state to adjust for housing appreciation. Since housing appreciation varies across states and neighborhoods, this is a very rough proxy.

[14] Borrower race is recoded into dummy variables for Asian, black, and Hispanic in the regressions.

[15] The variable for married was not used in the final model specification since there were too many missing values.

[16] New York and Florida are represented by dummy variables, since both states have mortgage recording taxes that increase transaction costs and depress refinancing activity. And a dummy is included for California, since that state is often thought to be a high prepayment state because of its high housing costs.

[17] Archer and Ling used the minimum, rather than the average, over the refinancing window but note that such an assumption may overstate the actual refinancing incentive since borrowers are unlikely to be able to time the market so closely. I use a 3-month lag to account for loan application and processing time.

1996–97.[18] In all regression results presented, the dependent variable is the binary outcome, prepayment, equal to one if the loan prepaid, and zero otherwise.

Models Estimated

Two specifications of the basic model, equation 5, are reported. The first is the base model, including only loan age (represented by dummy variables for the loan origination year and using 1992 as the holdout) and the measure for refinancing incentive. This specification is similar to that which might be used if only publicly available pool data were available. The second is the full model, including all loan, borrower, and property characteristics. To test the impact of borrower characteristics, I compare measures of model fit across the two specifications and, additionally, compare predicted probabilities at the loan level across models. Two measures of house-price volatility were tested for each model specification but were not statistically significant and did not materially affect the magnitudes of the other coefficients. Accordingly, they are not reported here but are available from the author on request.[19]

Results

The descriptive statistics in tables 1 and 2 show that, as expected, prepaid loans have higher average note rates (8.62 percent versus 8.09 percent) and larger average loan size ($162,000 versus $123,000) compared with loans not prepaid. Moreover, the borrower contemporaneous credit score is higher (740 versus 720) among prepaid loans than those not prepaid.

In table 3, results of the base logit model without borrower characteristics are shown. All variables are statistically significant at the 95–99 percent level, with refinancing incentive having the greatest effect. Pseudo R-squared is 0.32 and – 2 log likelihood is 524.8.

Table 4 extends model specification to include loan, borrower, and property characteristics. Pseudo R-squared is 0.43 and –2 log likelihood is 666.6, a significant improvement in the log likelihood ($p = 0.995$ with 13 degrees of freedom). In the base model, 84 percent of observations are correctly predicted; in the extended model, 90 percent.

Many, but not all, additional covariates included in the longer specification are statistically significant, and most are of the appropriate sign. Points paid by the borrower are not statistically significant, consistent with the view that such expenditures are sunk costs for the

[18] OFHEO publishes a quarterly index of housing price levels at the state level, the HPI, based on repeat sales transaction data obtained from Fannie Mae and Freddie Mac. OFHEO reportedly uses slightly different statistical procedures to produce the HPI, compared with the indices constructed by Fannie Mae and Freddie Mac. The principal difference is in the use of the median growth rate rather than the mean growth rate reportedly used by Fannie Mae and Freddie Mac. Since the OFHEO HPI is at the state level, this is a very rough proxy for housing price volatility at the local level.

[19] Some homeowners will not qualify for refinancing because of low house prices. The greater the volatility of house prices, the more likely it is that the appraised value will fail to meet the level required for refinancing.

Table 3. **Logit Model: Short Specification**

Variable	Estimate	Standard Error	Wald Chi-Square	Probability > Chi-Square
Intercept	−20.6	1.15	320.2	0.0001
Loan originated in 1993	2.72	0.33	69.1	0.0001
Loan originated in 1994	0.65	0.18	12.8	0.0003
Loan originated in 1995	0.38	0.19	4.03	0.044
Loan originated in 1996	0.92	0.20	20.8	0.0001
Incentive (ratio of contract rate to market rate)	17.9	0.99	323.9	0.0001

Note: −2 Log Likelihood = 524.8; Pseudo-R^2 = 0.32; Percent correctly predicted = 83.6.

