

# Bayesian Forecasting of Prepayment Rates for Individual Pools of Mortgages

Ivilina Popova\*, Elmira Popova† and Edward I. George‡

**Abstract.** This paper proposes a novel approach for modeling prepayment rates of individual pools of mortgages. The model incorporates the empirical evidence that prepayment is past dependent via Bayesian methodology. There are many factors that influence the prepayment behavior and for many of them there is no available (or impossible to gather) information. We implement this issue by creating a Bayesian mixture model and construct a Markov Chain Monte Carlo algorithm to estimate the parameters. We assess the model on a data set from the Bloomberg Database. Our results show that the burnout effect is a significant variable for explaining normal prepayment activities. This result does not hold when prepayment is triggered by non-pool dependent events. We show how to use the new model to compute prices for Mortgage Backed Securities. Monte Carlo simulation is the traditional method for obtaining such prices and the proposed model can be easily incorporated within simulation pricing framework. Prices for standard Pass-Throughs are obtained using simulation.

**Keywords:** mortgages, prepayment rates, MBS pricing, Bayesian forecasting

## 1 Introduction

Purchasing a house usually involves obtaining a loan (mortgage) originated by a financial institution. Any standard mortgage monthly payment consists of scheduled interest and principal. The borrower is also allowed to include additional payment toward the principal or early payoff of the whole mortgage. Refinancing of the mortgage is an example involving such a prepayment. The issuer of the mortgage usually sells the mortgages to another financial institution that pools them together and issues new securities, commonly known as mortgage backed securities (MBS). The MBS market is one of the largest bond markets in the United States (At the end of the second quarter of 2004 the outstanding stock of residential MBS in the United States was \$4.3 trillion.) The buyers of such structured products would like to know in advance the size of the incoming prepayments (if any). Improved forecasts of the prepayments would help them to better price such products.

One possible prepayment model is obtained by acting as if the borrower held a call option on the loan with exercise price equal to the outstanding balance (i.e. it is a time

---

\*Department of Finance, Albers School of Business and Economics, Seattle University, Seattle, WA, <mailto:popovai@seattleu.edu>

†Graduate Program in Operations Research and Industrial Engineering, The University of Texas at Austin, Austin, TX, <mailto:elmira@mail.utexas.edu>

‡Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, <mailto:edgerge@wharton.upenn.edu>

varying strike). Under optimal exercising conditions a mortgage can thus be priced as a callable bond. One would expect that the holder will prepay when the refinancing rate falls below the mortgage rate. However, the empirical evidence shows very different behavior, [Dunn and McConnell \(1981\)](#), that does not support such a model. It has been documented that homeowners often prepay when it is not optimal to do so and vice versa.

Recent literature attempts to model this partially “irrational” behavior. Prepayment activities can be classified as either interest rate related or non-interest rate related. Interest rate related activities (optimal prepayment) occur when the homeowners act in order to minimize the market value of the loan, whereas non-interest rate activities (suboptimal prepayments) occur for personal reasons of the borrowers, such as divorce, job change, etc.

Two approaches have been considered for modeling prepayments. [Downing et al. \(2005\)](#) characterize them as reduced-form and structural models. The structural approach assumes that mortgage termination is the optimal response of a rational borrower to changes in the values of variables like interest rates, housing prices, etc. It was first pioneered by [Dunn and McConnell \(1981\)](#) who introduced a model based on standard contingent claim pricing theory. In their model, prices of mortgage backed securities and prepayment behavior are determined together. They assume the Cox interest rate model, [Cox et al. \(1985\)](#), and introduce suboptimal prepayments as a Poisson events. Using a non-arbitrage argument, they derive a partial differential equation that can be solved numerically for the price of the security. Suboptimal prepayment behavior was first documented in their study. Other well-known structural models include [Timmis \(1985\)](#), [Dunn and Spatt \(2005\)](#), and [Stanton \(1995\)](#). [Stanton \(1995\)](#) presents a model that extends the option-theoretic approach. He models the transaction costs faced by mortgage holders and assumes that prepayment decisions occur at discrete times. This produces prepayment behavior that is consistent with the so-called burnout effect and usually occurs after the mortgages start to mature. If the interest rates decrease and percentage of the underlying loans fail to prepay constitutes the burnout. In other words, the borrowers who did not refinance during this interest rate drop period are less likely to do so if the interest rates drop again.

The reduced-form approach is empirical, mainly based on historical information. In the reduced-form approach, mortgage prepayment is modeled as a function of a set of variables that might influence the mortgage termination. [Schwartz and Torous \(1989\)](#) model the prepayment rate as a function of explanatory variables such as seasonality, burnout, difference between the contracting and re-financing rates, and speed of prepayment. Again, using a standard arbitrage argument they derive a partial differential equation that the MBS need to satisfy. Other well-known reduced-form models include [Deng et al. \(2000\)](#) and [Deng and Quigley \(2002\)](#).

Wall Street firms mainly use reduced-form models. [Richard and Roll \(1989\)](#) describe the prepayment model used by Goldman Sachs. Under their model, the conditional prepayment rate is determined by a product of functions of four important factors: refinancing incentive, age of the mortgage, seasonality and burnout. A non-linear least

squares optimization procedure is used to estimate the model parameters.

The reduced form approach is flexible and can closely mimic the historical data, however its out-of-sample performance can be weak. In comparison, the structural approach tends to perform well in *out-of-sample* settings since the main assumption is that mortgage prepayment arises from borrowers' optimizing behavior. However, due to the significant constraints on the relation between prepayments and the state variables, the model predictions diverge from the observed market prices.

All of the existing models (reduced-form and structural) estimate the prepayment function by using the information from all pools in the sample. In other words, they assume that all the pools manifest similar prepayment behavior. Stanton (1996) investigates the problem of predicting the prepayment for individual pools of mortgages. He reports that in 1,000 GNMA mortgage pools over a six and one-half year period, the unobservable heterogeneity is statistically significant. This could lead to very different prices of MBS backed by different pools. One of the latest models used by Wall Street firms (BlackRock) predicts individual pool prepayment rates based on detailed information about individual loans in the pools. Deng et al. (2005) use new heterogeneity-corrected 3-stage maximum likelihood specification to estimate the prepayment behavior of loan-level data from the Los Angeles metropolitan area.

Our contribution to the existing literature comes in several ways. First, we model the prepayment behavior on an individual pool level. Such modeling is important since the MBS and any other structured products are backed by individual pools of mortgages that exhibit very different prepayment behavior. Second, we use the Bayesian approach for parameter estimation. We believe this is the first article that uses such methodology. Our model incorporates the empirical evidence that prepayment is past dependent via Bayesian methodology. There are many factors that influence the prepayment behavior and for many of them there is no available (or impossible to gather) information. We implement this issue by creating a Bayesian mixture model and construct a Markov Chain Monte Carlo algorithm to estimate the parameters. We assess the model on a data set from the Bloomberg Database. We also show how to use the new model to compute prices for Mortgage Backed Securities. Monte Carlo simulation is the traditional method for obtaining such prices and the proposed model can be easily incorporated within simulation pricing framework. Prices for standard Pass-Throughs are obtained using simulation. In addition, we perform a very large numerical study by applying the Bayesian methodology to a set of 74 pools of data. In this process we observed several phenomena that (we believe) will be of interest to the statistical computation community.

The article is organized as follows. Section 2 describes the nature of the raw data and how the final data set is constructed. Section 3 describes the statistical prepayment model, Section 4 presents the empirical results, Section 5 shows how to use the model and price mortgage backed securities, and Section 6 concludes.

## 2 Prepayment Data Description

We consider historical data for 74 pools of mortgages, split between issues of Freddie Mac (a stockholder-owned corporation chartered by the US Congress to increase the supply of money that mortgage lenders can make available to homebuyers and multifamily investors.) and Ginnie Mae (the US Government National Mortgage Association.) To ensure homogeneity of the data, we exclusively focused on 30-year fixed-rate single-family 8% rate mortgages. For these pools, Bloomberg provided general and monthly pool information consisting of: issue date, maturity date, original amount, historical monthly prepayment (as percentage PSA). PSA stands for the Public Securities Association convention which assumes that 0.2% of the principal is paid in the first month, then increases by 0.2% for the following 29 months, and flattens at 6% until maturity. The PSA standard benchmark was introduced in 1985. It is not a model of prepayment but used as a benchmark in industry. Figure 1 shows the prepayment rates for 100% PSA.

We chose pools with original amounts of at least \$1 million. The age of these mortgages varies from 5 to 25 years, and consequently, the number of data points for each pool from 60 to 300. Tables 1 and 2 show the CUSIPs (a security identifier as defined by the Committee on Uniform Securities Identification Procedure) of the pools, their initiation date, initial dollar amount and current age (in months).

The average age for Freddie Mac pools is 122 months and the average initial dollar amount is approximately \$9 million. The average age for Ginnie Mae pools is 300 months and the average initial dollar amount is approximately \$3 million. In addition to the information provided by Bloomberg, we gathered historical long (30 year Treasury Bond) and short term (3 month Treasury Bill) interest rates from [Federal Reserve Bank \(2006\)](#).

The construction of the data set consists of several steps. For illustration purposes consider the mortgage pool with CUSIP 31340AC82. On July 1, 1984, Bloomberg reported that the prepayment activity for this pool is 300 PSA. As of the reporting date, the age of the mortgage pool was 7 months. The reported prepayment 300 PSA means that  $PSA(historical) = \frac{300}{100} = 3\%$  of  $PSA(standard)$ , where  $PSA(standard) = 6\% \times \frac{7}{30} = 1.4\%$ . Multiplying the two rates gives the actual historical prepayment rate. This annual rate is known as the Conditional Prepayment Rate (CPR) and is used to measure the speed of prepayments. For our example, the conditional prepayment rate (CPR) is:

$$CPR = PSA(standard) \cdot PSA(historical) = 1.4\% \times 3\% = 4.2\%$$

Given the annual CPR we can estimate the Single Monthly Mortality (SMM) rate. This measure assumes that there is a constant probability that the mortgage will be prepaid following the next month's scheduled payments. Given the definition of SMM, we can write that the probability the mortgage will survive a month is  $1 - SMM$ . For a period of one year, the probability of survival is  $(1 - SMM)^{12}$ . This is equal to  $1 - CPR$ . So, knowing the CPR we get the single monthly mortality rate as  $SMM =$

$$1 - (1 - CPR)^{\frac{1}{12}}.$$

Now we have all the ingredients to construct the monthly cash flows for the mortgage pools<sup>1</sup>. First, we compute the monthly payment,  $MP$ <sup>2</sup>, and the interest payment,  $IP$ <sup>3</sup>: For  $t = 0, \dots, 360$

$$MP = \frac{FaceValue \left[ \frac{0.08}{12} \right] \left[ 1 + \frac{0.08}{12} \right]^t}{\left[ 1 + \frac{0.08}{12} \right]^t - 1}$$

and

$$IP = FaceValue \left[ \frac{0.08}{12} \right]$$

Thus, the scheduled principal payment at time  $t$  is obtained by subtracting the interest payments from the total mortgage payments:  $SP = MP - IP$ . The nonscheduled prepayment (NPP) is computed by multiplying the reduced face value of the pool by the single monthly mortality rate:  $NPP = SMM(FaceValue - SP)$ , and the actual payment  $AP = SP + NPP$ . In our analysis we model the actual payment  $AP_t$  (in dollars) for each mortgage pool at the end of month  $t$ .

To assess the general structure of the data and to identify the presence of multimodality we used kernel density estimation, see [Silverman \(1981\)](#). We entered as an input the natural logarithm of the actual monthly payments for all 74 pools of mortgages. The produced density estimators revealed that the majority of the mortgage pools have at least 2 modes.

