Forthcoming: Journal of Real
Estate Finance and Economics

Unobservable Heterogeneity and Rational Learning: Pool Specific vs. Generic Mortgage-Backed Security Prices.

*Richard Stanton**

Haas School of Business
U.C. Berkeley
350 Barrows Hall
Berkeley, CA 94720
tel. (510) 642-7382
email: stanton@haas.berkeley.edu

This Draft: March 10, 1995

ABSTRACT

Previous mortgage prepayment and valuation models assume that two mortgage pools with the same observable characteristics should behave indistinguishably. However, even pools with apparently identical characteristics often exhibit markedly different prepayment behavior. The sources of this heterogeneity may be unobservable, but we can infer information about a pool from its prepayment behavior over time. This paper develops a methodology for using this information to calculate pool-specific mortgage-backed security prices. Knowledge of these prices is important both for portfolio valuation and for determining the cheapest pool to deliver when selling mortgage-backed securities. We find that unobservable heterogeneity between mortgage pools is statistically significant, and that pool-specific prices, calculated for a sample of outwardly identical mortgage pools between 1983 and 1989, may differ greatly from any single representative price.

---

*This paper is based on a chapter from my Ph.D. dissertation. I am grateful for helpful comments and suggestions from Jonathan Berk, Darrell Duffie, James Kau, David Modest, Chester Spatt, Nancy Wallace, Robert Whitelaw, Nicky King, two anonymous referees, and seminar participants at U.C. Berkeley, the 1993 meeting of the American Real Estate and Urban Economics Association, and the 1993 meeting of the Western Finance Association. I especially wish to thank the members of my dissertation committee, Allan Kleidon, Ken Singleton and Peter Reiss, for their enormous contributions. I thank Goldman, Sachs & Co. for providing the data used in this study, and I gratefully acknowledge financial assistance from the Berkeley Program in Finance, from the Dean Witter Foundation, and from the U.C. Berkeley Center for Real Estate and Urban Economics.

Most quoted mortgage-backed security prices are "generic", rather than pool-specific. A contract to deliver a certain face value of mortgage-backed securities on a particular date allows the seller wide latitude in choosing which pools are actually delivered. This gives him or her the option to deliver the least valuable pools. Generic prices are the prices at which these contracts are written, presumably taking into account the existence of this delivery option.[1] However, to determine the optimal pool(s) to deliver, while satisfying the terms of the contract, the seller needs to know the separate value of each deliverable pool. Similarly, investors in mortgage-backed securities need to know the values of the specific pools in their portfolios. It is thus important to be able to calculate a pool-specific price, the price of a mortgage-backed security backed by one specific mortgage pool.

Many researchers have developed models to predict the prepayment behavior of pools of fixed rate mortgages, and to value mortgage-backed securities. Dunn and McConnell (1981a, 1981b) model mortgage holders as rational holders of prepayment options, who refinance immediately when interest rates fall sufficiently. However, this model has two counterfactual implications. First, the price of a mortgage-backed securities can never exceed par. Second, since all mortgage holders are identical, they will all refinance simultaneously the instant interest rates fall below a certain critical value. To allow mortgage prices to exceed par, Dunn and Spatt (1986), Johnston and Van Drunen (1988) and Timmis (1985) add transaction costs payable on refinancing. With heterogeneous transaction costs, this also means that different classes of borrower will refinance at different times, though all members of each class will still refinance simultaneously. To relax this, Stanton (1994) restricts mortgage holders to making refinancing decisions only at discrete intervals, leading to prepayment behavior that exhibits the "burnout" characteristic of observed prepayment behavior.[2] Other authors developing rational mortgage pricing models include Kau, Keenan, Muller and Epperson (1992), who study the optimal exercise of the mortgage holder's right to default as well as prepay his or her mortgage. Examples of an alternative approach to prepayment modeling are the purely empirical models of Schwartz and Torous (1989), and many Wall Street firms. In these models, prepayment is fitted as a function of some set of (non-model based) explanatory variables, often including past

[1] There are also other options. For example, the seller may actually deliver any amount within a 5% range surrounding the agreed principal amount. See Bartlett (1989) for details.

[2] Burnout refers to the dependence of a pool's prepayment behavior on cumulative historical prepayment levels. The higher the fraction of the pool that has already prepaid, the less likely are those remaining in the pool to prepay this period, all else being equal. See Richard and Roll (1989) for a discussion.

prepayment or other endogenous variables. While there are many academic mortgage prepayment and valuation models, one feature which they all share is that they implicitly assume that two mortgage pools characterized by the same coupon rate, time to maturity, and other observable characteristics, should prepay at exactly the same rate. However, in reality, one mortgage pool often prepays at a very different rate from another, even after controlling for observable differences. This in turn leads to differences in the values of mortgage-backed securities backed by the pools.

The idea that mortgages are heterogeneous is not a new one. For example, the usual explanation for burnout is that mortgage holders are heterogeneous in their speeds of prepayment. As a result, if a large fraction of the pool has already prepaid, those remaining are likely to be slow prepayers.[3] However, previous academic models only address heterogeneity *within* a pool. In determining a pool-specific mortgage-backed security price, we are concerned instead with heterogeneity *between* pools. While these two types of heterogeneity are conceptually similar, they have different implications for pricing. With regard to heterogeneity within a pool, we care only about the overall distribution of the individual mortgage holders in that pool, since security holders receive a fraction of the cash flows from all the individual mortgage holders. In determining a pool-specific security price, taking into account heterogeneity between pools, we care not about the overall distribution of all pools, but about where within that distribution lies the specific pool we are pricing.

This paper extends previous research in two main directions. First, it presents a framework that adapts existing parametric prepayment models to handle heterogeneity between mortgage pools. Estimating such extended models using observed prepayment data yields an initial distribution of possible pool types from which each pool is assumed to be randomly drawn. Second, the paper develops an algorithm for using the information contained in monthly prepayment rates to calculate pool-specific mortgage backed security prices. As each monthly prepayment rate for a pool is observed, we perform a Bayesian update of the distribution from which that pool is drawn. This new distribution, combined with a valuation model, is used to calculate a pool-specific mortgage backed security price. Prepayment data on a sample of more than 1,000 GNMA mortgage pools over a 6 1/2 year period show that unobservable heterogeneity between pools is statistically significant. More important, implementation of the valuation algorithm shows that the value of securities backed by different, but outwardly identical, pools can differ widely. Failure to take this into account can lead to large pricing and hedging errors.

[3] This is modeled explicitly in Stanton (1994).

1 Mortgage Pool Heterogeneity

While many mortgage prepayment models assume that individual mortgages within a single pool may differ in their prepayment behavior,[4] no previous research has allowed for possible unobservable heterogeneity *between* mortgage pools characterized by the same coupon rate, issue date and other observable characteristics. However, not only do individual mortgages differ, but so do pools of mortgages. One major reason often cited for mortgage heterogeneity is geographical diversity.[5] Several Wall Street firms try to collect information about the location of the mortgages in a pool, and use this in their prepayment and valuation models. However, they cannot obtain perfect geographic information. For example, while they may know the servicing institution, this is often a national institution, and may in any case not be the original issuer of the mortgages. Besides, differences between pools will exist even after controlling for observable geographic variables, such as state of issue. Location within a state may be important; education levels may vary, which should have an impact on how ready mortgage-holders are to prepay should it be optimal to do so; mortgage-holders in two pools may have very different propensities to relocate for job-related reasons. More generally, without knowing everything about every individual mortgage holder in every mortgage pool, even after controlling for every observable factor, there will always be some *unobservable* heterogeneity between pools.