borrower once paid and should not enter into the refinancing calculation.[20] Loan size is highly statistically significant, consistent with the notion that for a given decline in rates the present value benefit increases with loan amount. LTV is statistically significant in both forms, though of opposite signs; that is, a higher initial LTV increases prepayment probability, but a higher current LTV decreases prepayment probability. I speculate that higher initial LTV may be correlated with the payment burden felt by the borrower (hence, at higher values, refinancing is more attractive), while high current LTV simply reflects a collateral constraint that will make refinancing more difficult. Borrower age is negative and weakly significant, perhaps because older borrowers have shorter time horizons over which to realize the benefits of refinancing. Income is negative and statistically significant but small in magnitude. This finding is somewhat puzzling, since many argue that higher-income borrowers would tend to be more financially sophisticated, have more alternatives available to them, and hence be more likely to refinance when there are incentives to do so. On the other hand, savings from refinancing may be less valuable to a higher-income borrower relative to a lower-income borrower, especially after controlling for loan size. Finally, recall that income is measured at the time of loan origination and contains a number of missing values, so the variable may simply be mis-measured. None of the race variables is statistically significant at the 95 percent level, though the coefficient for black borrowers is negative and significant at roughly the 90 percent level, perhaps because black borrowers may be less financially sophisticated[21] or less inclined to seek financing because of perceptions of potential discrimination by mortgage lenders. The state dummy variables included in the extended model specification are all negative, with only the variable for New York statistically significant. The negative effect for New York is consistent with the state mortgage recording tax that effectively increases borrower refinancing transaction costs.

[20] Of course, points paid at origination may reduce note rate and, to the extent the note rate is lower than a similar loan where no points were paid, prepayment probability is reduced for any given future market rate. Points may be a signal by borrower of expected mobility, but research in this area is hampered by channel differences; for example, under a corporate channel, a relocating employee may have her employer pay multiple points as, in effect, an employee benefit to help compensate for the costs of relocation.

[21] This argument is less plausible, however, given that the mean income of black households in this data set is $52,000 per year.

Table 4. **Logit Model: Long Specification**

Variable	Estimate	Standard Error	Wald Chi-Square	Probability > Chi-Square
Intercept	−33.2	2.26	216.3	0.0001
Loan originated in 1993	3.08	0.43	49.9	0.0001
Loan originated in 1994	0.93	0.23	15.8	0.0001
Loan originated in 1995	0.33	0.24	1.9	0.17
Loan originated in 1996	1.12	0.28	15.6	0.0001
Incentive (ratio of contract rate to market rate)	22.9	1.39	272.6	0.0001
Points paid	0.12	0.08	1.86	0.17
Loan size ($100,000)	0.93	0.14	41.8	0.0001
LTV	0.045	0.022	4.2	0.04
Estimated current LTV	−0.049	0.023	4.3	0.04
Borrower age	−0.01	0.008	2.1	0.15
Income ($1,000)	−0.005	0.001	10.7	0.001
FICO score in 1997	0.0085	0.0016	27.6	0.0001
Black borrower	−0.54	0.33	2.6	0.10
Hispanic borrower	−0.47	0.32	2.2	0.14
Asian borrower	0.45	0.39	1.4	0.24
Located in New York	−0.38	0.22	3.1	0.08
Located in Florida	−0.50	0.32	2.4	0.12
Located in California	−0.43	0.33	1.7	0.20

Note: −2 Log Likelihood = 666.6; Pseudo-R^2 = 0.43; Percent correctly predicted = 89.8.