Figures 2–5 show 8 such densities (all 74 densities are available upon request.) As we model the total payment from a pool the kernel density estimators are based on the  $N$  monthly total payments from that pool. The observed multimodality is intuitively reasonable - it seems likely that there are borrowers who prepay small amounts each month, as well as borrowers who prepay the whole mortgage. Indeed, there are probably certain events such as job loss, refinancing or house selling that trigger such “small” or “large” prepayment behavior. If we knew these events and could gather data associated with them, then we might be able to predict the next month prepayment. However, up to now, there is no research regarding such events, their existence and data availability.

### 3 A Bayesian Prepayment Model

We now proceed to propose and implement a model for  $y_t = \ln(AP_t)$ , the logarithm of the total money paid at the end of month  $t$ , for each of the 74 mortgage pools. As potential predictors for this model, we consider the following set of covariates which were found by [Schwartz and Torous \(1989\)](#) to be useful for modeling the prepayment

<sup>1</sup>For detailed numerical examples regarding constructing monthly mortgage cash flows see Chapter 9 of [Sundaesan \(1997\)](#)

<sup>2</sup>Equation 9.1, page 363 from [Sundaesan \(1997\)](#).

<sup>3</sup>Equation 9.4, page 364 from [Sundaesan \(1997\)](#).

(or hazard) function<sup>4</sup> in a proportional hazards model for the time until prepayment of a mortgage:

- $x_{t1}$  = the difference between the mortgage rate and the short term interest rate
- $x_{t2} = (x_{t1})^3$
- $x_{t3}$  captures the burnout effect - it is the logarithm of the ratio of the dollar amount of the pool  $i$  outstanding at time  $t$ , to the pool's principal which would prevail at  $t$  in the absence of prepayments
- $x_{t4}$  models the seasonality effect. It equals 1 for the months of May, June, July, and August, and 0 otherwise.
- $x_{t5}$  is the spread (difference) between the long and short term interest rates.

Although it would seem natural to consider a single regression of  $y_t$  on  $x_{t1}$ ,  $x_{t2}$ ,  $x_{t3}$ ,  $x_{t4}$ , and  $x_{t5}$  above, density plots of  $y_t$  such as those in Figures 2 – 5 suggest that the  $y_t$  distributions are multimodal, with at least two main groups of “small” and “large” prepayers. To account for this multimodality, we allow for several regression functions by considering a mixture of regressions of  $y_t = \ln(AP_t)$  on  $x_{t1}, x_{t2}, x_{t3}, x_{t4}, x_{t5}$ , (see Hurn et al. 2003, for an alternative presentation). More precisely, for each mortgage pool, we model  $y_t$ ,  $t = 1, \dots, N$  as realizations of independent draws from a  $k$ -component normal mixture

$$f(y_t) = \sum_{j=1}^k p_j f_j(y_t | \mu_{tj}, w_j), \quad (1)$$

where  $p_j$  is the probability that  $y_t$  belongs to component  $j$ ,  $f_j(\cdot | \mu_{tj}, w_j)$  is the density function of a normal distribution with mean  $\mu_{tj}$  and precision<sup>5</sup>  $w_j$ , and

$$\mu_{tj} = \sum_{l=0}^5 \beta_{jl} x_{tl} \quad (2)$$

with  $x_{t0} \equiv 1$ .

To fit and draw inference from this model, we take a Bayesian approach by assigning prior distributions to all the unknown parameters: the regression coefficients  $\beta = (\beta_1, \dots, \beta_k)$  where  $\beta_j = [\beta_{j0}, \dots, \beta_{j5}]$ , the precision parameters  $w = (w_1, \dots, w_k)$  and the component assignment probabilities  $p = (p_1, \dots, p_k)$ . For  $(\beta, w)$ , we assign the improper default prior  $\pi(\beta, w) = \left(\prod_{j=1}^k w_j\right)^{-1}$  for  $w_1, \dots, w_k > 0$ . For  $p = (p_1, \dots, p_k)$ , we assign a Dirichlet( $\alpha_1, \dots, \alpha_k$ ) distribution with the default choice  $\alpha_j \equiv 0.5$ .

For posterior calculation, we use a Gibbs sampler (Gelfand and Smith 1991) to simulate a Markov chain sample of parameter values from the full posterior  $\pi(\theta | y)$ , where

<sup>4</sup>Their prepayment function at time  $t$  is the probability that a prepayment occurs at time  $t + \Delta t$  given that no prepayment occurred by time  $t$ .

<sup>5</sup>Precision = 1/Variance

$\theta = (\beta, w, p)$  and  $y = (y_1, \dots, y_N)$ . For a mixture model such as (1), implementation of the Gibbs sampler is well known to be facilitated by the introduction of latent variables for component membership of all the  $y_t$  values, Diebolt and Robert (1994). More precisely, let  $z_t = (z_{t1}, z_{t2}, \dots, z_{tk})$  denote the vector of 0's and a single 1, i.e.  $z_{tj} \in \{0, 1\}$  and  $\sum_{j=1}^k z_{tj} = 1$ , such that  $z_{tj} = 1 \Leftrightarrow y_t$  was a draw from the density  $f_j$  of component  $j$ . Note that apriori  $P(z_{tj} = 1) = p_j$ . By considering  $z_t$  as missing data, the density under (1) of the completed data  $(y_t, z_t)$  can be expressed as

$$\prod_{j=1}^k [p_j f_j(y_t | \mu_{tj}, w_j)]^{z_{tj}}, \quad t = 1, \dots, N.$$

Letting  $z = (z_1, \dots, z_N)$  be the full component identification vector for  $y$ , Gibbs sampling is easily obtained by starting with initial value  $\theta^{(0)}$  and then simply iterating between the conditional posteriors  $\pi(z^{(m)} | \theta^{(m-1)}, y)$  and  $\pi(\theta^{(m)} | z^{(m)}, y)$  for  $m = 1, \dots, M$ . We now proceed to describe how we sample from each of these conditional posteriors.

To sample from  $\pi(z | \theta, y)$ , set the component  $z_{tj}$  of  $z_t$  equal to 1 (and the other components to 0) with probability

$$p_j^* = \frac{p_j f_j(y_t | \theta_j)}{\sum_{j=1}^k p_j f_j(y_t | \theta_j)}, \quad j = 1, \dots, k,$$

independently for  $t = 1, \dots, N$ .

To describe sampling from  $\pi(\theta | z, y)$ , note that

$$\pi(\theta | z, y) = \pi(p | z, y)\pi(w | z, y)\pi(\beta | w, z, y)$$

so it suffices to sample from each of these. Beginning with  $p$  and letting  $n_j$  be the number of  $y$  values assigned to component  $j$  by  $z$ ,  $p$  is simply obtained from

$$p \sim \text{Dirichlet}(n_1 + \alpha, \dots, n_k + \alpha_k).$$

To describe sampling  $w$  and then  $\beta$  we first need some notation. Let  $y_j$  be the subvector of  $y$  whose  $n_j$  values have been assigned to component  $j$  by  $z$ . Let  $X_j$  be the corresponding matrix whose  $n_j$  rows  $x_t = (x_{t0}, \dots, x_{t5})$  have also been assigned to component  $j$  by  $z$ . Let  $\hat{\beta}_j = (X_j' X_j)^{-1} X_j' y_j$  be the least squares estimate of  $\beta_j$ , and let  $s_j^2 = \frac{1}{n_j - 6} \|y_j - X_j \hat{\beta}_j\|^2$  be the usual unbiased estimate of the variance of  $f_j$ . Based on these statistics, each component of  $w$  is independently obtained from

$$w_j \sim \text{Gamma}\left(\frac{n_j - 6}{2}, \frac{n_j - 6}{2} \hat{s}_j^2\right)$$

and then given this choice of  $w$ , each component of  $\beta$  is independently obtained from

$$\beta_j \sim \text{Normal}(\hat{\beta}_j, (w_j X_j' X_j)^{-1}).$$

We remark to be able to maintain the propriety of the posterior distributions we include a condition in our algorithm where we reject a simulation step  $m$  if the number of data points in each of the clusters is less than or equal to 6.

## 4 Empirical Results

We implemented the above MCMC algorithm in C++ on a 2.40GHz Pentium using the uniform random number generator written by [Lecuyer \(2006\)](#) and non uniform random number algorithms from [Devroye \(1986\)](#). For each model and each pool, the length of our simulation runs was 5000. We estimated two models: the first one with two mixture components and the second one with three mixture components.

### 4.1 Model 1: Two Mixtures

For the two component mixture model, Figures 6, 7, 8, and 9 show the actual simulated points for  $w_1$ ,  $w_2$ ,  $\beta_{10}$ , and  $\beta_{20}$  for one of the mortgage pools. Based on the observed patterns of the simulated values, we dropped the first 1000 simulated values to eliminate the initial transient period, and used only the last 4000 simulated values for estimation and inference. We observed similar patterns and did the same with the remaining 73 pools of mortgages. For each pool, we used the average of the MCMC sample values to approximate the posterior mean estimates of the parameters, and their 95% posterior probability interval estimates.

Table 3 shows our results for all Freddie Mac pools and Tables 5 and 6 show our results for all Ginnie Mae pools. We reported only those parameters with probability intervals not covering zero for at least 50% of the pools. For Freddie Mac the relevant parameters are  $\beta_{10}$ ,  $\beta_{20}$  and  $\beta_{23}$ , where  $\beta_{10}$  and  $\beta_{20}$  are the intercepts and  $\beta_{23}$  is the variable measuring the burnout effect. For Ginnie Mae the parameters are  $\beta_{10}$ ,  $\beta_{13}$ ,  $\beta_{20}$  and  $\beta_{23}$ , where  $\beta_{10}$  and  $\beta_{20}$  are the intercepts and  $\beta_{13}$  and  $\beta_{23}$  measure the burnout effect.

Note that most of the coefficients for  $\beta_{13}$  and  $\beta_{23}$  are negative and the corresponding probability intervals do not include zero. This result is very intuitive since as the prepaid amount increases, a prepayment becomes less likely. There are only two Freddie Mac pools that have positive burnout coefficients. After examining their prepayment history we observed that the prepayment activity increased as the burnout effect increased, so the results are consistent with the historical prepayment behavior.

Table 7 summarizes the results from Tables 3, 5, and 6. We report the average parameter values across all pools. For Ginnie Mae pools the average burnout coefficient for the first cluster is  $-0.62$  and for the second cluster is  $-1.23$  with an average probability of 42% of choosing cluster one and 58% of choosing cluster two. The corresponding results for the Freddie Mac pools are  $-0.94$  for the burnout coefficient of the second cluster with an average probability of 48% of choosing cluster one and 52% of choosing cluster 2. The common theme here is that the coefficient is “more negative”



for the distribution that occurs more often, which represents the “normal prepayment behavior”. For normal prepayment activities it is to be expected that the prepayment will decrease as the principal amount left in the pool decreases. In contrast, the second distribution represents prepayment that are out-of-the-norm, they can be triggered by non-pool related variables such as selling the house, moving, bankruptcy, etc.

Based on the posterior mean parameter estimates, we are able to gauge the forecast quality for each of the pools and for all pools together. Figure 10 compares the actual prepayment with the forecasted prepayment for one of the mortgage pools.

The lower and upper solid curves are  $\sum_{l=0}^5 \tilde{\beta}_{1l} x_{tl}$  and  $\sum_{l=0}^5 \tilde{\beta}_{2l} x_{tl}$  respectively, where  $\tilde{\beta}_{1l}$  and  $\tilde{\beta}_{2l}$  are the parameter posterior mean estimates. The model will forecast an “average” prepayment with probability  $\hat{p}_1$  and a “high” prepayment with probability  $\hat{p}_2$ . Here  $\hat{p}_1$  and  $\hat{p}_2$  are the posterior mean estimates of the parameters  $p_1$  and  $p_2$ . To compute the associated error (residual) for each point, we calculate it’s distance to the weighted average of the upper and lower curves (the middle line).

