

Bayesian Modeling of Mortgage Prepayment Rates ¹

Elmira Popova²
Operations Research
The University of Texas at Austin
Austin, TX 78712, USA
elmira@mail.utexas.edu

Ivilina Popova
Ceiba Asset Management, LLC
106 E 6th Street, Suite 650
Austin, TX 78701, USA
ipopova@ceibacapital.com

Edward George
Department of Statistics
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104, USA
edgeorge@wharton.upenn.edu

Abstract

This paper proposes a novel approach for modeling prepayment rates of pools of mortgages. Our goal is to establish a model that will give a good prediction for prepayment rates for 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 mixture model and construct a Markov Chain Monte Carlo algorithm to estimate the parameters. We assess the model on a large data set from the Bloomberg Database.

Keywords: Finance, Markov Chain Monte Carlo, Bayesian statistics

¹This research has been partially supported by grant #003658-0763 from the State of Texas Advanced Research Program

²Corresponding author

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 principle or early payoff 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. The buyers of such structured products would like to know in advance the size of the incoming prepayments (if any). To obtain the correct price one needs to know a forecast of the prepayment.

One possible model of the prepayment is if we pretend that the borrower holds a call option on the loan with exercise price equal to the outstanding balance (i.e. it is a time varying strike). Under optimal exercising conditions a mortgage can be priced as a callable bond. One would expect that the holder will prepay when the refinancing rate is below the mortgage rate. However, the empirical evidence, which shows very different behavior [4], does not support such a model. As has been documented, homeowners often prepay when it is not optimal to do so and vice versa.

The recent literature attempts to model this partially “irrational” behavior. The prepayment activities can be classified as either interest rate related or non-interest rate related. The interest rate related activities (optimal prepayment) occur when the homeowners act in order to minimize the market value of the loan. The 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 prepayment. Dunn and McConnell [4] pioneered a model based on standard contingent claim pricing theory. In their model the prices of the mortgage backed securities (MBS) and prepayment behavior are de-

terminated together. They assume Cox-Ingersoll-Ross interest rate model [5], introduce a suboptimal prepayment as a Poisson event, and using non-arbitrage argument derive a partial differential equation for the price of the security that can be solved numerically. Suboptimal prepayment behavior was first documented in this study.

A second approach is empirical - based on historical information. Schwartz and Torous [10] model the prepayment rate as a function of different explanatory variables, like 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.

Richard and Roll [8] describe the prepayment model used by Goldman Sachs. They consider four important effects: refinancing incentive, age of the mortgage, seasonality and burnout. The conditional prepayment rate is determined as a product of functions of the above factors. A non-linear least squares optimization procedure gives the estimated values of the parameters. Stanton [11] 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 time. This produces prepayment behavior that is consistent with the burnout effect.

All of the existing models 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. Recently, Stanton [12], 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 that are 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.

None of the previous models has dealt with pool level predictions, and this is our main

research objective. Section §2 describes the nature of the raw data and how the final data set is constructed. In sections §3 and §4 we present two prediction models: a Bayesian mixture of regressions and a Bayesian probit mixture of regressions. In each of them we present an extensive empirical study.

2 Data Description

We consider historical data for 78 pools of mortgages equally split between issues of Freddie Mac and Ginnie Mae, the two major mortgage providers. To ensure the homogeneity of the data, we exclusively focused on 30 year fixed rate single family 8% coupon 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 principle is paid in the first month, increase by 0.2% for the following 29 months, and flattens at 6% until maturity, see [13].

The age of the mortgages varies from 5 to 25 years, and consequently, the number of data points for each pool from 60 to 300. In addition to the information provided by Bloomberg, we gathered historical long and short term interest rates from <http://www.stld.frb.org/fred/>. The steps to construct the final data set are given in the Appendix, see §6.

In our analysis we model the actual payment (AP) (in dollar amount) at the end of each month t , $AP(t)$. Preliminary analysis of the data indicates that for the most of the pools the data exhibits a split in two groups.

INSERT FIGURE 1 HERE

Figure 1 plots histograms of the logarithm of the total money paid at the end of the month for four pools of mortgages. Note that there is a group of “small” prepayments

concentrated on the lower part of each graph and a group of “large” prepayments on the upper part of each plot.

This grouping 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 (refinancing, selling the house, etc.) There are probably certain events that trigger the “small” or the “large” prepayment behavior. If we knew these events and could gather data associated with them, then we might be able to reasonably well predict the next month prepayment. However, up to now, there is no research regarding such events, their existence and data availability.