To give an indication of the significance of heterogeneity between apparently identical mortgage pools, figure 1 shows pool factors[6] for five 12.5% GNMA-I mortgage pools between January 1983 and December 1989. Apart from different pool numbers, these pools have identical observable characteristics. Each contains mortgages issued in the same month (January 1983), with the same interest rate (12.5%), and of the same type (30 year single family mortgages). However, the pools exhibit very different prepayment behavior.

[4] Usually captured via a "burnout" factor.

[5] For example, Becketti and Morris (1990) present evidence that prepayment behavior varies by state.

[6] A pool factor is the ratio of the current remaining principal in a pool to its initial principal. It is *not* adjusted for amortization, so even with no unscheduled prepayment it drops to zero over the life of the mortgage.

1.1 Modeling Prepayment with Heterogeneous Mortgage Pools

1.1.1 Parametric Prepayment Models

By a parametric prepayment model we mean any model which predicts the prepayment behavior of a mortgage, or pool of mortgages, over the next period in terms of some set of (possibly endogenous) state variables, and a set of parameters. Write $f_\pi(\pi \mid I_t; \boldsymbol{\theta})$ for the probability density function for $\boldsymbol{\pi}_t$, the proportion of a pool which prepays during month t. This is a function of I_t, the "information set" of all relevant variables known at the start of month t. It also depends on a vector of parameters, $\boldsymbol{\theta}$. The expected prepayment level next period (the usual point estimate) is then

$$\overline{\pi}_t \equiv E[\boldsymbol{\pi}_t \mid I_t] = \int_0^1 \pi\, f_\pi(\pi \mid I_t; \boldsymbol{\theta})\, d\pi. \tag{1}$$

Almost every mortgage prepayment model used in practice has a reduced form representation of this type. For example, it includes the many "hazard rate" models of mortgage prepayment, such as Schwartz and Torous (1989). This model expresses the likelihood of prepayment over the next instant via a hazard function, $\pi(t; \underline{v}, \underline{\theta})$. Here, $\underline{v}$ is a vector of explanatory variables, and $\underline{\theta}$ is a vector of parameters. Informally, the probability of prepayment in a time interval of length δt, conditional on not having prepaid already, is approximately $\pi\, \delta t$. Each hazard function corresponds to a unique likelihood function, and vice versa.[7] An example of a model which does not at first seem to be of this form, but which nevertheless possesses a reduced form representation of this type, is the rational prepayment model of Stanton (1994).

1.1.2 Extension to Heterogeneous Mortgage Pools

Continuing to assume that prepayment for each pool is described by a parametric prepayment model with associated probability density function f_π, we now introduce (unobservable) heterogeneity between pools. Assume that each pool has its own set of parameters, say $\boldsymbol{\theta}_i$ for pool i, but that we cannot ex ante identify which pool is which. The probability density function for $\boldsymbol{\pi}_{it}$, the proportion of pool i which prepays during month t, is

$$f_\pi^i(\pi) \equiv f_\pi(\pi \mid I_t; \boldsymbol{\theta}_i), \tag{2}$$

[7] For more details on hazard functions, see Cox and Oakes (1984), or Kalbfleisch and Prentice (1980).

and the expected prepayment level in period t is

$$\overline{\pi}_{it} \equiv E[\boldsymbol{\pi}_{it} \mid I_t] = \int_0^1 \pi \, f_\pi(\pi \mid I_t; \boldsymbol{\theta}_i) \, d\pi. \tag{3}$$

One way to think of this is as a separate model for each pool. We can estimate one set of parameters per pool in exactly the same way as we would estimate a model with a single set of parameters governing all pools. The only difference is that more parameters need to be estimated. In addition, we shall usually want to impose some cross-pool restrictions on the parameters $\boldsymbol{\theta}_i$. For example, one natural restriction is to assume that, if $\boldsymbol{\theta}_i$ is a vector of parameters, all but one of the parameters are the same for all pools ("common" parameters), and that just one parameter is "pool-specific". There is no theoretical problem with estimating the model in the presence of such restrictions. However, there are potential computational difficulties as the dimensionality of the problem gets large. This is discussed further in section 3.

Rather than yielding a single set of parameter values that hold for every pool, estimating the extended model described here yields a single set of common parameter values, and one set of pool-specific parameters per pool used in the estimation process.

1.2 Valuation with Unobservable Heterogeneity

If we knew in advance the value of the parameters $\boldsymbol{\theta}_i$ for a given pool, valuation would be the same as for a model with no pool-specific parameters. We would insert the parameter vector $\boldsymbol{\theta}_i$ into the appropriate parametric prepayment model, and this would give us prepayment predictions for pool i, which could then be used to price the pool. However, in practice we cannot do this. We do *not* know the true value of θ_i. We know only the distribution from which $\boldsymbol{\theta}_i$ is drawn.[8] Valuation given a known distribution of parameter values can be accomplished via a slight extension of the usual valuation approach. Suppose, at time t, we know that the prepayment parameters for mortgage pool i are drawn from some distribution $f^i_{\theta,t}$.[9] Write $V_{\theta,t}$ for the value of a mortgage-backed security backed by a pool with *known* prepayment parameter(s) $\boldsymbol{\theta}$.[10] Then, if we assume risk neutrality with respect to our uncertainty over $\boldsymbol{\theta}_i$,[11] the value of the pool is just the expected value of $V_{\theta,t}$ taken over all possible

[8] In practice, this distribution would be estimated as described above, using historical prepayment data.

[9] For notational simplicity, we shall assume that this is a univariate distribution. However, there is no theoretical reason why the distribution should not be over several parameters.

[10] Calculated by inserting the value $\boldsymbol{\theta}$ into the appropriate prepayment/valuation model.

[11] This can be justified by arguing that there are many mortgage pools, so all idiosyncratic risk can be diversified away.

values of $\boldsymbol{\theta}_i$.

$$\begin{aligned} V_t^i &= E^i[V_{\theta,t}] \\ &\equiv \int_{-\infty}^{\infty} V_{\theta,t}\, f_{\theta,t}^i \, d\boldsymbol{\theta}. \end{aligned} \tag{4}$$

The initial parameter distribution, $f_{\theta,0}^i$, is estimated using historical prepayment data, as above. However, the distribution changes over time. Each period we observe a new monthly prepayment rate for mortgage pool i. This conveys further information about the likely value of the parameter vector $\boldsymbol{\theta}_i$. Intuitively, if the observed prepayment history for a pool is "relatively likely" for one parameter value, $\boldsymbol{\theta}_1$, and "relatively unlikely" for another, $\boldsymbol{\theta}_2$, we should revise our beliefs after seeing the new prepayment rate to attach a higher relative likelihood to $\boldsymbol{\theta}_1$ than to $\boldsymbol{\theta}_2$. More formally, this revision in beliefs can be expressed as a Bayesian updating rule. For month t, call the distribution of $\boldsymbol{\theta}_i$ before we observe that month's prepayment rate ($f_{\theta,t}^i$) the "prior" distribution of $\boldsymbol{\theta}$. If the distribution governing prepayment for a particular value of $\boldsymbol{\theta}_i$ is $f_\pi(\pi \mid I_t; \boldsymbol{\theta}_i)$, as above, and we observe a prepayment rate π_{it}, then by Bayes' theorem, the "posterior" distribution for $\boldsymbol{\theta}_i$, $f_{\theta,t+1}$, is given by

$$f_{\theta,t+1}(\theta_i) = f_{\theta,t}(\theta_i) \frac{f_\pi(\pi_{it} \mid I_t; \theta_i)}{\int_{\theta=-\infty}^{\infty} f_{\theta,t}(\theta)\, f_\pi(\pi_{it} \mid I_t; \theta)\, d\theta}. \tag{5}$$

Inserting this distribution into equation 4 gives a pool-specific mortgage-backed security price which takes into account both the initial distribution of prepayment parameters, and the information contained in the sequence of monthly prepayment rates for pool i.