## 4.2 Comparison with Model 2: Three Mixtures Model

Following the steps presented in §3 we estimated the model with three components for each of the 74 mortgage pools. For comparison with the previous two component model, we calculated Bayes factors for each pool using the methodology described by Kass and Raftery (1995). The null hypothesis is that the data are coming from the two components mixture model vs the alternative of three components mixture. We sampled from the posterior distributions and computed the harmonic mean of the likelihood values, (see Kass and Raftery 1995, equation 11). The Bayes factor ( $B_{10}$ ) is simply the ratio of the estimated likelihoods. Table 8 shows the results for all 74 mortgage pools. The tables contain two columns for the Bayes factor - one that is the ratio defined above ( $B_{10}$ ), and the second one is  $2 \log B_{10}$ . Kass and Raftery (1995) recommend the second one since it is on the same scale as the likelihood ratio test statistics. They also discuss the proper threshold to be used for rejection of the null hypothesis. The default value is 150 (first recommended by Jeffreys in 1961) but they argue that if this is to be used as a forensic evidence on a criminal trial then the threshold has to be at least 1000.

As a threshold for  $B_{10}$  to reject the null hypothesis, we used the threshold of 150, and found that 6 out of the 74 pools reject the null hypothesis of a two component mixture model. We obtained the same result using  $2 \log B_{10}$  and a threshold of 10.

One might also consider goodness-of-fit tests using the approach described by Robert and Rousseau (2007). The null hypothesis is  $H_0 : Y \sim F_\theta$  for an unknown  $\theta$ , i.e. there exists  $\theta$  such that  $F_\theta(Y) \sim U(0, 1)$ . We illustrate this approach by applying it to one of the pools. We used the average expected posterior value for each parameter across the data set and estimated the cumulative probabilities for both models. The Kolmogorov-Smirnov statistic for the two component model equals to 0.15 comparing to 0.55 for the three component model. We cannot reject the null hypothesis at 1% level of significance for the two component model but we reject it for the three component model. Figure 11 shows the fitted and the Uniform(0,1) cumulative probabilities for the

two component model.

## 5 Pricing Mortgage Backed Securities

Mortgage backed securities are usually priced by Monte Carlo simulation. First, a sample path for the interest rates is generated by using some model explaining the dynamic behavior of the interest rates. Each month, expected prepayments are calculated from the current yield curve and the prepayment model. These prepayments determine the expected cash flows to the holder of the MBS and the cash flows are discounted to time zero to obtain a sample value of the MBS. An estimate of the price of the MBS is obtained by performing many simulation runs.

We will use as an example Mortgage-Pass-Through securities. They are the simplest form of MBS. A pass-through represents a share of the underlying mortgage pool. The cash flows generated by the pool are passed on to the security holders on a pro-rata basis. They are delivered monthly and consist of three components - interest payment, scheduled principal payment and prepayment.

We will price a pass-through security by making the following assumption: there are 200 units of the pass-through backed by a pool. For example, if we assume that the value of the pool is \$100 million, each unit amounts to \$500,000 and is entitled to 0.5% of the cash flows.

In the United States, there are two types of pass-throughs: agency and non-agency pass-throughs. Agency pass-throughs are issued by Ginnie Mae, Freddie Mac and Fannie Mae. We will use one of the Freddie Mac pools for our illustration, in particular pool with CUSIP 31340CEJ2. The initial amount of the pool is \$93,073,150 and the pool has 250 months left till maturity. The assumption is that there are 200 units of the pass-through on pool 31340CEJ2, resulting in a 0.5% claim of the cash flow for every unit.

In order to generate paths for the interest rates we will simulate from one factor Cox et al. (1985) model for the dynamics of the short term interest rate:

$$dr = a(b - r)dt + \sigma(t, T)\sqrt{r}dW \quad (3)$$

This model belongs to the class of equilibrium models and is one of the standards in interest rates modeling. Here the short rate is mean-reverting and the process for the rates is always non-negative. Cox, Ingersoll and Ross show that the bond prices have the following form:

$$P(t, T) = A(t, T)e^{-B(t, T)r(t)} \quad (4)$$

where,  $P(t, T)$  is the price of a pure discount bond at time  $t$  that matures at time  $T$ . Also,

$$B(t, T) = \frac{2(e^{\gamma(T-t)} - 1)}{(\gamma + a)(e^{\gamma(T-t)} - 1) + 2\gamma}$$

and

$$A(t, T) = \left[ \frac{2\gamma e^{(a+\gamma)(T-t)/2}}{(\gamma + a)(e^{\gamma(T-t)} - 1) + 2\gamma} \right]^{2ab/\sigma^2}$$

with  $\gamma = \sqrt{a^2 + 2\sigma^2}$ . The long rate is given by:

$$R(t, T) = \frac{1}{T-t} [\ln A(t, T) + B(t, T)r(t)]$$

We can see that the long rate is linearly dependent on  $r(t)$ . This means that the value of  $r(t)$  determines the level of the term structure at time  $t$ .

For simulation purposes we chose the following values of the parameters:  $a = 8\%$ ,  $b = 8\%$ ,  $\sigma = 2\%$ ,  $r(0) = 8\%$ . Figure 6 shows 10 simulated short rate paths.

Along each path we can compute the corresponding long rate and use the simulated covariates to compute an estimated prepayment from the Bayesian model. Since there are 250 months till maturity we simulated 250 months of data. Figure 12 shows that the range of the simulated rates is quite wide, from 1% to 20%. At every point along the simulated path we used the covariates and produced the corresponding prepayment amount based on the mixture model. Figure 13 shows the historical prepayment for the pool as well as 10 simulated prepayment paths.

The final price of the Pass-Through is computed by discounting the actual pool payment along the simulated paths. Based on 1,000 simulation runs the price of 1 unit of the Pass-Through on Pool 31340CEJ2 is \$131,054.05.

## 6 Conclusion

In this paper, we have proposed using two or three component normal regression mixture models to describe the apparent multimodal distribution of prepayments within mortgage pools over time. Distinct linear regression functions of observed covariates are used to model the means of the components. We consider a model where the mixture probabilities are constrained to be fixed over time. To fit this multiparameter nonlinear model, we take a Bayesian statistical approach where the uncertainty about all the unknown parameters is described by prior distributions. For this setup, a Markov Chain Monte Carlo algorithm is constructed and used to carry out all the computations. Empirical results show that the fixed mixture weight model appears to fit the observed data reasonably well. We also show how to use the new model to compute prices for mortgage backed securities. Monte Carlo simulation is the traditional method for obtaining such prices and the proposed model can be easily incorporated within simulation pricing framework.

We feel that our model is a good start towards the modeling of the individual pool prepayment rate process. Naturally, the extent to which this model can be effective depends on the quality of the available covariates. More and better covariate information will likely lead to better fits and forecasts. Another future direction for potential improvements will be to consider elaborations to larger models that can exploit infor-

mation across similar mortgage pools. One such elaboration would be a hierarchical Bayes model that treats the parameters of each pool as a sample from a superpopulation model. Such an elaboration would be particularly natural given the Bayesian treatment we have here considered.

## References

- Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). "A theory of the term structure of interest rates." *Econometrica*, 53(2): 385–407. 394, 402
- Deng, Y., Pavlov, A., and Yang, L. (2005). "Spatial Heterogeneity in Mortgage Terminations by Refinance, Sale and Default." *Real Estate Economics*, 33: 739–764. 395
- Deng, Y. and Quigley, J. M. (2002). "Woodhead Behavior and the Pricing of Residential Mortgages." *Working Paper, University of Southern California, Los Angeles*. 394
- Deng, Y., Quigley, J. M., and Order, R. V. (2000). "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options." *Econometrica*, 68: 275–307. 394
- Devroye, L. (1986). *Non uniform random variate generation*. Springer Verlag. 400
- Diebolt, J. and Robert, C. P. (1994). "Estimation of finite mixture distributions through Bayesian sampling." *Journal of the Royal Statistical Society, B* 56(2): 363–375. 399
- Downing, C., Stanton, R., and Wallace, N. (2005). "An Empirical test of a Two-Factor Mortgage Valuation Model: How much Do House Prices Matter?" *Real Estate Economics*, 33: 681–710. 394
- Dunn, K. and McConnell, J. (1981). "Valuation of GNMA mortgage-backed securities." *The Journal of Finance*, XXXVI(3): 599–616. 394
- Dunn, K. and Spatt, C. (2005). "The Effect of Refinancing Costs and Market Imperfections on the Optimal Call Strategy and the Pricing of Debt Contracts." *Real Estate Economics*, 33(4): 595–617. 394
- Federal Reserve Bank (2006). <http://www.stld.frb.org/fred>. 396
- Gelfand, A. E. and Smith, A. F. (1991). "Gibbs sampling for marginal posterior expectations." *Communications in Statistics - Theory and Methods*, 20(5): 1747 – 1766. 398
- Hurn, M., Justel, A., and Robert, C. P. (2003). "Estimating mixtures of regressions." *Journal of Computational and Graphical Statistics*, 12(1): 1–25. 398
- Kass, R. E. and Raftery, A. E. (1995). "Bayes factors." *Journal of the American Statistical Association*, 90(430): 773–795. 401
- Lecuyer, P. (2006). <http://www.iro.umontreal.ca/lecuyer>. 400

- Richard, R. and Roll, R. (1989). “Modeling Prepayments on Fixed-Rate Mortgage-Backed Securities.” *Journal of Portfolio Management*, 15(3): 73–82. 394
- Robert, C. P. and Rousseau, J. (2007). “A mixture approach to Bayesian goodness of fit.” Working paper, <http://www.ceremade.dauphine.fr/~xian/rrapr03.pdf>. 401
- Schwartz, E. and Torous, W. (1989). “Prepayment and the Valuation of Mortgage-Backed Securities.” *Journal of Finance*, 44(2): 375–392. 394, 397
- Silverman, B. W. (1981). “Using kernel density estimates to investigate multimodality.” *Journal of the Royal Statistical Society*, 43(1): 97–99. 397
- Stanton, R. (1995). “Rational Prepayment and the Valuation of Mortgage-Backed Securities.” *Review of Financial Studies*, 8(3): 677–708. 394
- (1996). “Unobservable Heterogeneity and Rational Learning: Pool-Specific versus Generic Mortgage-Backed Security Prices.” *Journal of Real Estate and Economics*, 12: 243–263. 395
- Sundaresan, S. M. (1997). *Fixed income markets and their derivatives*. South-Western College Publishing. 397
- Timmis, G. C. (1985). “Valuation of GNMA Mortgage-Backed Securities with Transaction Costs, Heterogeneous Households and Endogenously Generated Prepayment Rates.” *Working Paper, Carnegie-Mellon University, Pittsburgh, PA*. 394

#### Acknowledgments

This research has been partially supported by grant #003658-0763 from the State of Texas Advanced Research Program and by the National Science Foundation grants CMMI-0457558 and DMS-0605102.