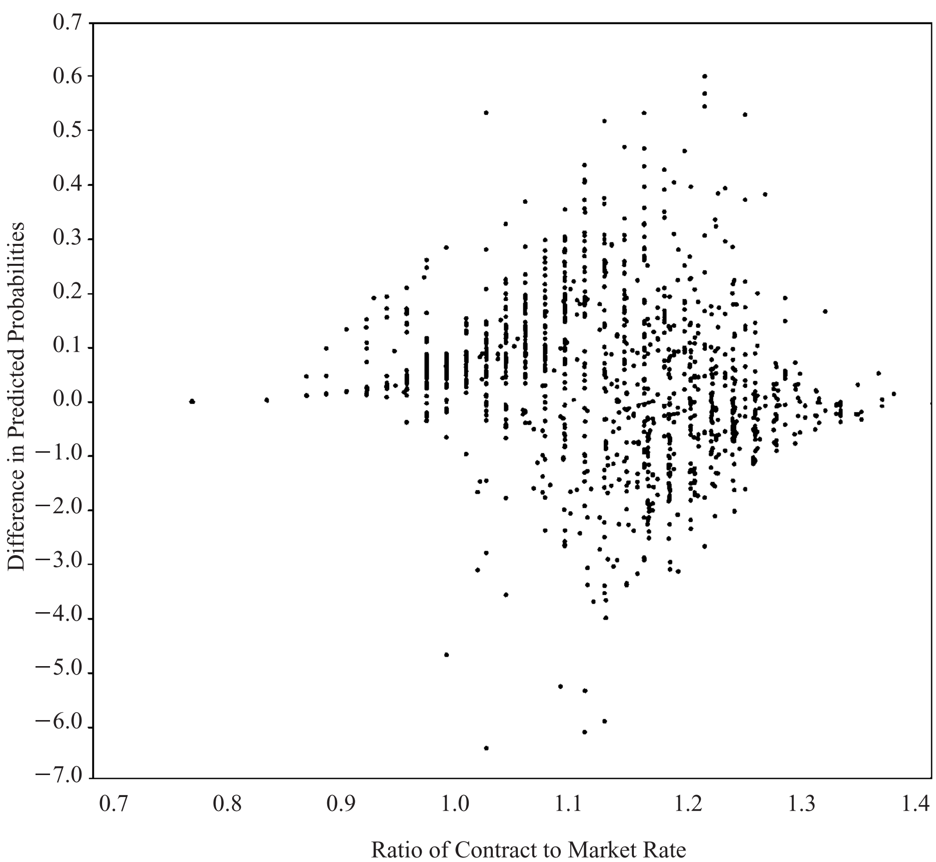
A second approach to answering the fundamental question of whether borrower characteristics matter is to compare predicted probabilities across models. To do this, predicted values are generated loan by loan and their difference compared (table 5). In general, the extended model produces *lower* predicted probabilities of prepayment for any given loan age and coupon level. In 90 percent of the cases, the extended model's prediction is lower than that of the base model. The mean difference in predicted values is 0.043; the median difference is 0.036. Figures 2 and 3 depict the difference, defined as base model minus extended model specification prediction, first by refinancing incentive and then by note rate. The elliptical shape generated suggests that at low note rates borrower characteristics alter predicted

Table 5. **Comparison of Average Predicted Probabilities**

Note Rate (percent)	Short Model (origination year and rate only)	Long Model (all loan and borrower information)
Less than 7.00	0.06	0.02
7.01–8.00	0.22	0.16
8.01–9.00	0.60	0.56
More than 9.00	0.87	0.83