2 Implementation

The methodology described in the previous section can be applied to any parametric prepayment and valuation model. We shall implement the procedure here, starting with the rational prepayment model developed by Stanton (1994). This model has the attractive features of retaining the optimal option exercise framework of rational models such as Dunn and McConnell (1981a, 1981b), while fitting observed prepayment behavior as well as purely empirical models such as Schwartz and Torous (1989). A summary of the main points of the model is presented here. For more details, see Stanton (1994).

2.1 Mortgage Holders' Prepayment Decisions

Assume that each mortgage holder i must pay a transaction cost equal to some constant X_i times the remaining principal balance on the mortgage, should he or she prepay early. This cost in part reflects the direct monetary costs of refinancing (points, inspections, etc.), but it also stands in for non-monetary costs associated with the difficulty and inconvenience of filling out forms, lost productivity etc.

Mortgage prepayment arises either through interest rate driven exercise of the prepayment option, or for some exogenous reason, such as forced relocation. Exogenous prepayment is governed by a hazard rate λ (in words, the probability of prepayment for exogenous reasons in a time interval of length δt, conditional on not having previously prepaid, is approximately $\lambda\,\delta t$). Mortgage holders follow the prepayment strategy which minimizes the value of their mortgage liabilities, subject to the transaction cost X_i, and to being able to make prepayment decisions only at random discrete intervals,[12] governed by the constant hazard rate ρ. In deciding when to prepay, mortgage holders need to take into account possible movements in interest rates. We shall use the Cox, Ingersoll and Ross (1985) one factor interest rate model, in which the instantaneous risk-free interest rate satisfies the stochastic differential equation,

$$dr_t = \kappa(\mu - r_t)\,dt + \sigma\sqrt{r_t}\,dz_t. \tag{6}$$

The long run mean interest rate is μ, with speed of convergence governed by κ. The volatility of interest rates is $\sigma\sqrt{r_t}$. One further parameter, q, summarizing risk preferences of the representative individual, is needed to price interest rate dependent assets. We use the parameter values estimated by Pearson and Sun (1989).

$$\begin{aligned}
\kappa &= 0.29368, \\
\mu &= 0.07935, \\
\sigma &= 0.11425, \\
q &= -0.12165.
\end{aligned}$$

In this setting, each mortgage holder's optimal exercise strategy takes a simple form. For a given coupon rate, parameters λ and ρ, and transaction cost X_i, there is a critical interest rate, r^*_{it}, such that if $r_t \leq r^*_{it}$ mortgage holder i optimally refinances at time t; if $r_t > r^*_{it}$, mortgage holder i optimally chooses not to refinance.[13] The

[12] This would result, for example, from mortgage holders facing a fixed cost associated with the time and difficulty of making each prepayment decision.

[13] The optimal prepayment strategy for each mortgage holder is yielded as a by-product of the

higher the transaction cost X_i, the lower r^*_{it}, since the benefits to refinancing must be greater before they offset the costs incurred. For each mortgage holder, this leads to a hazard function governing prepayment, which takes on the value

$$\begin{cases} \lambda & \text{if} \quad r_t > r^*_{it}, \\ \lambda + \rho & \text{if} \quad r_t \leq r^*_{it}. \end{cases} \tag{7}$$

The hazard function governing prepayment for the pool as a whole takes on a value somewhere between λ (when r_t is so high that no mortgage holder finds it optimal to prepay) and $\lambda + \rho$ (when r_t is so low that all mortgage holders find it optimal to prepay), the exact value depending on the proportion of mortgage holders in the pool for whom $r_t \leq r^*_{it}$, which depends in turn on the value of r_t and on the distribution of mortgages within the pool. Assume the initial distribution of transaction costs in a pool is a beta distribution, with parameters α and β.[14] This implies that mortgage holders face costs between 0 and 100% of their remaining principal balance, with a mean of $\alpha/(\alpha + \beta)$.

2.2 Borrower Heterogeneity and Mortgage Pools

We have described the main features of the prepayment model of Stanton (1994). We now extend the model to introduce heterogeneity between apparently identical pools. Maintaining the assumption of a beta distribution of costs within each pool, we shall allow the parameters of the distribution to differ across pools. One way to achieve this would be to allow the mean of the distribution to vary across pools, keeping the variance fixed. However, for a given variance, the mean of a beta distribution may take on only a restricted range of possible values. We shall instead assume that all pools share a common parameter α, but that each pool i has a parameter β_i which is specific to that pool. This allows the mean of the transaction cost distribution to take on its full range of possible values, as β_i ranges from 0 (all mortgage holders face transaction costs equal to 100% of the remaining principal balance, and all interest rate driven prepayment is effectively precluded) to ∞ (no mortgage holders face any transaction costs, so they can efficiently exercise their prepayment options).

The prepayment behavior of each pool is thus determined by the values of four parameters. Three of these are common across pools. They are α (one of the two parameters governing the initial distribution of transaction costs in the pool), ρ (gov-

usual binomial tree or finite difference algorithms used to value the mortgages.

[14]The beta distribution has the desirable feature that it can take on a wide variety of different shapes, yet is fully described by the values of only two parameters.

erning how long mortgage holders wait, on average, between successive prepayment decisions), and λ (governing how likely mortgage holders are to prepay for non-interest rate related reasons). The fourth is the pool-specific parameter, β_i (which, together with α, determines the initial distribution of transaction costs in the pool).

3 Estimation of the Model

Estimating a prepayment model consists of finding parameter values that predict prepayment behavior that is close to observed behavior. To do this, we need to be able to determine the optimal exercise strategies for mortgage holders with different transaction costs, for given values of the parameters λ and ρ. We use the fact that, under the Cox, Ingersoll and Ross interest rate model, the value $V(r,t)$ of an interest rate contingent claim paying coupons/dividends at rate $C(r,t)$ is the solution to the partial differential equation

$$\frac{1}{2}\sigma^2 r V_{rr} + [\kappa\mu - (\kappa+q)r]\, V_r + V_t - rV + C = 0, \tag{8}$$

subject to appropriate boundary conditions. For any given transaction cost, we use the Crank-Nicholson finite difference algorithm[15] to solve equation 8 on a rectangular grid of possible interest rate and time values. A by-product of the valuation process is the optimal exercise strategy for the mortgage holder. Repeating this for every possible transaction cost yields a set of valuation grids, from which we can determine the value of a mortgage, and whether its holder optimally chooses to prepay, for every possible transaction cost, for every interest rate level, and for every time since the issue date. Combining this with knowledge of the initial distribution of transaction costs in the pool (determined by the parameters α and β_i), allows us to calculate the value of the pool and its aggregate prepayment behavior. This gives a set of predicted prepayment rates for each pool i (as a function of the parameters α, λ, ρ and β_i), which can be compared with the observed prepayment behavior of the pool. The empirical data used for estimation are monthly prepayment rates for 12.5% GNMA 30 year single family mortgages over the period July 1983 – December 1989, a total of 78 months. A few pools with missing data were excluded, leaving 1,156 pools in the sample used for estimation. To focus on long-term interest rate movements,[16] rather

[15] See McCracken and Dorn (1969).