Table 1: Data summary, Ginnie Mae

CUSIP	Origination Date	Initial Amount	Age in months
362027BX1	1/1/1973	\$ 5,388,014.99	308
362027JP0	1/1/1973	\$ 2,303,787.79	308
362027J47	1/1/1973	\$ 2,156,245.31	308
362027KB9	1/1/1973	\$ 1,994,981.34	308
362026VD5	2/1/1973	\$ 4,057,248.04	304
362027RC0	4/1/1973	\$ 4,199,561.59	305
362027R55	4/1/1973	\$ 1,709,405.67	305
362027SS4	4/1/1973	\$ 3,203,047.00	305
362027UJ1	6/1/1973	\$ 3,040,361.00	304
362027UK8	6/1/1973	\$ 3,812,288.98	304
362027V76	7/1/1973	\$ 2,818,570.48	303
362027FQ2	11/1/1973	\$ 2,249,925.76	298
362027HZ0	11/1/1973	\$ 2,029,476.62	299
362027KW3	11/1/1973	\$ 2,015,190.05	299
362027N26	11/1/1973	\$ 3,642,915.05	299
3620274Q4	11/1/1973	\$ 2,001,116.95	297
362028A77	11/1/1973	\$ 2,020,035.56	298
362028BT8	11/1/1973	\$ 2,007,416.51	299
362028FM9	11/1/1973	\$ 5,000,616.25	299
362027HS6	12/1/1973	\$ 2,020,564.41	298
362027QY3	12/1/1973	\$ 2,010,207.98	296
362027T95	12/1/1973	\$ 7,912,679.24	298
3620274L5	12/1/1973	\$ 2,015,578.53	298
3620274N1	12/1/1973	\$ 2,194,180.55	298
3620276T6	12/1/1973	\$ 3,011,643.62	298
362028AZ5	12/1/1973	\$ 5,000,764.88	298
362028BX9	12/1/1973	\$ 2,008,693.49	298
362028DY5	12/1/1973	\$ 5,009,615.27	298
362028EQ1	12/1/1973	\$ 10,234,823.28	298
362028FB3	12/1/1973	\$ 4,021,388.17	297
362028F31	12/1/1973	\$ 2,000,630.91	296
362028HB1	12/1/1973	\$ 2,014,895.05	298
362028HU9	12/1/1973	\$ 2,000,041.48	298
362028JF0	12/1/1973	\$ 2,499,138.92	298
362028JK9	12/1/1973	\$ 4,017,510.77	297

Table 2: Data summary, Freddie Mac

CUSIP	Origination Date	Initial Amount	Age in months
31340B5T2	12/1/1988	\$4,022,391.00	118
31340B5Y1	12/1/1988	\$3,425,491.00	118
31340CBW6	11/1/1992	\$33,317,099.00	71
31340CB55	5/1/1993	\$9,342,593.00	65
31340CEJ2	9/1/1989	\$93,073,150.00	109
31340CEL7	9/1/1989	\$1,199,680.00	109
31340CER4	9/1/1989	\$4,111,316.00	109
31340CE60	10/1/1989	\$1,312,970.00	108
31340CFR3	12/1/1989	\$3,012,953.00	106
31340CBK2	11/1/1991	\$33,884,130.00	83
31340CBL0	12/1/1991	\$12,484,193.00	82
31340CBP1	1/1/1992	\$14,765,342.00	81
31340CBR7	2/1/1992	\$5,805,280.00	80
31340CBS5	2/1/1992	\$9,970,966.00	80
31340CBA4	7/1/1989	\$20,690,305.00	111
31340CBG1	9/1/1991	\$20,843,194.00	85
31340CBJ5	10/1/1991	\$28,367,731.00	84
31340CAH0	3/1/1989	\$1,068,345.00	115
31340CAJ6	3/1/1989	\$4,053,510.00	115
31340CAN7	3/1/1989	\$3,157,634.00	115
31340B6B0	1/1/1989	\$3,029,407.00	117
31340B6J3	1/1/1989	\$1,000,811.00	117
31340B6V6	2/1/1989	\$1,198,354.00	116
31340B7L7	2/1/1989	\$1,472,666.00	116
31340CAA5	2/1/1989	\$1,007,985.00	116
31340CAC1	2/1/1989	\$1,234,631.00	116
31340CAF4	3/1/1989	\$1,711,867.00	115
31340B5P0	12/1/1988	\$1,209,824.00	118
31340B5H8	12/1/1988	\$2,191,105.00	118
31340AEU1	7/1/1984	\$3,924,740.00	170
31340AEL1	6/1/1984	\$3,328,475.00	171
31340AD40	4/1/1984	\$1,687,252.00	173
31340ADW8	3/1/1984	\$2,856,860.00	174
31340ADH1	1/1/1984	\$9,493,396.00	174
31340ADG3	1/1/1984	\$8,615,957.00	176
31340ADF5	1/1/1984	\$1,081,768.00	176
31340ADC2	12/1/1983	\$1,031,959.00	177
31340AC82	12/1/1983	\$1,443,159.00	177
31340ABB6	9/1/1983	\$1,354,125.00	180

Table 3: Results from two component model, Freddie Mac. Here LPB stands for the left probability bound, and RPB for the right probability bound

CUSIP	$\beta_{10}$				$\beta_{20}$			
	Avg	SD	LPB	RPB	Avg	SD	LPB	RPB
31340abb6	9.66	0.05	9.57	9.75	10.87	0.08	10.72	11.02
31340ac82	9.29	0.11	9.07	9.50	11.40	0.45	10.52	12.28
31340ad40	9.95	0.19	9.57	10.33	9.88	0.21	9.48	10.28
31340adc2	9.13	0.96	7.26	11.01	9.24	0.39	8.48	10.00
31340adf5	9.46	0.29	8.90	10.03	9.92	0.43	9.08	10.77
31340adg3	11.19	0.05	11.09	11.29	11.78	0.14	11.50	12.06
31340adh1	11.59	0.14	11.31	11.86	11.15	0.19	10.77	11.53
31340adw8	9.90	0.14	9.63	10.17	11.40	0.33	10.75	12.05
31340ael1	10.13	0.40	9.34	10.92	11.27	0.35	10.58	11.96
31340aeu1	10.33	0.04	10.25	10.42	11.01	0.03	10.95	11.08
31340b5h8	9.77	0.05	9.67	9.87	11.09	0.31	10.47	11.70
31340b5p0	9.62	0.13	9.37	9.88	9.26	0.12	9.02	9.51
31340b5t2	10.98	0.42	10.15	11.81	10.59	0.49	9.62	11.55
31340b5v7	10.10	0.08	9.95	10.25	1.50	3.34	-5.05	8.05
31340b5y1	10.17	0.10	9.98	10.35	10.40	0.26	9.89	10.91
31340b6b0	10.06	0.09	9.90	10.23	10.21	0.34	9.55	10.88
31340b6j3	8.96	0.10	8.77	9.15	8.00	0.55	6.92	9.09
31340b6v6	9.27	0.09	9.10	9.44	9.72	0.31	9.11	10.32
31340b7l7	9.36	0.08	9.20	9.51	10.86	0.45	9.98	11.73
31340caa5	8.82	0.05	8.72	8.91	9.57	0.16	9.25	9.88
31340cac1	9.31	0.08	9.16	9.46	10.50	3.54	3.56	17.45
31340caf4	9.60	0.24	9.12	10.08	10.09	0.66	8.80	11.37
31340cah0	9.07	0.12	8.85	9.30	8.95	0.53	7.91	9.98
31340caj6	10.47	0.12	10.24	10.71	11.01	0.12	10.76	11.25
31340can7	10.25	0.07	10.11	10.39	10.81	0.10	10.61	11.01
31340cb55	11.06	0.20	10.68	11.45	13.41	1.18	11.09	15.73
31340cba4	12.31	0.21	11.89	12.72	12.15	0.16	11.85	12.46
31340cbg1	10.61	0.53	9.57	11.65	13.31	0.30	12.71	13.90
31340cbj5	11.20	0.75	9.74	12.67	13.00	0.28	12.44	13.55
31340cbk2	11.64	0.82	10.04	13.24	12.34	0.31	11.73	12.94
31340cbl0	11.46	0.93	9.63	13.29	11.60	0.47	10.68	12.51
31340cbp1	9.36	0.83	7.73	11.00	13.26	0.53	12.22	14.31
31340cbr7	9.62	1.02	7.61	11.63	11.19	0.28	10.63	11.75
31340cbs5	12.21	0.58	11.06	13.35	11.41	0.27	10.87	11.94
31340cbw6	13.11	0.43	12.27	13.95	13.09	0.43	12.24	13.94
31340ce60	9.17	0.56	8.08	10.26	9.30	0.98	7.38	11.22
31340cej2	13.34	0.15	13.03	13.64	13.44	0.19	13.06	13.82
31340cel7	9.14	0.11	8.92	9.35	3.03	3.84	-4.50	10.57
31340cer4	10.40	0.08	10.25	10.55	10.67	0.22	10.24	11.10
31340cfr3	9.56	0.30	8.96	10.16	11.51	0.92	9.71	13.32



Table 4: Table 3 cont.

CUSIP	$\beta_{23}$			
	Avg	SD	LPB	RPB
31340abb6	-1.26	0.03	-1.32	-1.20
31340ac82	-0.81	0.24	-1.28	-0.33
31340ad40	-1.58	0.16	-1.89	-1.28
31340adc2	-1.21	0.53	-2.25	-0.16
31340adf5	-1.81	0.36	-2.50	-1.11
31340adg3	-0.81	0.11	-1.02	-0.59
31340adh1	-1.03	0.12	-1.27	-0.79
31340adw8	-1.26	0.07	-1.40	-1.11
31340ael1	-1.22	0.09	-1.39	-1.05
31340aeu1	-1.25	0.05	-1.36	-1.14
31340b5h8	-0.56	0.14	-0.84	-0.27
31340b5p0	-1.77	0.70	-3.14	-0.39
31340b5t2	-0.84	0.47	-1.75	0.07
31340b5v7	3.26	0.35	2.57	3.94
31340b5y1	-1.58	0.34	-2.26	-0.91
31340b6b0	-1.14	0.10	-1.34	-0.94
31340b6j3	-1.15	0.40	-1.94	-0.36
31340b6v6	-1.28	0.23	-1.73	-0.83
31340b7l7	-1.03	0.28	-1.57	-0.49
31340caa5	-1.93	0.10	-2.14	-1.73
31340cac1	1.02	0.49	0.06	1.99
31340caf4	-1.38	0.25	-1.88	-0.88
31340cah0	-1.76	0.31	-2.36	-1.16
31340caj6	-0.66	0.13	-0.92	-0.40
31340can7	-1.17	0.44	-2.03	-0.32
31340cb55	-1.45	0.08	-1.60	-1.30
31340cba4	-0.82	0.17	-1.15	-0.49
31340cbg1	-1.23	0.10	-1.43	-1.03
31340cbj5	-1.13	0.08	-1.28	-0.98
31340cbk2	-1.18	0.09	-1.36	-0.99
31340cbl0	-0.97	0.14	-1.24	-0.70
31340cbp1	-1.38	0.12	-1.62	-1.15
31340cbr7	-1.12	0.12	-1.35	-0.89
31340cbs5	-1.27	0.15	-1.57	-0.97
31340cbw6	-1.25	0.13	-1.50	-1.01
31340ce60	-1.15	0.18	-1.49	-0.80
31340cej2	-1.14	0.20	-1.53	-0.76
31340cel7	4.75	0.11	4.52	4.97
31340cer4	-1.26	0.13	-1.51	-1.01
31340cfr3	-2.72	0.72	-4.14	-1.30

Table 5: Results from two component model, Ginnie Mae. Here LPB stands for the left probability bound, and RPB for the right probability bound