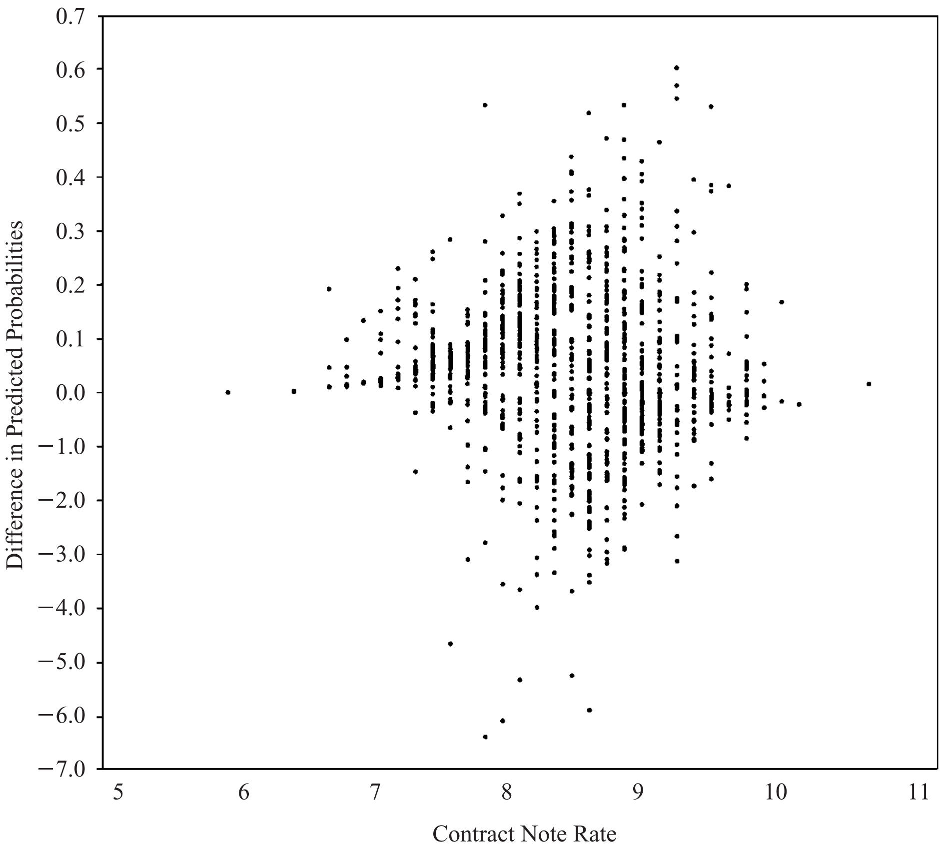
Note: Dependent variable is probability of refinancing at any time during the five-quarter study period, January 1, 1997 to March 31, 1998.

Figure 2. **Difference in Predicted Probabilities of Refinancing by Level of Incentive**



Note: This figure shows the difference in predicted probability of refinancing, defined as base model minus extended model specification predicted values, as a function of r_c/r_m. The largest differences occur at values of r_c/r_m of about 1.1; this corresponds to a borrower with an 8.25 percent note rate in a 7.5 percent market, for example.

Figure 3. **Difference in Predicted Probabilities of Refinancing by Level of Contract Note Rate**



Note: This figure shows the same difference in predicted probabilities as figure 2, but as a function of the level of r_c only. Both figures suggest that borrowers with note rates around 8.5 percent appear to have the greatest heterogeneity in probability of refinancing.

prepayment probability only slightly; likewise, at high note rates, borrower characteristics have little effect. In the mid-range (the area where the prepayment option is only slightly in- or out-of-the-money), however, inclusion of borrower characteristics can dramatically alter predicted prepayment probability.

Conclusions

Previous research on mortgage prepayments has been hampered by data limitations, particularly the inability to distinguish the varied sources of prepayment: borrower mobility, liquidity demand, and interest-rate driven rate-term refinancing. Moreover, proxies for the credit and collateral constraints have been imperfect. This study has almost entirely remedied these problems, accomplishing the first major empirical analysis of pure refinancing behavior, while explicitly modeling both credit and collateral constraints. When non–interest-rate driven prepayments are excluded from the data, the role of borrower and loan characteristics can be seen much more clearly. Results indicate that borrower characteristics do affect mortgage prepayment risk, though primarily in the region where the prepayment option is at-the-money. When the prepayment option is substantially in- or out-of-the-money, borrower and loan characteristics are largely irrelevant. Heterogeneity in borrower transaction costs, broadly defined, is one possible explanation; heterogeneity in borrower expected tenure is another. Future research should further explore these subtleties.

References

Allison, Paul D. 1995. Survival Analysis Using the SAS System. SAS Institute.

Archer, Wayne, and David Ling. 1993. Pricing Mortgage-Backed Securities: Integrating Optimal Call and Empirical Models of Prepayment. *Journal of the American Real Estate and Urban Economics Association* 21(4):373–404.