[16] Since mortgages are long-term instruments, their value tends to move with long-term interest rates. The correlation between 30 year mortgage rates and the yield on 10 year Treasury securities is 0.98.

than using a short rate directly, each month the Salomon Brothers yield on newly issued 20 year Treasury bonds was used to derive a short rate series by inverting the Cox, Ingersoll and Ross bond pricing formula.

3.1 Non-linear least squares

The model's prepayment predictions can be written in the form

$$\begin{aligned}
\pi_{it} &= \overline{\pi}_{it}(\rho, \lambda, \alpha, \beta_i) + v_{it}, && (9)\\
E_t[v_{it} \mid I_t] &= 0, && (10)\\
E_t[v_{it} v_{i\tau} \mid I_t] &= 0. \quad (\tau > t) && (11)\\
E_t[v_{it} v_{jt} \mid I_t] &= 0, \quad (i \neq j), && (12)
\end{aligned}$$

for $i = 1, 2, \ldots, 1156$ (number of pools) and $t = 1, 2, \ldots, 78$ (number of months). Here π_{it} is the observed prepayment rate for pool i in month t, and $\overline{\pi}_{it}$ is the expected prepayment rate generated by the model, conditional on the history of interest rates. Write $\boldsymbol{y}$ for the vector formed by stacking the 1156 vectors π_i on top of each other, and $\boldsymbol{u}$ for the vector formed similarly from the matrix v_{it}. The new series are thus defined by

$$\begin{aligned}
y_{78(i-1)+t} &= \pi_{it}, && (13)\\
u_{78(i-1)+t} &= v_{it}. && (14)
\end{aligned}$$

All observations for a single pool appear in time order, followed by all residuals for the next pool. Thus $y_1, y_2, \ldots y_{78}$ are the 78 monthly prepayment rates for pool 1; $y_{79}, y_{80}, \ldots y_{156}$ are the 78 monthly prepayment rates for pool 2, and so on. The model can now be written in the form

$$\begin{aligned}
y_k &= \pi_k(\boldsymbol{\theta}) + u_k, && (15)\\
E[u_k] &= 0, && (16)\\
E[u_k u_l] &= 0, \quad (k \neq l), && (17)
\end{aligned}$$

for $k = 1, 2, \ldots 78 \times 1156 = 90,168$, where $\boldsymbol{\theta} = (\rho, \lambda, \alpha, \beta_1, \beta_2, \ldots, \beta_{1156})$. Given such a (non-linear) regression equation, the non-linear least squares (NLLS) estimator of the true parameter value $\boldsymbol{\theta_0}$, $\widehat{\boldsymbol{\theta}}$, is the value which minimizes the sum of squared residuals,

$$S_K(\boldsymbol{\theta}) = \sum_{k=1}^{K} u_k^2 = \sum_{k=1}^{K} [y_k - \pi_k(\boldsymbol{\theta})]^2 . \tag{18}$$

Assuming homoscedastic residuals with constant variance σ_0^2, the NLLS estimator of σ_0^2, is

$$\widehat{\sigma}^2 = \frac{S_K(\widehat{\boldsymbol{\theta}})}{K}, \tag{19}$$

and the asymptotic variance-covariance matrix for the estimated parameter values is

$$\frac{S_K(\widehat{\boldsymbol{\theta}})}{K} \left[\sum_{k=1}^{K} \frac{\partial \pi_k}{\partial \boldsymbol{\theta}} \frac{\partial \pi_k}{\partial \boldsymbol{\theta}'} \right]^{-1} . \tag{20}$$

Estimating the model requires the minimization of $S_K(\widehat{\boldsymbol{\theta}})$, a function of 1,159 variables. This is not a theoretical problem, but it is computationally burdensome, if not infeasible, for general functions of this many variables. However, one feature of the model greatly simplifies its estimation. Looking at equations 9–12, we see that the residuals for pool i depend only on the value of β_i, and *not* on the value of β_j for any other pool. Thus we do not need to consider the interaction between changes in different β values. For given values of ρ, λ and α, changing a particular β_i will only affect the residuals for pool i, and the size and direction of this change is independent of the values of all other β parameters. Write the sum of squares as $S_K(\boldsymbol{\theta}^*, \boldsymbol{\beta})$, where

$$\begin{aligned} \boldsymbol{\theta}^* &= (\rho, \lambda, \alpha), \\ \boldsymbol{\beta} &= (\beta_1, \beta_2, \ldots, \beta_{1156}). \end{aligned}$$

Rather than minimizing S_K over 1,159 parameters simultaneously, we can minimize it in two nested steps. Define the concentrated sum of squares

$$S_K^*(\boldsymbol{\theta}^*) = S_K(\boldsymbol{\theta}^*, \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}^*)), \tag{21}$$

where $\widehat{\beta}(\boldsymbol{\theta}^*)$ is the set of β_i values which minimizes the sum of squared residuals conditional on the given values of ρ, λ, α. Choose $\widehat{\boldsymbol{\theta}}^*$ to minimize $S_K^*(\boldsymbol{\theta}^*)$.

The independence of the residuals for each pool from the β value for all other pools implies that

$$\min_{\boldsymbol{\beta}} S_K(\boldsymbol{\theta}^*, \boldsymbol{\beta}) = \sum_{i=1}^{I} \min_{\beta_i} S_i(\boldsymbol{\theta}^*, \beta_i), \tag{22}$$

where $S_i(\boldsymbol{\theta}^*, \beta_i)$ is the sum of squares for pool i only. Thus, in calculating the concentrated sum of squares for a given set of parameters ρ, λ, α, we can minimize the sum of squared residuals for each pool separately, performing a one dimensional minimization to estimate one β_i parameter at a time. We have effectively reduced the solution of a problem in 1,159 dimensions to the solution of a problem in 4 dimensions.

3.2 Estimation Results

Table 1 shows the results of NLLS estimation of the model described in section 2. Individual estimates for the parameters ρ, λ and α are shown, as well as summary statistics for the estimated distribution of β_i parameters, and some descriptive measures calculated from the estimated parameters. According to the parameter estimates obtained, mortgage holders wait an average of 1.28 years before prepaying their mortgage once it becomes optimal to do so. They have a probability of 4.37% per year of prepaying their mortgages for exogenous reasons, and face an overall average transaction cost equal to 48% of the remaining principal balance on their loans.[17] Figure 2 shows the full estimated empirical distribution of the pool specific parameter, β_i. This has a wide support, and translates into very different transaction cost distributions for different pools. While the overall average transaction cost level is 48%, the average transaction cost in a single pool ranges between 20% and 100% of the remaining principal balance. As a diagnostic check, table 2 shows autocorrelations and cross-correlations for both the original prepayment rates, and the model's prediction errors. The errors exhibit little serial correlation or contemporaneous correlation across pools.[18] As a test for homoscedasticity, table 3 shows the results of regressing the model's squared prediction errors, $\left[y_k - \pi_k(\widehat{\boldsymbol{\theta}})\right]^2$, against several likely explanatory variables. None of the estimated coefficients (except the constant term) is significant at the 1% level, and the R^2 of the regression is less than 0.0002. This is a rather unexpected result. One might expect the variance to be roughly proportional to the inverse of the number of mortgages in the pool, itself proportional to the dollar balance in the pool. This same lack of dependence on pool size was also documented by Becketti and Morris (1990) for mortgage pools underlying FNMA mortgage-backed securities.

[17] This is rather high. See Stanton (1994) for some possible explanations.

[18] Regressing the errors simultaneously against lags 1–6, plus the contemporaneous residuals from the pool with the next higher pool number, results in an R^2 of 0.0016.