CUSIP	$\beta_{10}$				$\beta_{13}$			
	Avg	SD	LPB	RPB	Avg	SD	LPB	RPB
362026yj9	9.39	0.41	8.58	10.20	-0.73	0.45	-1.62	0.15
3620274l5	9.65	0.10	9.46	9.83	-0.94	0.05	-1.04	-0.83
3620274n1	9.74	0.02	9.70	9.77	-0.94	0.02	-0.97	-0.90
3620274q4	10.11	0.05	10.01	10.20	-0.24	0.08	-0.39	-0.09
3620276t6	11.14	0.25	10.64	11.63	-0.67	0.16	-0.99	-0.35
362027bx1	11.38	0.04	11.30	11.46	-0.76	0.03	-0.82	-0.70
362027fq2	9.85	0.11	9.63	10.08	-0.69	0.35	-1.37	-0.01
362027hs6	9.72	0.22	9.29	10.15	0.28	1.32	-2.30	2.87
362027hz0	10.81	0.06	10.70	10.93	-0.58	0.05	-0.68	-0.49
362027j47	10.63	0.04	10.55	10.70	-0.87	0.02	-0.90	-0.83
362027jp0	10.46	0.04	10.39	10.54	-0.67	0.04	-0.76	-0.58
362027kb9	10.49	0.03	10.44	10.55	-0.77	0.03	-0.83	-0.72
362027ke3	8.60	0.11	8.39	8.81	-0.75	1.84	-4.36	2.85
362027kw3	9.61	0.02	9.58	9.65	-0.93	0.02	-0.96	-0.89
362027n26	11.19	0.04	11.12	11.26	-0.88	0.03	-0.95	-0.82
362027qy3	9.77	0.16	9.46	10.08	-0.26	0.54	-1.31	0.80
362027r55	9.89	0.09	9.72	10.06	-0.73	0.25	-1.21	-0.25
362027rc0	11.22	0.03	11.17	11.27	-0.75	0.02	-0.80	-0.71
362027ss4	10.77	0.02	10.73	10.81	-0.61	0.04	-0.68	-0.54
362027t95	11.58	0.03	11.52	11.63	-0.93	0.05	-1.02	-0.84
362027uj1	10.91	0.03	10.84	10.97	-0.90	0.04	-0.98	-0.83
362027uk8	11.22	0.16	10.91	11.52	-0.84	0.09	-1.02	-0.65
362027v76	10.74	0.05	10.65	10.84	-0.70	0.04	-0.77	-0.63
362028af9	9.54	0.11	9.32	9.76	1.40	1.16	-0.88	3.68
362028az5	11.30	0.16	10.98	11.62	-0.72	0.34	-1.39	-0.06
362028bt8	10.72	0.17	10.37	11.06	-0.59	0.10	-0.79	-0.39
362028bx9	10.99	0.09	10.81	11.16	-0.65	0.08	-0.80	-0.49
362028dy5	10.72	0.11	10.50	10.93	-0.35	0.49	-1.32	0.62
362028eq1	12.19	0.03	12.13	12.24	-1.18	0.05	-1.29	-1.08
362028f31	9.79	0.37	9.07	10.51	-0.39	0.60	-1.57	0.79
362028fb3	10.64	0.08	10.48	10.79	-0.26	0.20	-0.66	0.13
362028fm9	11.32	0.12	11.08	11.57	-0.71	0.22	-1.14	-0.29
362028hb1	11.02	0.05	10.92	11.13	-0.88	0.04	-0.97	-0.80
362028hu9	9.63	0.03	9.58	9.69	-0.95	0.02	-0.98	-0.92
362028jf0	10.81	0.05	10.71	10.92	-0.49	0.05	-0.59	-0.39
362028jk9	10.47	0.03	10.41	10.52	-0.65	0.06	-0.78	-0.53

Table 6: Table 5 cont.

CUSIP	$\beta_{20}$				$\beta_{23}$			
	Avg	SD	LPB	RPB	Avg	SD	LPB	RPB
362026yj9	8.76	0.18	8.41	9.12	-0.71	0.06	-0.83	-0.60
3620274l5	10.75	0.14	10.48	11.03	-0.53	0.08	-0.68	-0.37
3620274n1	10.67	0.04	10.58	10.76	-0.75	0.04	-0.83	-0.68
3620274q4	10.11	0.05	10.02	10.20	-1.52	0.05	-1.62	-1.43
3620276t6	10.09	0.16	9.77	10.41	-0.97	0.08	-1.13	-0.81
362027bx1	10.85	0.03	10.78	10.91	-0.93	0.04	-1.00	-0.85
362027fq2	10.46	0.29	9.90	11.02	-1.14	0.29	-1.72	-0.57
362027hs6	10.09	0.57	8.98	11.20	-1.20	0.34	-1.86	-0.54
362027hz0	9.61	0.03	9.56	9.66	-0.92	0.02	-0.95	-0.88
362027j47	9.93	0.04	9.85	10.01	-1.11	0.02	-1.15	-1.06
362027jp0	9.78	0.02	9.73	9.83	-0.78	0.02	-0.82	-0.75
362027kb9	9.64	0.01	9.61	9.67	-0.76	0.02	-0.79	-0.73
362027ke3	10.01	0.67	8.69	11.33	-3.10	0.91	-4.88	-1.31
362027kw3	10.74	0.03	10.67	10.80	-0.99	0.02	-1.03	-0.94
362027n26	10.35	0.03	10.30	10.40	-1.03	0.02	-1.06	-1.00
362027qy3	9.98	0.31	9.38	10.57	-1.30	0.28	-1.84	-0.76
362027r55	8.69	0.41	7.88	9.49	-0.38	0.26	-0.89	0.13
362027rc0	10.50	0.02	10.46	10.54	-0.81	0.01	-0.84	-0.78
362027ss4	9.96	0.02	9.92	10.00	-0.63	0.01	-0.66	-0.60
362027t95	11.06	0.02	11.02	11.10	-0.98	0.02	-1.02	-0.94
362027uj1	10.49	0.02	10.44	10.53	-1.45	0.04	-1.53	-1.37
362027uk8	10.50	0.05	10.40	10.59	-0.93	0.03	-0.98	-0.87
362027v76	10.11	0.02	10.07	10.15	-0.89	0.01	-0.90	-0.87
362028af9	10.85	0.24	10.38	11.31	-7.79	1.69	-11.10	-4.48
362028az5	10.64	0.10	10.45	10.83	-1.02	0.09	-1.19	-0.85
362028bt8	9.66	0.09	9.48	9.85	-0.97	0.03	-1.04	-0.91
362028bx9	9.64	0.02	9.60	9.68	-0.96	0.02	-1.00	-0.92
362028dy5	11.32	0.16	11.02	11.63	-1.46	0.12	-1.70	-1.22
362028eq1	11.49	0.02	11.45	11.52	-1.07	0.02	-1.11	-1.04
362028f31	10.18	0.38	9.43	10.92	-1.25	0.15	-1.55	-0.96
362028fb3	11.27	0.12	11.03	11.51	-1.53	0.06	-1.66	-1.41
362028fm9	10.60	0.07	10.46	10.74	-0.95	0.03	-1.01	-0.90
362028hb1	9.55	0.02	9.52	9.58	-0.76	0.01	-0.79	-0.74
362028hu9	10.65	0.04	10.57	10.72	-0.50	0.03	-0.56	-0.44
362028jff0	9.87	0.04	9.79	9.94	-0.92	0.02	-0.97	-0.88
362028jkk9	11.33	0.04	11.25	11.41	-1.34	0.02	-1.38	-1.31

Table 7: Results from two component model, summary

Ginnie Mae Pools			
Parameter	Average	LPB	RPB
$\beta_{10}$	10.50	10.31	10.69
$\beta_{13}$	-0.62	-1.11	-0.13
$\beta_{20}$	10.28	10.04	10.53
$\beta_{23}$	-1.23	-1.50	-0.96

Freddie Mac Pools			
Parameter	Average	LPB	RPB
$\beta_{10}$	10.27	9.69	10.84
$\beta_{20}$	10.58	9.39	11.77
$\beta_{23}$	-0.94	-1.38	-0.50

Table 8: Bayes Factor

CUSIP	$B_{10}$	$2 \log B_{10}$	CUSIP	$B_{10}$	$2 \log B_{10}$
362026VD5	1.78E-01	-3.45	31340ABB6	4.75E+23	109.04
3620274L5	4.37E-04	-15.47	31340AC82	1.04E-02	-9.14
3620274N1	3.57E-03	-11.27	31340AD40	5.82E-01	-1.08
3620274Q4	2.62E-02	-7.28	31340ADC2	6.82E-03	-9.98
3620276T6	1.32E-03	-13.26	31340ADF5	2.00E-02	-7.83
362027BX1	7.90E-02	-5.08	31340ADG3	4.15E-02	-6.37
362027FQ2	1.14E-01	-4.35	31340ADH1	4.57E-02	-6.17
362027HS6	1.95E-02	-7.88	31340ADW8	3.54E+00	2.53
362027HZ0	7.27E-03	-9.85	31340AEL1	3.91E-02	-6.48
362027J47	1.07E-01	-4.47	31340AEU1	2.96E-01	-2.43
362027JP0	1.83E-02	-8.00	31340B5H8	8.16E-02	-5.01
362027KB9	1.87E-02	-7.96	31340B5P0	4.80E-01	-1.47
362027KW3	3.19E-03	-11.50	31340B5T2	1.04E-02	-9.14
362027N26	2.56E-03	-11.94	31340B5Y1	2.57E-01	-2.72
362027QY3	3.33E+00	2.40	31340B6B0	1.33E-01	-4.03
362027R55	4.74E-01	-1.49	31340B6J3	1.20E-01	-4.23
362027RC0	7.12E-04	-14.49	31340B6V6	2.13E-01	-3.09
362027SS4	1.53E-02	-8.36	31340B7L7	1.06E-01	-4.49
362027T95	1.31E+03	14.36	31340CAA5	2.94E+48	223.21
362027UJ1	3.47E-02	-6.72	31340CAC1	4.06E-01	-1.80
362027UK8	5.24E-03	-10.50	31340CAF4	1.43E-01	-3.88
362027V76	2.88E-02	-7.09	31340CAH0	4.29E-01	-1.69
362028A77	1.68E-01	-3.57	31340CAJ6	3.92E-02	-6.48
362028AZ5	3.61E+03	16.38	31340CAN7	5.23E-02	-5.90
362028BT8	4.55E-04	-15.39	31340CB55	3.81E-01	-1.93
362028BX9	5.94E-03	-10.25	31340CBA4	6.61E-02	-5.43
362028DY5	8.07E+01	8.78	31340CBG1	3.22E-02	-6.87
362028EQ1	1.78E-02	-8.05	31340CBJ5	9.66E-03	-9.28
362028F31	8.50E-01	-0.32	31340CBK2	5.89E-02	-5.66
362028FB3	3.44E-01	-2.13	31340CBL0	1.54E-02	-8.35
362028FM9	1.95E-03	-12.48	31340CBP1	6.19E-01	-0.96
362028HB1	3.61E-03	-11.25	31340CBR7	9.31E-02	-4.75
362028HU9	5.64E-02	-5.75	31340CBS5	1.74E-02	-8.10
362028JF0	5.10E-03	-10.56	31340CBW6	1.17E-02	-8.90
362028JK9	6.18E+03	17.46	31340CE60	5.18E-01	-1.32
31340CEJ2	3.45E-01	-2.13	31340CER4	1.24E-01	-4.18
31340CEL7	1.98E-01	-3.24	31340CFR3	2.07E+07	33.69