Archer, Wayne, David Ling, and Gary McGill. 1997. Demographic Versus Option-Drive Mortgage Terminations. *Journal of Housing Economics* V(4):1051–377.

Archer, Wayne, David Ling, and Gary McGill. 1996. The Effect of Income and Collateral Constraints on Residential Mortgage Terminations. *Regional Science and Urban Economics* 26:235–61.

Bennett, Paul, Richard Peach, and Stavros Peristiani. 1998. Structural Change in the Mortgage Market and the Propensity to Refinance, Federal Reserve Bank of New York Staff Reports No. 45 (September).

Cox, David, and D. Oakes. 1985. *Analysis of Survival Data*. New York: Chapman & Hall.

Crawford, Gordon, and Hsiu-Wen Wu. 1998. On the Rationality of Non-refinancing Behavior. Paper read at the annual meeting of the American Real Estate and Urban Economics Association. Chicago.

Deng, Yongheng. 1997. Mortgage Termination: An Empirical Hazard Model with a Stochastic Term Structure. *Journal of Real Estate Finance and Economics* 14(3):310–31.

Follain, James R., Louis O. Scott, and Tyler Yang. 1992. Microfoundations of a Mortgage Prepayment Function. *Journal of Real Estate Finance and Economics* 5(2):197–217.

Foster, Charles, and Robert Van Order. 1990. Estimating Prepayments. *Secondary Mortgage Markets*. Winter.

Green, Jerry, and John B. Shoven. 1986. The Effects of Interest Rates on Mortgage Prepayments. *Journal of Money, Credit, and Banking* 18(1):41–50.

Green, Richard K., and Michael LaCour-Little. 1999. The Truth About Ostriches: Who Never Refinances Their Mortgage and Why They Don't. *Journal of Housing Economics,* forthcoming.

Kalbfleisch, J. D. and R. L. Prentice. 1980. The Statistical Analysis of Failure Time Data. New York: Wiley & Sons.

Kau, James B., Donald Keenan, Walter Mueller, and James Epperson. 1992. A Generalized Valuation Model for Fixed-Rate Residential Mortgages. *Journal of Money, Credit, and Banking* 24(3):279–99.

Kau, James B., and Kim Taewon. 1992. The Timing of Prepayment: A Theoretical Analysis. *Journal of Real Estate Finance and Economics* 7(3):221–28.

Kiefer, N. 1988. Economic Duration Data and Hazard Functions. *Journal of Economic Literature* 26:646–79.

McConnell, John J., and Manoj Singh. 1994. Rational Prepayments and the Valuation of Collateralized Mortgage Obligations. *Journal of Finance* 49(3):891–921.

Peristiani, Stavros, Paul Bennett, Gordon Monsen, Richard Peach, and Jonathan Raiff. 1997. Effects of Household Creditworthiness on Mortgage Refinancings. *Federal Reserve Bank of New York Economic Policy Review* (July), 83–103.

Peters, Helen F., Scott M. Pincus, and David J. Askin. 1984. Prepayment Patterns of Conventional Mortgages: Experience from the Freddie Mac Portfolio. *Secondary Mortgage Markets* (February).

Quigley, John M. 1987. Interest Rate Variations, Mortgage Prepayments, and Household Mobility. *Review of Economics and Statistics* 69:636–42.

Quigley, John, and Robert Van Order. 1990. Efficiency in the Mortgage Market: The Borrower's Perspective. *Journal of the American Real Estate and Urban Economics Association* 18(3):237–52.

Quigley, John, and Robert Van Order. 1995. Explicit Tests of Contingent Claims Models of Mortgage Default. *Journal of Real Estate Finance and Economics* 11(2):99–117.

Richard, Scott F., and Richard Roll. 1989. Prepayments on Fixed Rate Mortgage-Backed Securities. *Journal of Portfolio Management* (spring):73–82.

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

Schwartz, Eduardo S., and Walter N. Torous. 1992. An Empirical Investigation of Mortgage Prepayment and Default Decisions. Finance working paper. University of California at Los Angeles.