3.3 Significance of Heterogeneity

To test the significance of the estimated degree of heterogeneity between outwardly identical pools, the model was estimated again with no heterogeneity between pools, i.e. subject to the constraint that

$$\beta_1 = \beta_2 = \ldots = \beta_{1,156}.$$

The estimated values and standard errors, reported in table 4, are close to those reported in Stanton (1994).[19] Figure 3 plots average monthly prepayment rates for the 1,156 mortgage pools between July 1983 and December 1989, together with average predicted prepayment rates generated by the model with and without heterogeneity.

Both models fit the average prepayment rates over this period closely. The R^2 values[20] for the two models are both over 91%, with that for the model with heterogeneity being slightly higher than that for the constrained model. However, we are not here primarily interested in the performance averaged across all pools, but rather in the performance of the two models relative to specific pools. Comparing the results in tables 1 and 4, the model with heterogeneity explains 15.6% of the variation in monthly prepayment rates across pools.[21] The restricted model, with no heterogeneity, explains only 12.6%. To test the statistical significance of this difference, we test the statistical hypothesis,

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_{1,156}. \tag{23}$$

We can construct a test statistic for this hypothesis as a generalization of the χ^2 test

[19] This restricted model is exactly equivalent to the model in Stanton (1994). The source of the slightly different parameter estimates is that we are minimizing a slightly different objective function here.

[20] Calculated from the 78 data points plotted in the graph, using the formula

$$R^2 = 1 - \frac{\text{var(Average Fitting Error)}}{\text{var(Average Prepayment Rate)}}.$$

[21] Calculated from all $1,156 \times 78$ individual pools' monthly prepayment rates, using the formula

$$R^2 = 1 - \frac{\text{var(Fitting Error)}}{\text{var(Prepayment Rate)}}.$$

used in OLS.[22] We shall use

$$(T-K)\frac{S_T(\overline{\boldsymbol{\theta}})-S_T(\widehat{\boldsymbol{\theta}})}{S_T(\widehat{\boldsymbol{\theta}})} \overset{A}{\sim} \chi_q^2, \tag{24}$$

where q is the number of restrictions, T is the total sample size, $\widehat{\boldsymbol{\theta}}$ is the unconstrained estimator, $\overline{\boldsymbol{\theta}}$ is the constrained estimator, and K is the number of parameters being estimated. The χ^2 test statistic is given by

$$\begin{aligned}\chi^2_{1155} &= [(1156\times 78)-1159]\times\frac{(142.608-137.807)}{137.807}\\ &= 3101.\end{aligned}$$

This is significant at the 1% level (critical value for χ^2_{1155} is 1270), so we can reject the hypothesis that there is no heterogeneity among the pools, though we must note that the χ^2 distribution of the test statistic is a large sample result. The next section analyzes the economic significance of this heterogeneity.

3.4 Valuation

We have seen that, after controlling for observable differences, the remaining heterogeneity between pools is statistically significant. In this section, we implement the valuation algorithm described above, to test the *economic* significance of this unobservable heterogeneity. To implement the valuation procedure, we use the values of α, ρ and λ estimated in section 3, and assume that each β_i is randomly drawn from the distribution estimated there (and plotted in figure 2).[23] We also need to know the distribution of the pool's prepayment level for each possible value of β_i, $f_\pi(\pi \mid I_t;\beta_i)$. The model we have been using would tell us this if we knew the number of mortgages in each pool. However, we do *not* know the number of mortgages in each pool, only the total principal balance. We could assume some fixed mortgage size, but the posterior distribution would depend on this assumed size. Intuitively, the larger the number of mortgages we assume are in a pool, the less likely we are to observe large deviations from the expected prepayment level each month, the less noise there is in the observed signal, and the more information we can extract from each monthly

[22]See Amemiya (1985), page 136.

[23]This distribution is estimated using the entire sample period, which includes information that was not known to market participants. However, we only use this information to estimate the model's parameters. Market participants could have estimated these same parameters using additional historical prepayment data that we do not have access to. In the updating procedure, we only use information that was available to market participants.

prepayment level.

We shall instead, therefore, use an empirically derived distribution function, assuming that the prepayment level for a pool with prepayment parameter β_i, π_t, is normally distributed, with mean

$$\mu_\pi^i = \overline{\pi}_{it}(\rho, \lambda, \alpha, \beta_i), \tag{25}$$

and variance

$$\sigma_\pi^2 = \widehat{\sigma}^2 = 1.53 \times 10^{-3}, \tag{26}$$

the estimated residual variance from table 1. So now we have

$$f_\pi(\pi_{it} \mid I_t; \rho, \lambda, \alpha, \beta_i) = N(\mu_\pi^i, \sigma_\pi^2), \tag{27}$$

where N represents the normal probability density function.

The detailed implementation of the valuation methodology described in section 1 can be summarized by the following algorithm:

1. Initialize:
   (a) Set $t = 0$.
   (b) Choose a set of values $\beta_1, \beta_2, \ldots, \beta_{n_\beta}$ and initial probabilities $p_1^0, p_2^0, \ldots, p_{n_\beta}^0$ to approximate the initial distribution of β values.[24]
   (c) Pick a set of possible transaction costs $K_1, K_2, \ldots, K_{n_k}$.
2. Value security at time t:
   (a) For each $\beta_i, i = 1, 2, \ldots, n_\beta$, calculate w_{ij}^t, the expected proportion of a pool with prepayment parameter β_i with transaction cost K_j at time t.
   (b) Calculate w_j^t, the overall expected proportion of the pool at time t with transaction cost K_j, using the equation

$$w_j^t = \sum_{i=1}^{n_\beta} p_i^t w_{ij}^t. \tag{28}$$

   (c) For each transaction cost level, calculate $V(K_j, r_t, t)$, the value at time t of a mortgage with transaction cost K_j with interest rate r_t.

[24]Using a discrete distribution makes keeping track of changes in the distribution over time merely a matter of updating the n_β probabilities $p_1^t, p_2^t, \ldots$. By making the values denser, we can approximate any arbitrary continuous distribution as closely as we like.

(d) Calculate V_t, mortgage-backed security value, as a weighted average of these values:

$$V_t = \sum_{j=1}^{n_K} V(K_j, r_t, t). \tag{29}$$

3. Update beliefs (the probabilities $p_1^t, p_2^t, \ldots, p_{n_\beta}^t$) in light of the observed prepayment rate π_t.

   (a) Calculate likelihood for each β_i,[25]

   $$\mathcal{L}(\beta_i) = e^{-\frac{1}{2}\left(\frac{\pi_t - \mu_\pi^i}{\widehat{\sigma}}\right)^2}.$$

   (b) Update probabilities $p_1^t, p_2^t, \ldots, p_{n_\beta}^t$,

   $$p_i^{t+1} = p_i^t \frac{\mathcal{L}(\beta_i)}{\sum_{\iota=1}^{n_\beta} \mathcal{L}(\beta_\iota)}$$

4. Go back to step 2

Figure 4 shows the results of implementing this valuation process for the five mortgage pools whose prepayment behavior was shown in figure 1. The line labelled "Naive" shows the value that would be obtained in the absence of unobservable heterogeneity, i.e. assuming the same parameters for all pools, and not updating our beliefs as prepayment information becomes available.[26] The graph exhibits several important features. The first is that between January 1983 (when the pools were issued) and early 1986, all prices were almost exactly the same. The updating rule made almost no difference over this period, even though figure 1 shows that the pools did prepay at different speeds. The reason for this is that interest rates remained relatively high over this period, with the result that the prepayment model predicted no prepayment for rational (interest rate driven) reasons. Any prepayment must therefore have been for exogenous reasons. This does not tell us anything about the value of β_i, which only determines how much prepayment occurs for *interest rate* reasons.