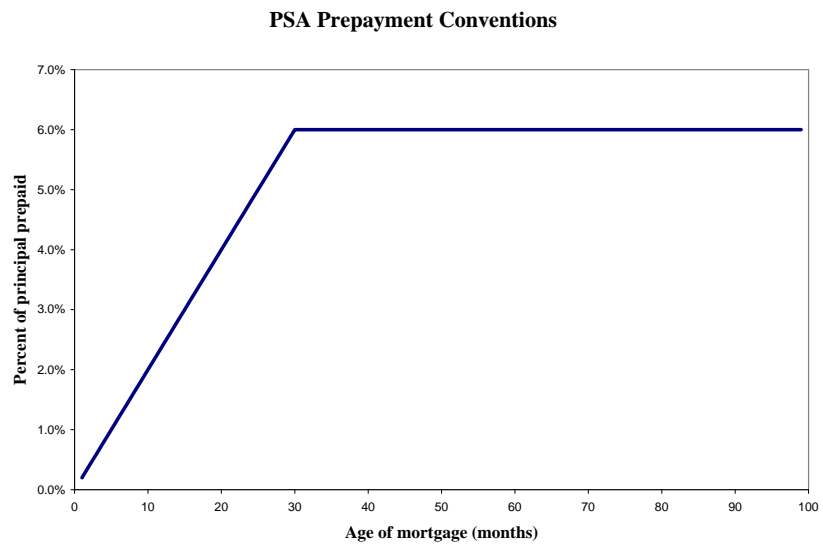


Figure 1: PSA Prepayment Conventions

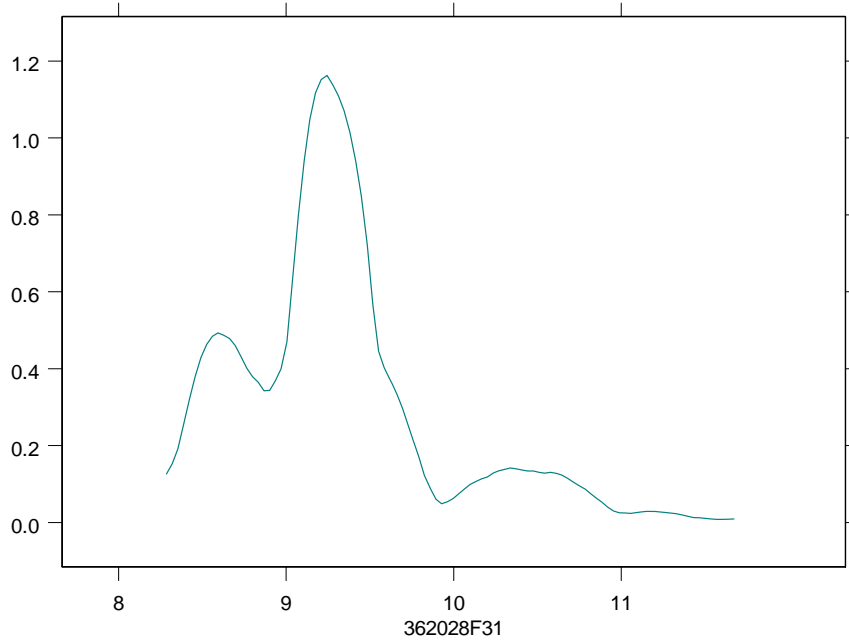
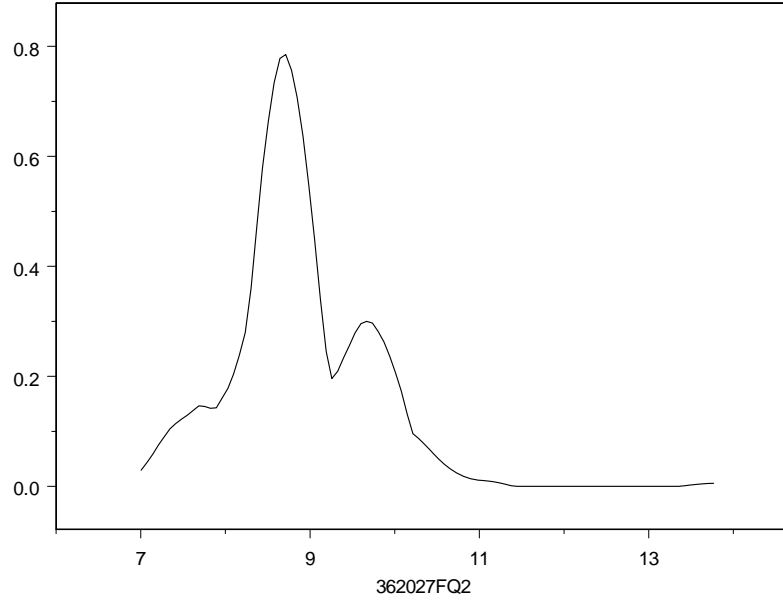


Figure 2: Kernel density plots. The horizontal axis is  $y_t = \ln(AP_t)$ , the logarithm of the total money paid at the end of month  $t$ . The number below the horizontal axis is the pool CUSIP identifier.

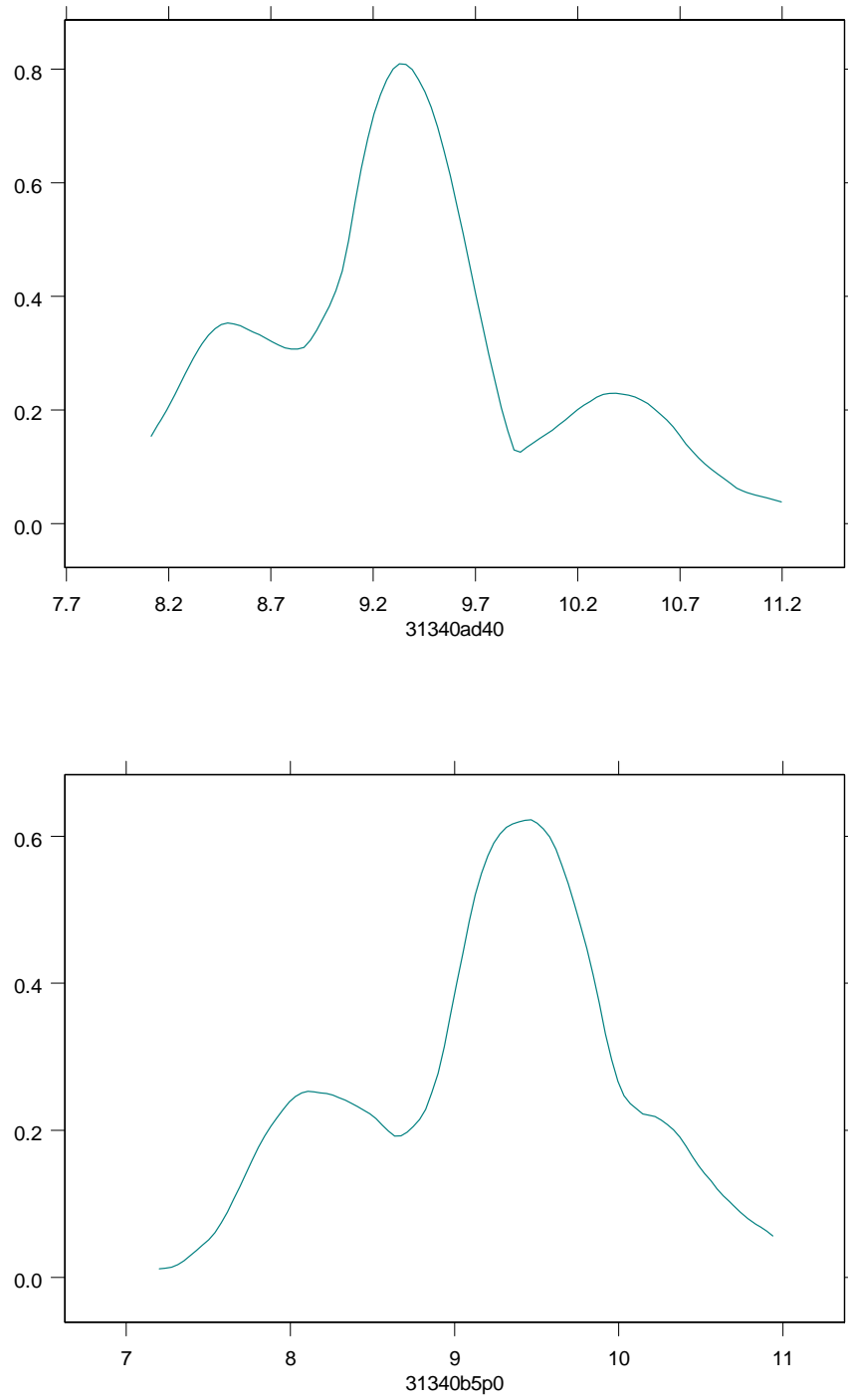


Figure 3: Kernel density plots cont.



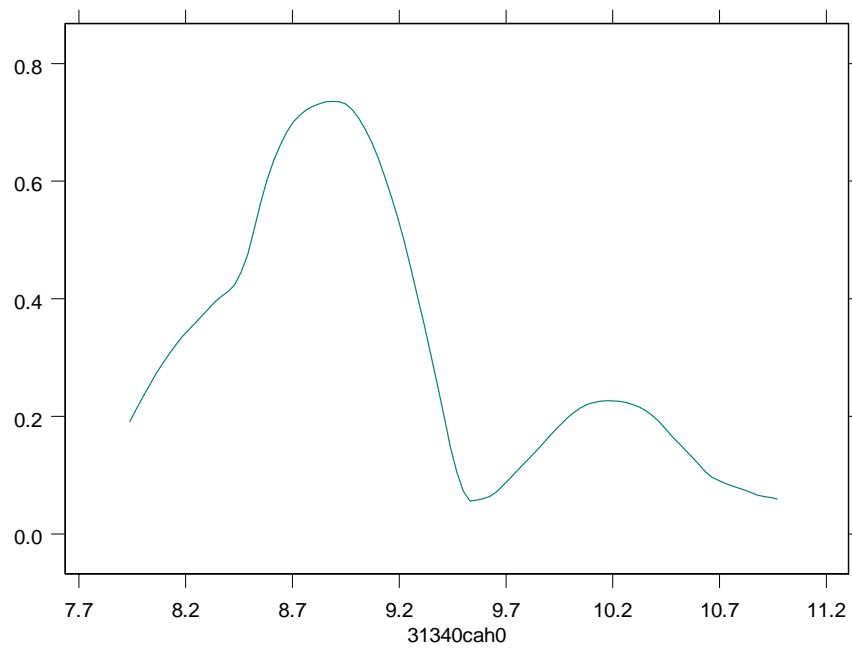
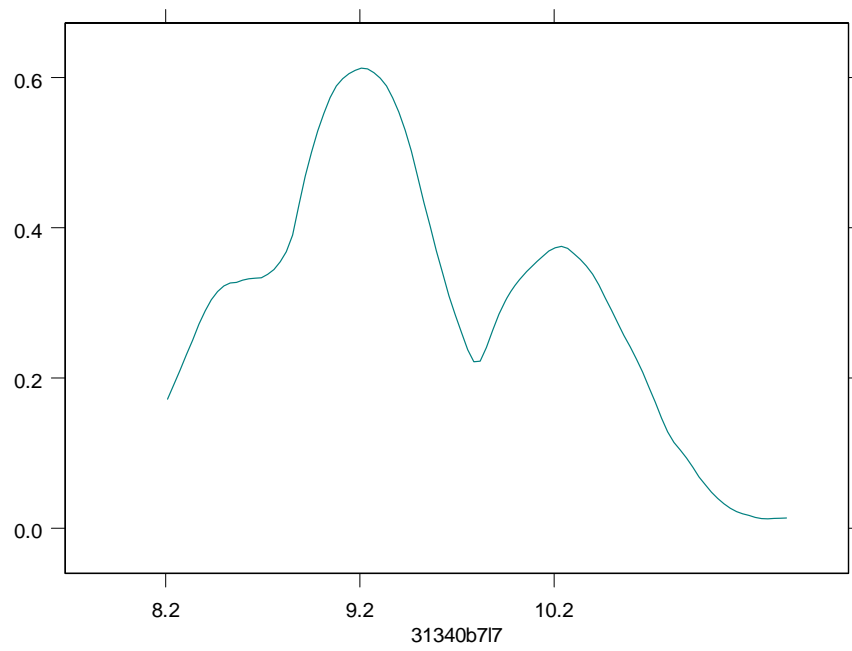


Figure 4: Kernel density plots cont.