The Figure 1 histograms suggest that the distribution of AP is not unimodal and may be better described with a bimodal model such as a mixture of normals. In particular, we consider model the AP values as the realizations of n independent random variables, x_1, x_2, \dots, x_n from a 2-component mixture

$$f(x_i) = pf_1(x_i) + (1 - p)f_2(x_i), i = 1, \dots, n$$

where $f_i(\cdot)$ is the probability density function of a Normal distribution with parameters μ_i and σ_i , $i = 1, 2$. We denote the unknown parameters of the mixture by $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p)$.

To fit and draw inference from such a model, we take a Bayesian approach which describes the uncertainty about unknown parameters with prior probability distributions. We proceed as in Diebolt, J. and C.P. Robert [9] and Robert, Hurn and Justel [7]. As is well-known, Dempster, Laird and Rubin [1], any mixture model can be expressed in terms of *missing* or *incomplete* data as follows. If, for $1 \leq i \leq n$, z_i is a 2 - dimensional vector indicating to which component x_i belongs, such that $z_{ij} \in \{0, 1\}$, $i = 1, \dots, n$, $j = 1, \dots, 2$, the density of the completed data (x_i, z_i) is

$$\prod_{j=1}^2 p_j^{z_{ij}} f_j^{z_{ij}}(x_i)$$

Using this representation, the parameters, θ , of the mixture model can be straightforwardly estimated by a Markov Chain Monte Carlo (MCMC) algorithm, namely the Gibbs Sampler. The general description of such an algorithm is (see [14]):

Starting with an initial value θ_0

• **at step m :**

Generate $z^{(m)} \sim f(z|x, \theta^{(m)})$,

Generate $\theta_1^{(m+1)} \sim \pi(\theta_1|x, z^{(m)}, \theta_2^{(m)}, \dots, \theta_s^{(m)})$,

\vdots

Generate $\theta_s^{(m+1)} \sim \pi(\theta_s|x, z^{(m)}, \theta_1^{(m+1)}, \dots, \theta_{s-1}^{(m+1)})$

In the above description, x is the observed data and $\pi(\theta|x, z)$ and $f(z|x, \theta)$ are the full conditional distributions. The above algorithm is geometrically convergent, see Diebolt and Robert, [9].

3 Bayesian Mixture of Regressions

The preliminary analysis of the data suggests that there are two main groups - "small" and "large" prepayers. We would like to model this structure by incorporating covariates described in the literature as important. Schwartz and Torous [10] define the following set of variables which are significant in predicting prepayment behavior:

- X_1 = the difference between the mortgage rate and the short term interest rate
- $X_2 = X_1^3$
- X_3 captures the burnout effect - it is the logarithm of the ratio of the dollar amount of the pool outstanding at time t , to the pool's principle which would prevail at t in the absence of prepayments

- X_4 models the seasonality effect. It equals 1 for the months of May, June, July, and August, and 0 for September, October, November, December, January, February, March, and April.
- X_5 is the spread (difference) between the long and short term interest rates.

As a forecasting model we define a Bayesian mixture of regressions similar to [7], where $Y = \ln(AP(t))$ is the logarithm of total money paid at the end of the month, and X_1, X_2, X_3, X_4, X_5 are the covariates defined above. The model is:

$$Y \sim pN\left(\sum_{i=0}^5 U_i X_i, W_1\right) + (1-p)N\left(\sum_{i=0}^5 V_i X_i, W_2\right)$$

with parameters $\{U_0, U_1, U_2, U_3, U_4, U_5, V_0, V_1, V_2, V_3, V_4, V_5, W_1, W_2, p\}$. Note that we use W_1 and W_2 to represent the precision parameters, rather than the variances, of the two normal distributions that constitute the mixture. (*Precision* = $1/\text{Variance}$).

Letting $M_1 = (U_0, U_1, U_2, U_3, U_4, U_5)$ and $M_2 = (V_0, V_1, V_2, V_3, V_4, V_5)$ denoted covariate coefficients corresponding to each of the mixture components, we consider the following default priors for the unknown parameters. We assumed improper joint prior distribution for (M, W) , $\xi(m, w) = 1/w, w > 0$ and $\text{Beta}(\alpha, \beta)$ as a prior distribution for p . Let X be the matrix

$$X = \begin{Bmatrix} 1 & X_1^1 & X_2^1 & X_3^1 & X_4^1 & X_5^1 \\ 1 & X_1^2 & X_2^2 & X_3^2 & X_4^2 & X_5^2 \\ \dots & & & & & \\ 1 & X_1^N & X_2^N & X_3^N & X_4^N & X_5^N \end{Bmatrix}$$