During 1986, interest rates declined, and the model started to predict prepayment for interest rate reasons. As a result, differences in observed prepayment *did* affect

[25]Ignoring the constant term, which does not affect the results.

[26]This is a price obtained using an "average" set of parameter values. Ignoring Jensen's Inequality, we can think of this as roughly the price of an "average" pool. It is important to note that this may *not* be the same as the quoted generic price, unless traders believe all pools to be homogeneous. Even ignoring other delivery options, the fact that the seller can choose which pools to deliver implies that the generic price ought to reflect an estimate of the price of some of the *worst* available pools, rather than the average. This is similar to Akerlof (1970).

the posterior distribution for β_i, in turn resulting in pool specific prices that started to diverge significantly from mid-1986 onward. By the end of 1989, the smallest pool specific price was \$128.40 per \$100 of principal, and the largest was \$136.41, a difference of \$8.01 per \$100 of remaining principal, or more than 6%.

Table 5 shows a more comprehensive summary. Both naive and pool-specific prices are calculated for all 1,156 pools, for each month that the pool was in existence between January 1980 and December 1989. Table 5 shows the distribution of the difference between each month's pool-specific price, calculated using the algorithm above, and the naive price. One striking feature of the table is that for every pool, within the first 2 years of its issue date, the pool-specific and naive prices differ by less than \$0.50. There are two reasons for this. First, with only a short series of prepayment rates, the Bayesian updating rule can have only a small effect on the distribution of the pool-specific parameter β_i for any given pool. Second, as discussed above, interest rates were relatively high prior to early 1986, leading to little predicted interest rate driven refinancing, and hence little revision in beliefs. For older pools, both because more data goes into the updating process and because interest rates dropped from 1986 on (so that more interest rate driven prepayment occurred), the difference between pool-specific and naive prices is often much larger. For pools between 2 and 5 years from issue, only 48% of prices are within \$0.50 of the naive price, and 5% of prices are more than \$5.00 higher or lower than the naive price. For pools older than 5 years, the proportion of prices within \$0.50 of the naive price drops to 22%, and the proportion more than \$5.00 higher or lower than the naive price increases to 7% of all calculated values. This shows that, when trading on a pool-specific basis, it is of vital importance to look not only at initially observable characteristics of each pool, but also to incorporate information about how the pool prepays over time. A single representative price, no matter how this is calculated, can be a long way from the true value of any individual pool.

4 Summary and Concluding Remarks

This paper shows that, even after controlling for observable differences, there still remains statistically significant heterogeneity in the prepayment behavior of different mortgage pools. It shows how existing parametric prepayment models can be adapted to handle this unobservable heterogeneity between mortgage pools with identical observable characteristics, and develops a valuation algorithm for calculating pool-specific mortgage-backed security prices.

Estimating the extended prepayment model using historical prepayment data yields an initial distribution of possible types from which each pool is drawn. Each period, a Bayesian update of this distribution is performed, to incorporate the information contained in the observed prepayment history of a pool. This distribution can then be used to calculate pool-specific mortgage-backed security prices.

Pool-specific prices calculated in this way are important for valuing portfolios of mortgage-backed securities, and for determining which pools are the cheapest to deliver when selling mortgage-backed securities. This paper shows that the values of two pools with initially identical observable characteristics often differ greatly. Ignoring pool-specific information may thus lead to large pricing and hedging errors.

The methodology developed here has application in the valuation of any asset or liability whose value depends on the average behavior of a group of individuals whose individual differences are unobservable, but whose behavior is heterogeneous. This includes bank liabilities such as Certificates of Deposit and insurance contracts such as Single Premium Deferred Annuities (whose value depends on their holders' propensities to surrender their policies when alternative investments offer a relatively higher return) and convertible bonds (whose holders do not all convert at the same time). In each case, a basic valuation and option exercise/surrender model could be combined with a distribution of possible investor types to calculate values that take into account the information contained in the observed behavior of the policy/bond holders over time to learn more about their likely behavior in future.

References

Akerlof, G. "The Market for Lemons: Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 89, 488–500 (1970).

Amemiya, T. *Advanced Econometrics*. Harvard University Press, Cambridge (1985).

Bartlett, W. *Mortgage-Backed Securities: Products, Analysis, Trading*. Prentice Hall, New Jersey (1989).

Becketti, S. and Morris, C. "The Prepayment Experience of FNMA Mortgage-Backed Securities." Working paper, New York University Salomon Center (1990).

Cox, D. R. and Oakes, D. *Analysis of Survival Data*. Chapman and Hall, New York (1984).

Cox, J. C., Ingersoll, J. E. and Ross, S. A. "A Theory of the Term Structure of Interest Rates." *Econometrica* 53, 385–467 (1985).

Dunn, K. B. and McConnell, J. J. "A Comparison of Alternative Models for Pricing GNMA Mortgage-Backed Securities." *Journal of Finance* 36, 471–483 (1981a).

Dunn, K. B. and McConnell, J. J. "Valuation of Mortgage-Backed Securities." *Journal of Finance* 36, 599–617 (1981b).

Dunn, K. B. and Spatt, C. S. "The Effect of Refinancing Costs and Market Imperfections on the Optimal Call Strategy and the Pricing of Debt Contracts." Working paper, Carnegie-Mellon University (1986).

Johnston, E. and Van Drunen, L. "Pricing Mortgage Pools with Heterogeneous Mortgagors: Empirical Evidence." Working paper, University of Utah (1988).

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

Kau, J. B., Keenan, D. C., Muller, III, W. J. and Epperson, J. F. "A Generalized Valuation Model for Fixed-Rate Residential Mortgages." *Journal of Money, Credit and Banking* 24, 279–299 (1992).

McCracken, D. and Dorn, W. *Numerical Methods and FORTRAN Programming*. John Wiley, New York (1969).

Pearson, N. D. and Sun, T. "A Test of the Cox, Ingersoll, Ross Model of the Term Structure of Interest Rates Using the Method of Maximum Likelihood."

Working paper, MIT (1989).

Richard, S. F. and Roll, R. "Prepayments on Fixed Rate Mortgage-Backed Securities." *Journal of Portfolio Management* 15, 73–82 (1989).

Schwartz, E. S. and Torous, W. N. "Prepayment and the Valuation of Mortgage-Backed Securities." *Journal of Finance* 44, 375–392 (1989).

Stanton, R. H. "Rational Prepayment and the Valuation of Mortgage-Backed Securities." Working paper, U. C. Berkeley (1994). Forthcoming, Review of Financial Studies.

Timmis, G. C. "Valuation of GNMA Mortgage-Backed Securities with Transaction Costs, Heterogeneous Households and Endogenously Generated Prepayment Rates." Working paper, Carnegie-Mellon University (1985).

Table 1: Parameter estimates for extended model

Parameter	Estimated Value (Standard Error)
ρ	0.7792 (0.0069)
λ	0.0447 (0.0003)
α	2.9605 (0.0094)
Mean estimated β_i	3.154 (0.093)
Standard deviation of β_i	1.326
Minimum estimated β_i	0.0001
Maximum estimated β_i	11.597
Sum of squared residuals, $S_K(\widehat{\boldsymbol{\theta}})$	137.807
Variance of residuals, $\widehat{\sigma}^2$	.00153
Variance of dependent variable, $\mathrm{var}(y_k)$	.00181
R^2, $1 - \widehat{\sigma}^2/\mathrm{var}(y_k)$	.156

In this table, non-linear least squares is used to estimate the three common parameters, ρ (which measures how likely mortgage holders are to make a prepayment decision in any given period), λ (which measures the probability of prepayment for exogenous (non-interest rate) reasons), and α (one determinant of the distribution of transaction costs among mortgage holders in each pool), plus 1,156 pool specific parameters β_i (which, together with α, determines the distribution of transaction costs in pool i). Data used for estimation are monthly prepayment rates for 1,156 12.5% GNMA-1 mortgage pools between July 1983 and December 1989.