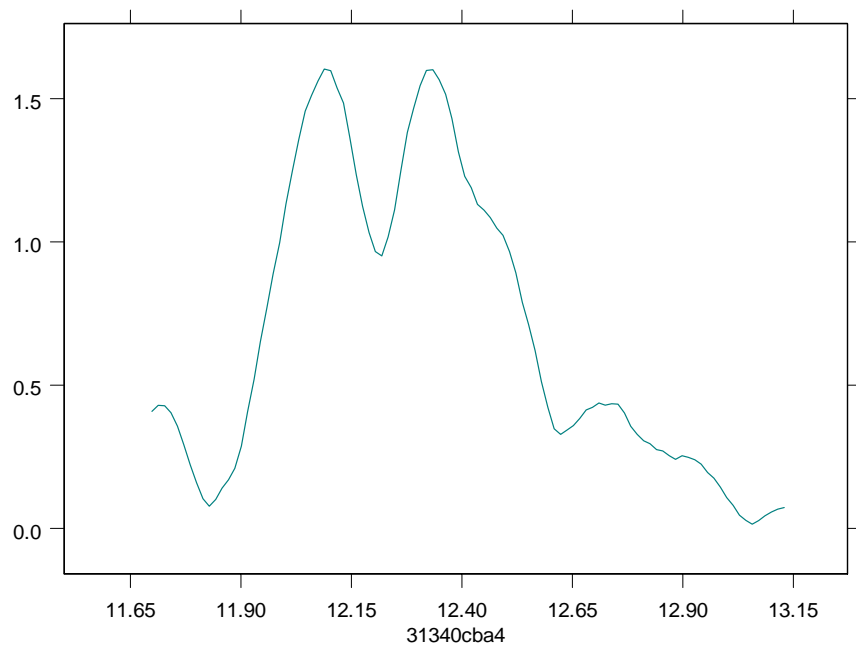
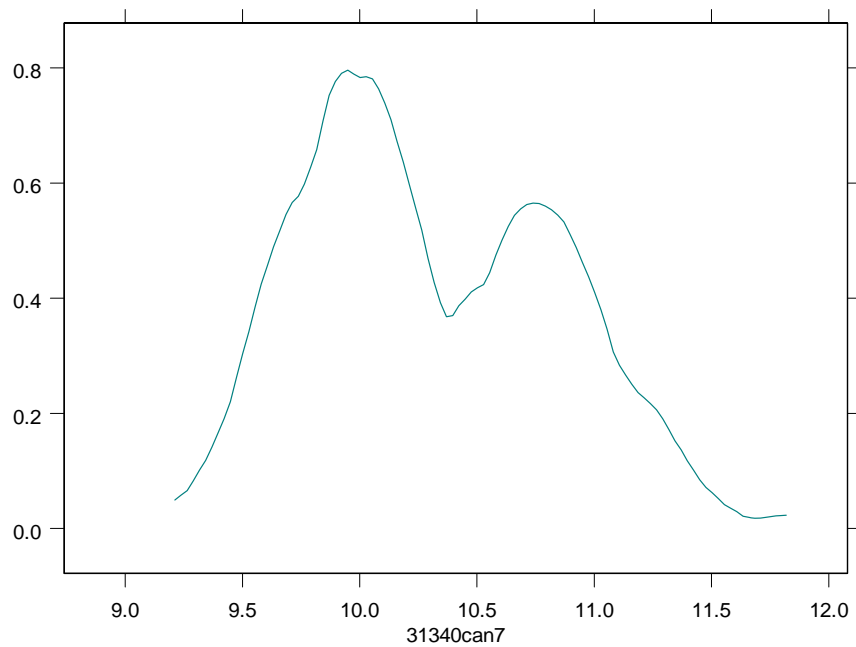


Figure 5: Kernel density plots cont.

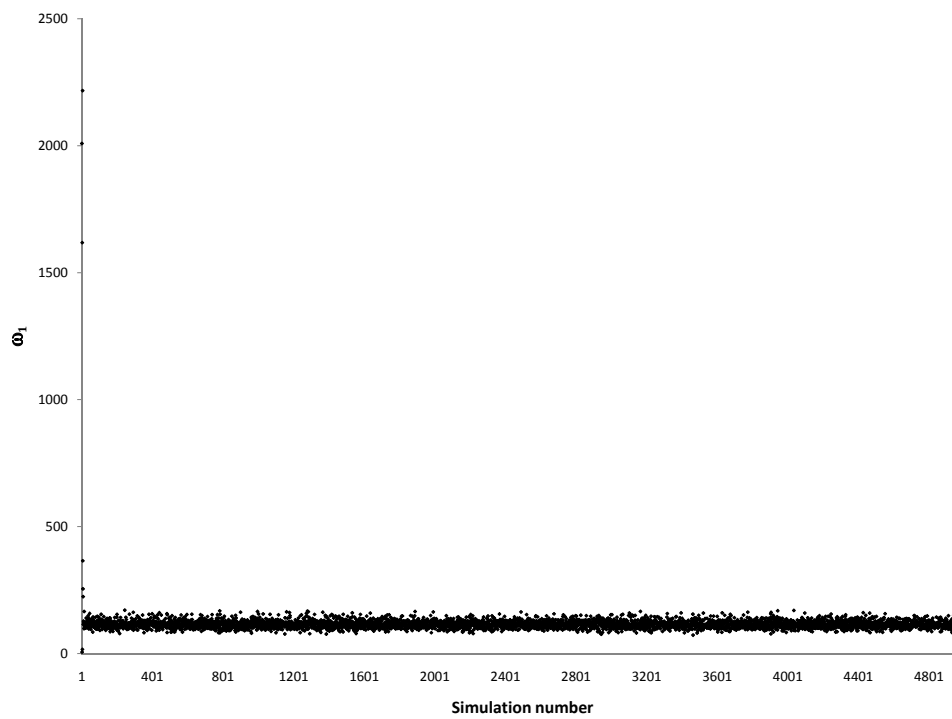


Figure 6: Actual simulated points for  $\omega_1$  for one of the mortgage pools for the two component model

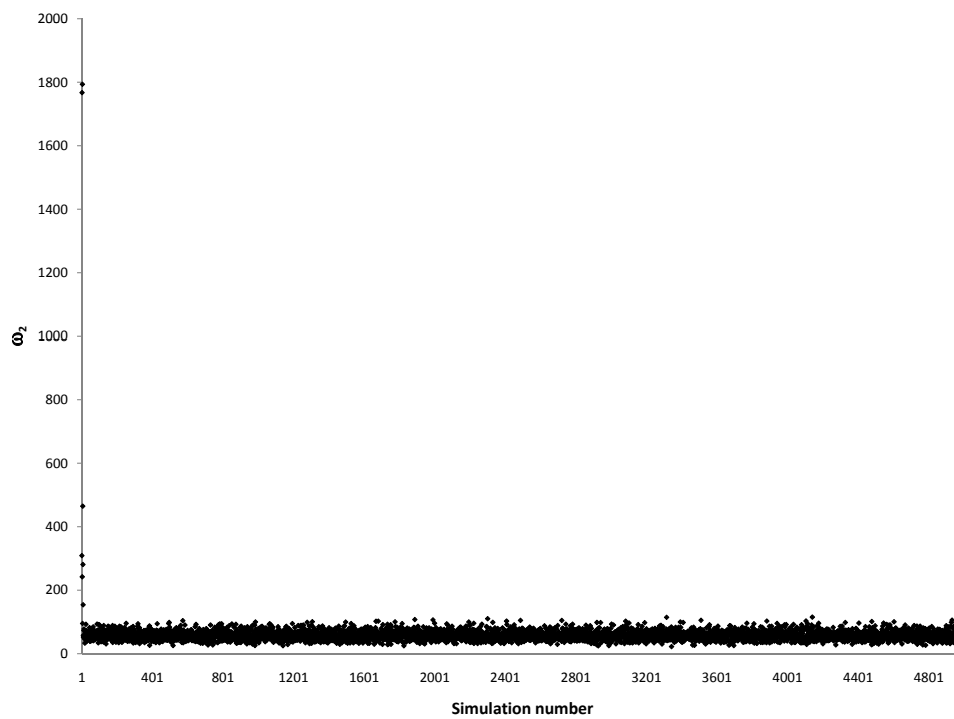


Figure 7: Actual simulated points for  $\omega_2$  for one of the mortgage pools for the two component model

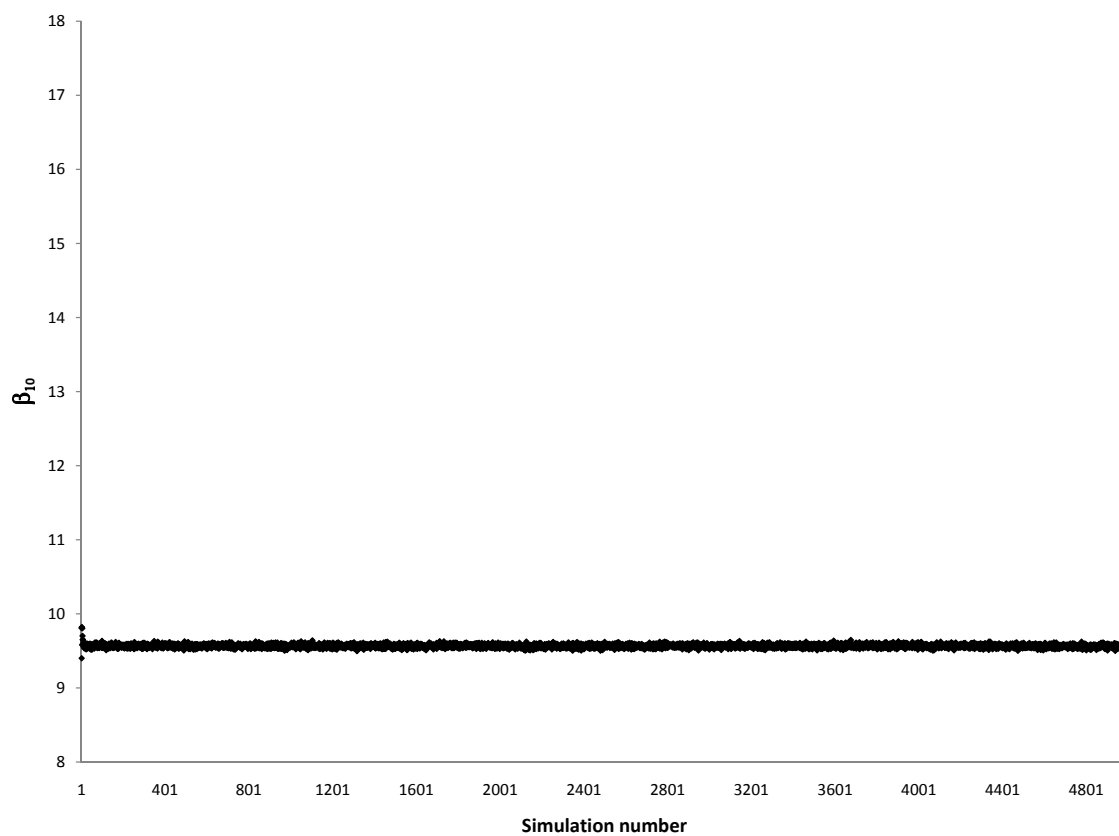


Figure 8: Actual simulated points for  $\beta_{10}$  for one of the mortgage pools for the two component model