Defining

$$\hat{m} = (X'X)^{-1} X'Y \quad (1)$$

$$\hat{s}^2 = \frac{1}{n-2} (Y - X\hat{m})' (Y - X\hat{m}) \quad (2)$$

$$n_1 = \text{Number of data points in cluster 1}$$

$$n_2 = \text{Number of data points in cluster 2}$$

$$n_1 + n_2 = N$$

then under our priors, the marginal posterior distributions of the mixture parameters are (see [9]):

$$\begin{aligned}
 M_i &\sim \text{Multivariate } t \left(n_i - 2, \hat{m}_i, \frac{1}{\hat{s}_i^2} X' X \right) \\
 W_i &\sim \text{Gamma} \left(\frac{n_i - 2}{2}, \frac{n_i - 2}{2} \hat{s}_i^2 \right) \\
 p &\sim \text{Beta} (n_1 + \alpha, n_2 + \beta) \\
 &i = 1, 2
 \end{aligned} \tag{3}$$

The Gibbs sampling algorithm for this setup consists of the following steps (see [9])

1. Start with initial values of the parameters $\theta^0 = (M_1^0, M_2^0, W_1^0, W_2^0, p^0)$
2. Allocate each Y_i to the first or second mixture based on odds ratio

$$\frac{P[z_i = 0 | \theta, y]}{P[z_i = 1 | \theta, y]}$$

This part of the algorithm will generate $Z_i = 0$ if the odds ratio is less than 1 and $Z_i = 1$ if it is greater than 1.

3. Now each the points has been assigned either to cluster 1 or cluster 2. Given that we are in cluster $i, i = 1, 2$, we simulate θ from the marginal distributions (3)
4. Repeat the above steps M times (number of simulation runs)

We used the uniform random number generator written by P. L'Ecuyer (see [6]) and non uniform random number algorithms from Devroye, [3].

The length of our simulation runs is 5000, the algorithm is implemented in C++ and run on 2.37GHz Pentium. As with any MCMC algorithm, there is a transient period at the beginning of the simulation. Figure 2 shows the actual simulated points for one of the parameters of one of the pools of mortgages. Based on the observed patterns for the moving average process of the simulated values we dropped the first 1000 simulated values

and used only the last 4000 simulated values for estimation and inference. We observed similar patterns with the remaining 77 pools of mortgages.

INSERT FIGURE 2 HERE

Before presenting our model estimates, we emphasize that we have treated all the parameters as random variables and the estimates that we present here are the expected values of the corresponding posterior distributions. Note that a further advantage of the Bayesian approach is that it also produces full posterior and predictive distributions, which then can be used to construct future point forecasts, probability intervals, or to be used in simulation studies. In our current analysis the predictions are based on expected posterior values only.

Given that we were able to construct and compute the expected posterior values for each of the model's parameters, we examine the forecast quality for each of the pools and for all pools together. Figure 3 compares the actual prepayment with the predicted prepayment for one pool of mortgages.

INSERT FIGURE 3 HERE

The lower solid curve is $\sum_{i=0}^5 \hat{U}_i X_i$, where \hat{U}_i are the expected values of the parameters posterior distributions. The upper solid curve is $\sum_{i=0}^5 \hat{V}_i X_i$, where \hat{V}_i are the expected values of the parameters posterior distributions. The model will forecast an "average" prepayment with probability \hat{p} and a "high" prepayment with probability $1 - \hat{p}$. A point forecast could be the weighted average of these two points (the middle line). We use that forecast from the mixture to compute the associated error.

For every pool of mortgages we compute the mean of the associated residuals. Next we construct the distribution of the calculated means. Figure 4 shows the box plot of the mean of the residuals (measured as the actual minus the predicted).

The descriptive statistics of the set of mean residuals are given in Table 1. Note that the mean error distribution exhibits a long right tail. This suggests that the model may

Mean	-0.05
Median	-0.09
1st Quartile	-0.26
3rd Quartile	-0.01
Standard Deviation	0.5
Skewness	3.17
Kurtosis	13.75

Table 1: Descriptive statistics for the distribution of the error means for the Bayesian mixture of regressions model

be systematically underestimating the largest prepayment months. In the next Section, we consider an elaboration of the model which yields a more symmetric distribution of the error means, and produces varying forecasts for the probability of prepayment.

INSERT FIGURE 4 HERE

4 Probit mixture of regressions

The model described in the previous section gives interesting insight into prepayment behavior: it is either “high” or “average” and these two categories are very distinct. This is an important observation, as none of the existing models incorporate such a split. Given future values of the covariates we can produce a forecast that will say : with probability p it will be a “high” prepayment point and with probability $1 - p$ it will be an “average” prepayment point. However, it may be the case that this probability is changing over time. For example, in Figure 3, the frequency of larger prepayment months appears to be larger in the middle months and smaller at the extremes. To allow for this feature, we elaborate our previous model to allow for such changing probabilities. As we will see, such an elaboration allows us to forecast p as well (given the covariates).