Table 2: Autocorrelations and cross-correlations

Autocorrelations	**Prepayment Rates**	**Model's Residuals**
ρ_1	0.163153	0.008299
ρ_2	0.144951	-0.006560
ρ_3	0.123448	-0.020467
ρ_4	0.113847	-0.016373
ρ_5	0.102803	-0.014801
ρ_6	0.094024	-0.011551
ρ_7	0.085350	-0.009399
ρ_8	0.077198	-0.008245
ρ_9	0.060412	-0.016121
ρ_{10}	0.055121	-0.014379
ρ_{11}	0.043137	-0.018770
ρ_{12}	0.035045	-0.016000
Cross correlation, pool i, $i+1$	0.150243	0.024494
Cross correlation, pool i, $i+2$	0.139543	0.013916

First twelve rows give monthly autocorrelation coefficients for observed monthly prepayment rates (column 2), and model's prediction errors (column 3). These are calculated from the stacked series for the 1,156 pools, y_k, and u_k, as $\mathrm{corr}(y_k, y_{k-1})$, $\mathrm{corr}(y_k, y_{k-2})$ etc. The last two rows give the contemporaneous correlation between different pools, calculated as $\mathrm{corr}(y_k, y_{k-78})$ and $\mathrm{corr}(y_k, y_{k-156})$ etc.

Table 3: OLS regressions of squared prediction errors

Independent Variables	**Estimated Coefficient (t-statistic)**
Intercept	-0.00156^{***} (5.072)
INVSIZE, 1 / (1 + Remaining Principal Balance)	51.06^{**} (2.56)
TIME, Months since issue	$-1.04e-06$ (-0.528)
SHORT, One month T-Bill return	$5.51e-05^{**}$ (2.44)
LONG, 20yr T-Bond yield	$-4.59e-05$ (-1.446)
F-statistic	3.747
Durbin-Watson	1.839
R^2	0.00017

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

OLS regression of model's squared prediction errors, $\left[y_k - \pi_k(\widehat{\boldsymbol{\theta}})\right]^2$, against explanatory variables INVSIZE (1/(1+Remaining balance)), TIME (months since issue), SHORT (1 month T-Bill return) and LONG (yield on newly issued 20 year Treasury Bonds).

Table 4: Parameter estimates for restricted model

Parameter	**Estimated Value (Standard Error)**
ρ	0.6615 (0.0056)
λ	0.0336 (0.0004)
α	2.9716 (0.0073)
β	4.0237 (0.0196)
Sum of squared residuals, $S_K(\overline{\boldsymbol{\theta}})$	142.608
Variance of residuals, $\overline{\sigma}^2$	.00158
Variance of dependent variable, $\mathrm{var}(y_k)$	.00181
R^2, $1 - \overline{\sigma}^2/\mathrm{var}(y_k)$	.126

In this table, non-linear least squares is used to estimate the model subject to the restriction $\beta_1 = \beta_2 = \ldots = \beta_{1,156} = \beta$. The parameters estimated are ρ (which measures how likely mortgage holders are to make a prepayment decision in any given period), λ (which measures the probability of prepayment for exogenous (non-interest rate) reasons), α and β (the two parameters in the beta distribution which describes the distribution of transaction costs in each pool). Data used for estimation are monthly prepayment rates for 1,156 12.5% GNMA-1 mortgage pools between July 1983 and December 1989.

Table 5: Distribution of pool-specific prices

	Months Since Issue			
	0 – 12	**13 – 24**	**25 – 60**	**61+**
Mean Naive Price	108.25	103.11	120.87	129.56
σ(pool-specific - naive)	0.0002	0.0001	6.15	7.20
Proportion of prices within \$0.50	1.00	1.00	0.48	0.22
Proportion between \$0.51 and \$1.00	0.00	0.00	0.09	0.22
Proportion between \$1.01 and \$2.00	0.00	0.00	0.15	0.27
Proportion between \$2.01 and \$3.00	0.00	0.00	0.12	0.12
Proportion between \$3.01 and \$4.00	0.00	0.00	0.07	0.06
Proportion between \$4.01 and \$5.00	0.00	0.00	0.04	0.04
Proportion outside \$5.01	0.00	0.00	0.05	0.07

Distribution of pool-specific prices calculated for 1,156 12% GNMA-1 mortgage pools over the period January 1980 – December 1989. For pools of different ages, the first row shows the average "naive" price, calculated assuming all pools to be homogeneous. The second row gives the standard deviation (in \$) of the difference between the naive price and the pool-specific price for each pool. The remaining rows give the proportion of calculated pool-specific prices within different distances from the naive price.

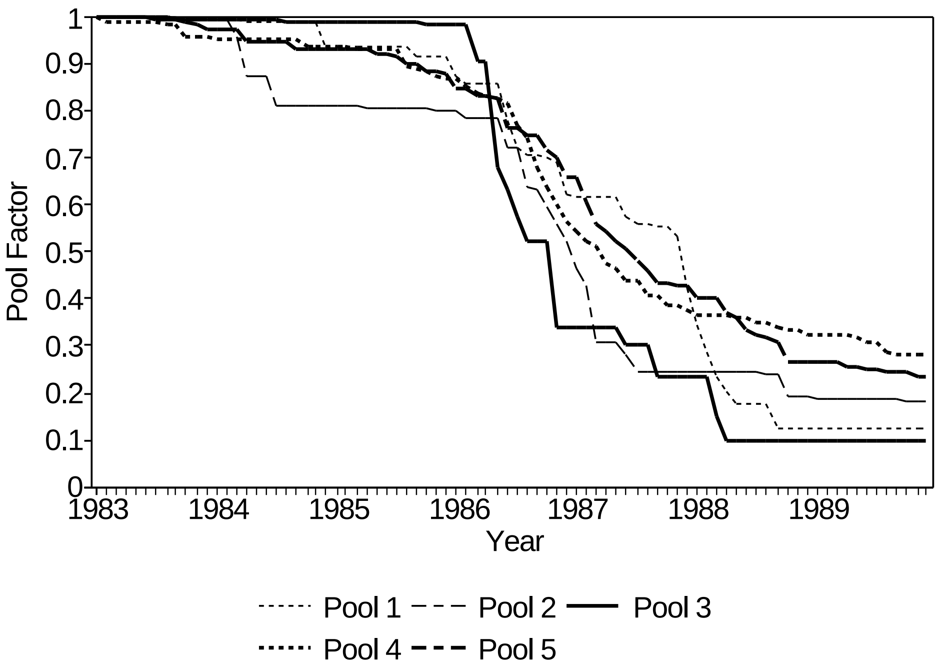

Figure 1: Monthly pool factors (proportion of initial principal still remaining, with no adjustment for amortization) for a sample of five GNMA-1 mortgage pools. Each pool contains only 12.5%, 30 year, single family mortgages. Each was issued in January 1983.