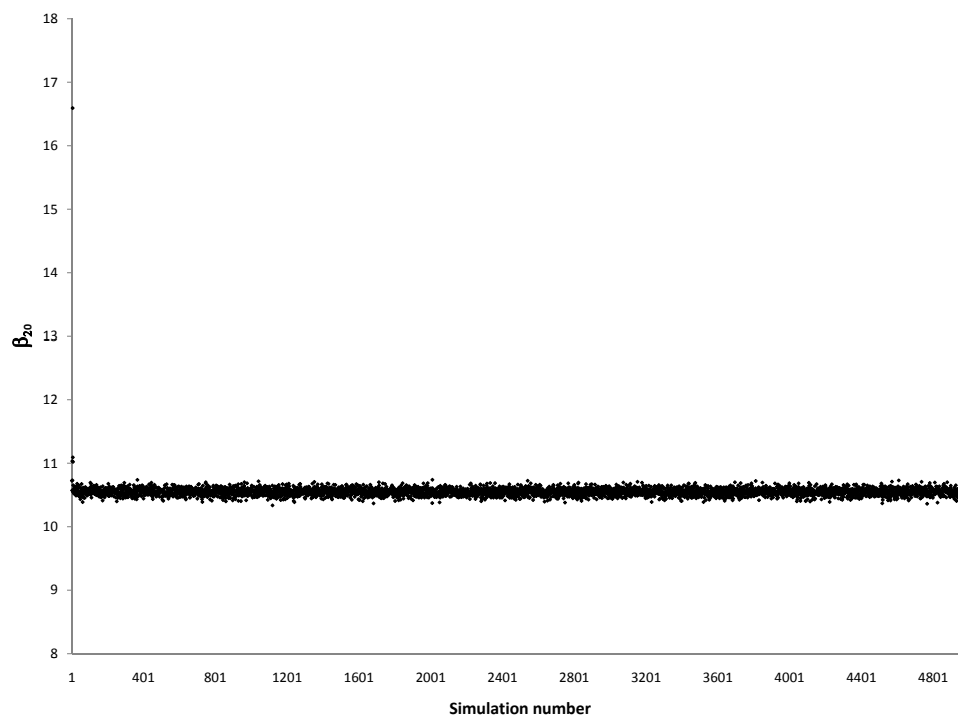


Figure 9: Actual simulated points for  $\beta_{20}$  for one of the mortgage pools for the two component model

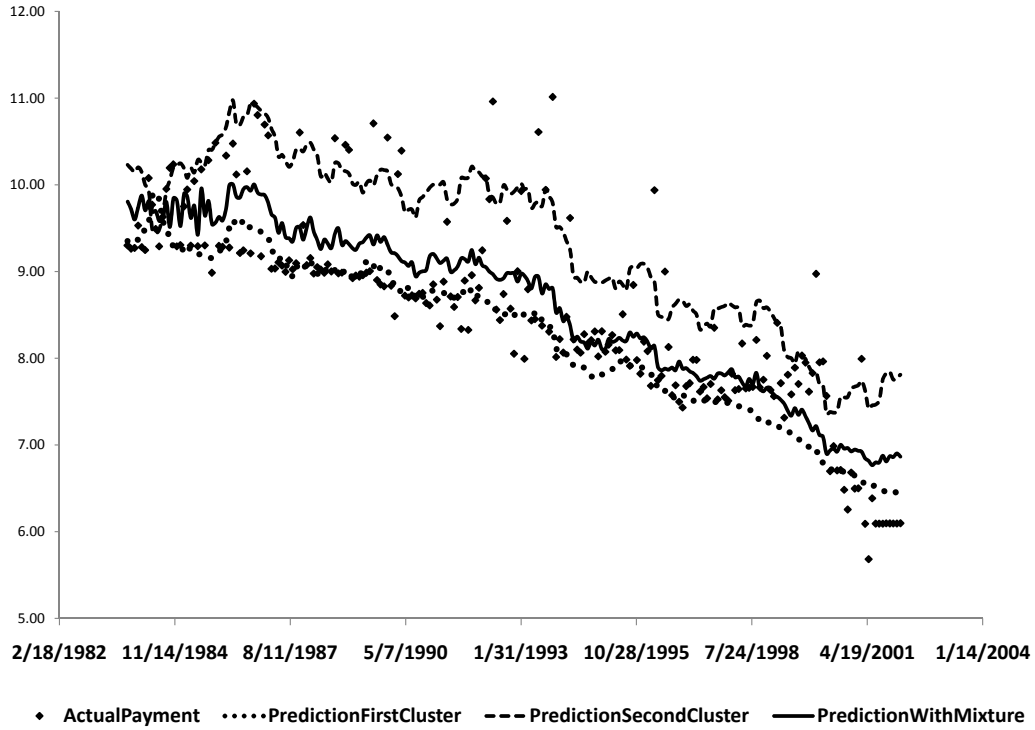


Figure 10: Actual vs forecasted prepayment for one of the mortgage pools for the two component model. PredictionFirstCluster is the estimated prepayment using the expected posterior parameters of the first mixture component. PredictionSecondCluster is the estimated prepayment using the expected posterior parameters of the second mixture component. PredictionWithMixture is the estimated prepayment using the weighted estimated prepayments from both components.

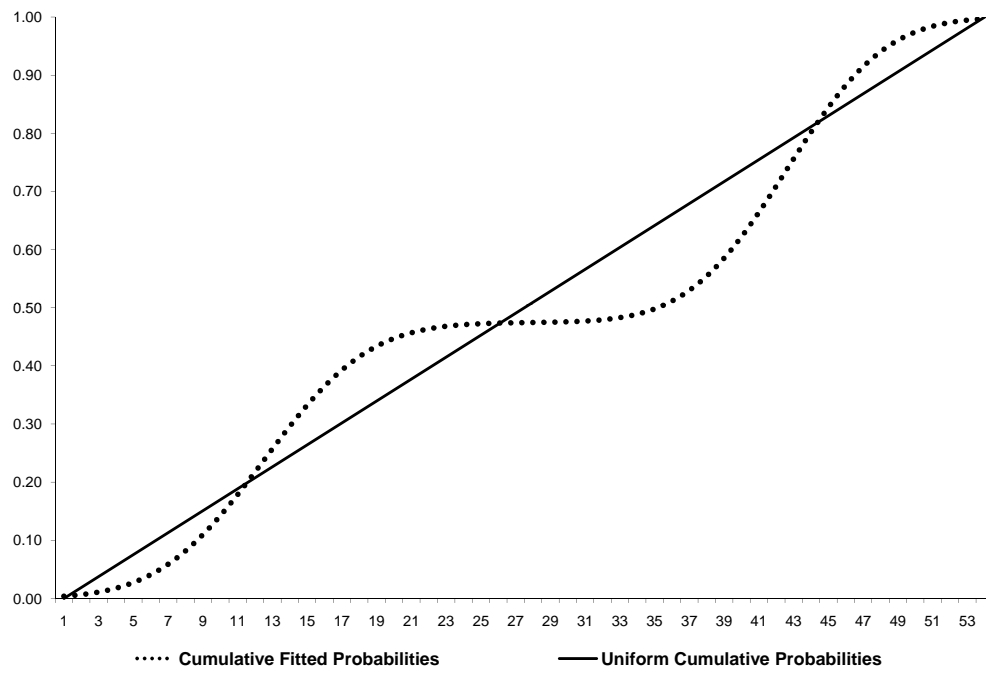


Figure 11: Fitted and uniform(0,1) cumulative probabilities for the two component model



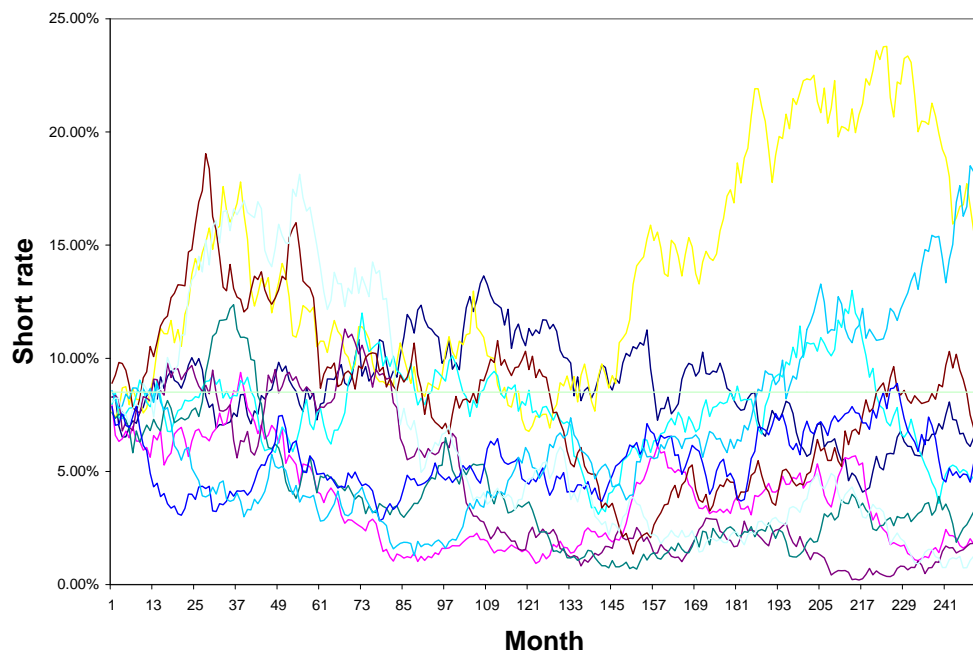


Figure 12: Paths of simulated short rates

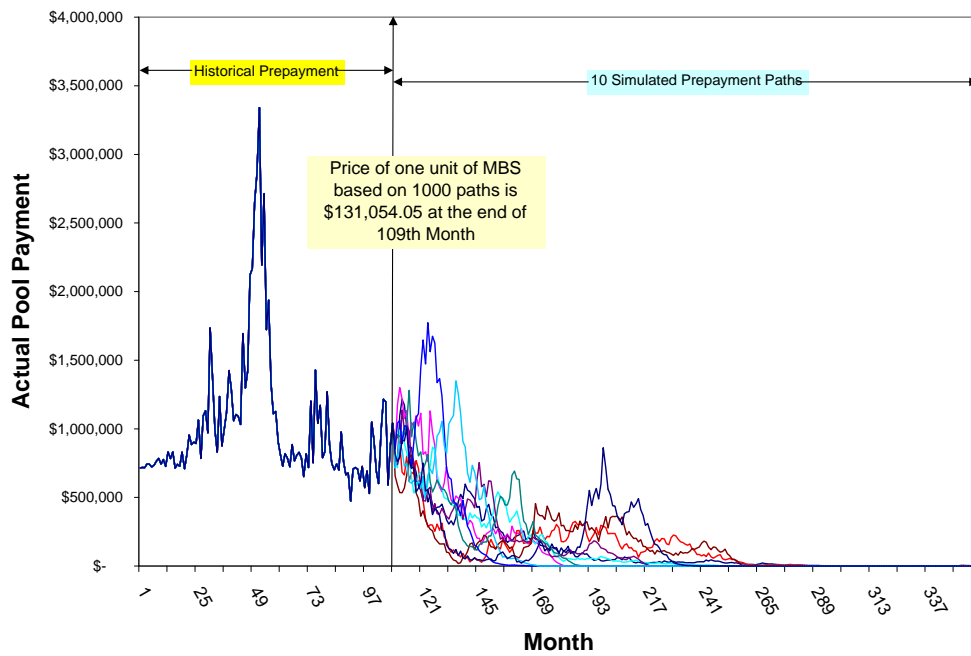


Figure 13: Freddie Mac pool 31340CEJ2. Initial pool amount is \$93,073,150. 1 Unit MBS = 0.5% claim on the cash flow from the pool.