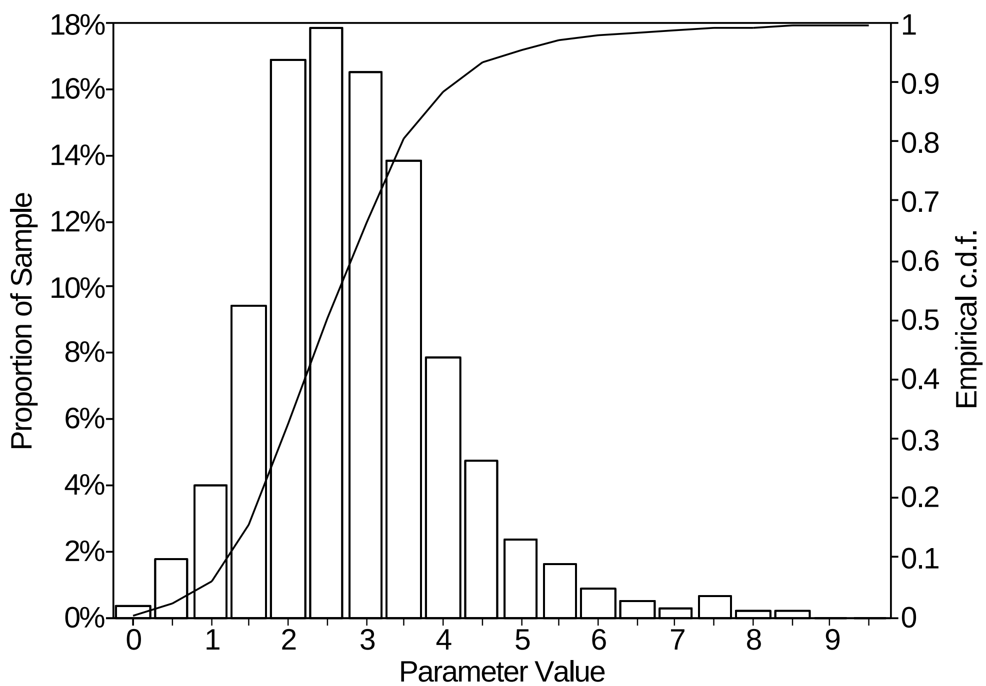

Figure 2: Empirical distribution of pool-specific parameter β_i. Histogram shows proportion of pools in sample (total 1,156) with estimated β_i falling within $\pm$ 0.25 of value on X-axis. The solid line depicts the empirical c.d.f. This parameter, together with the common parameter α, determines the initial distribution of transaction costs in pool i. The average cost level in pool i is $\alpha/(\alpha + \beta_i)$.

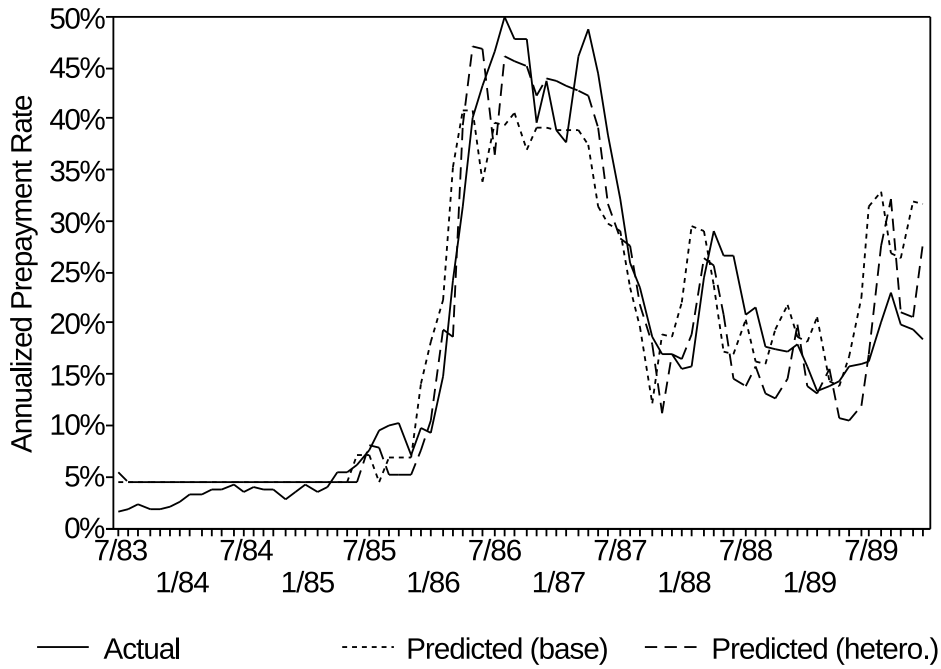

Figure 3: Average monthly prepayment rates from July 1983 to December 1989 for a sample of 1,156 12.5% GNMA-1 mortgage pools, plotted against fitted values from base prepayment model (all pools share a common β parameter), and extended model (each pool i has its own pool-specific parameter, β_i).

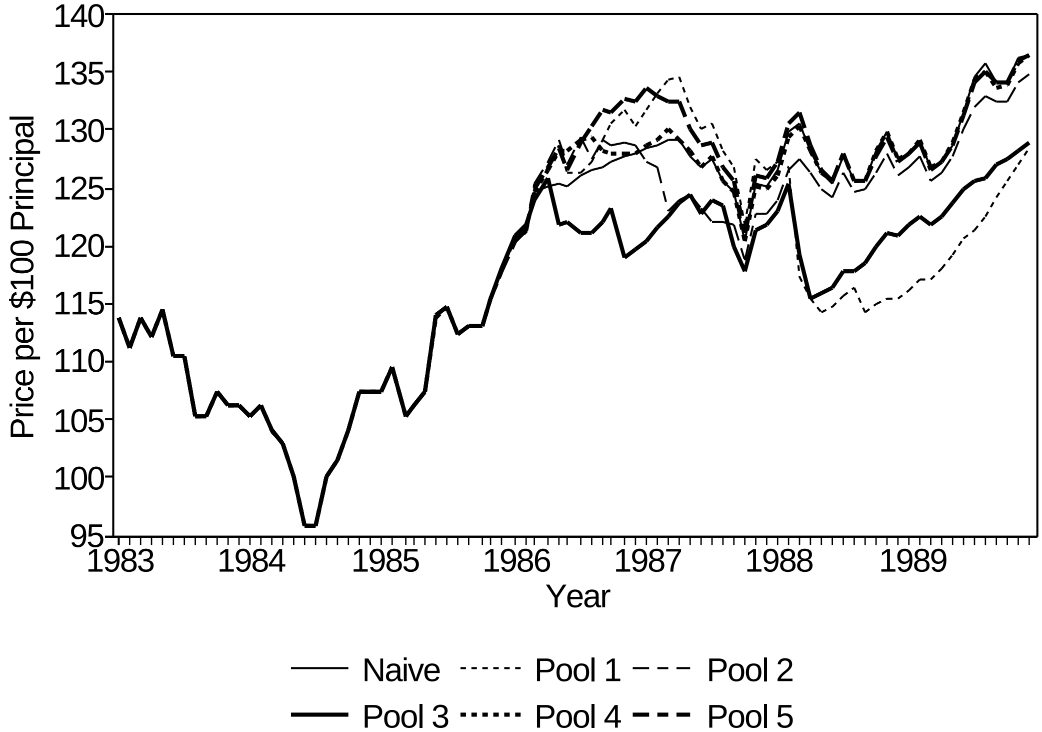

Figure 4: Pool-specific mortgage-backed security values from January 1983 to December 1989, for securities backed by five GNMA-1 mortgage pools. Each pool was issued in January 1983, and contains only 12.5%, 30 year, single family mortgages. Prices are calculated using a Bayesian updating rule to infer further information about the distribution of costs in each pool from the prepayment history of that pool. The line labelled "Naive" shows prices calculated assuming all pools to be homogeneous.