Our model elaboration entails incorporating a probit regression model for p within the

mixture of two regressions described in the previous section.

$$Y \sim \Phi \left(\sum_{i=0}^5 \beta_i X_i \right) N \left(\sum_{i=0}^5 U_i X_i, W_1 \right) + \left[1 - \Phi \left(\sum_{i=0}^5 \beta_i X_i \right) \right] N \left(\sum_{i=0}^5 V_i X_i, W_2 \right)$$

with a parameter set $\{U_0, U_1, U_2, U_3, U_4, U_5, V_0, V_1, V_2, V_3, V_4, V_5, W_1, W_2, \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ for which we define improper joint prior distribution for (M, W) and improper prior for β .

The idea is to extend the MCMC algorithm to include the estimation of the probit regression parameters. The only reference in the literature is the work by Albert and Chib, [2] where they construct a MCMC algorithm to estimate the parameters of a probit regression model $p_i = \Phi(XB)$, where Φ is the standard Normal cumulative distribution function, X are the covariates, and $p_i, i = 1, \dots, N$ is the probability of success of N independent Bernoulli random variables Y_1, \dots, Y_N . The algorithm is once again based on the introduction of N latent variables $Z_1, \dots, Z_N, Z_i \sim N(XB, 1)$. If we define

$$Y_i = \begin{cases} 1 & \text{if } Z_i > 0 \\ 0 & \text{if } Z_i \leq 0 \end{cases}$$

then $p_i = P[Y_i = 1] = \Phi(XB)$. They shown that the marginal distribution of (B, Z) given the data (Y) is proportional to

$$\prod_{i=1}^N \{I(Z_i > 0) I(Y_i = 1) + I(Z_i \leq 0) I(Y_i = 0)\} \phi(Z_i; XB, 1)$$

where $I(A)$ is the indicator function for the event A , i.e. it equal to 1 if the event A occurred and 0 otherwise, and $\phi(Z_i; XB, 1)$ is the *Normal* $(XB, 1)$ probability density function.

A distinctive difference in our case is the fact that we do not observe the class participation but rather the continuous variable that equals to the logarithm of the total money paid at the end of each month. We will take full advantage of the Markov Chain Monte Carlo modeling and will simulate the participation using the following model:

Define

$$Y \sim \text{Normal}(\mu_\delta, \sigma_\delta^2) \quad (4)$$

$$\delta = 1(Z > 0) = \begin{cases} 1 & \text{if } Z > 0 \\ 0 & \text{if } Z \leq 0 \end{cases} \quad (5)$$

$$Z \sim \text{Normal}(BX, 1) \quad (6)$$

$$\mu_0 = \sum_{i=0}^5 U_i X_i \quad (7)$$

$$\mu_1 = \sum_{i=0}^5 V_i X_i \quad (8)$$

$$\sigma_i^2 = 1/W_i, i = 1, 2 \quad (9)$$

Then given $\delta_i, Y, i = 1, \dots, n$ we simulate $U_0, U_1, U_2, U_3, U_4, U_5, V_0, V_1, V_2, V_3, V_4, V_5, W_1, W_2$; given $U_0, U_1, U_2, U_3, U_4, U_5, V_0, V_1, V_2, V_3, V_4, V_5, W_1, W_2, Y$ we simulate $\delta_1, \dots, \delta_n$.

If we denote by $\Theta = (\mu_0, \mu_1, \sigma_0^2, \sigma_1^2)$ the marginal distribution of (Z, B) given (Y, Θ) is

$$\pi(Z, B|Y, \Theta) \propto C\pi(B) \prod_{i=1}^N \left\{ 1(Z_i > 0) \phi_{(\mu_1, \sigma_1^2)}(y_i) + 1(Z_i \leq 0) \phi_{(\mu_0, \sigma_0^2)}(y_i) \right\} \phi_{(XB, 1)}(Z_i)$$

where $\phi_{(\mu, \sigma^2)}(\bullet)$ is the *Normal* (μ, σ^2) probability density function.

We need to be able to simulate from that distribution. In Albert and Chib case it is the truncated Normal either at the left or right by 0 depending on the observed value of the binary data (see Figure 5).

INSERT FIGURE 5 HERE

In our case the distribution to the left of 0 does not have to be the same Normal distribution. We propose the following sampling algorithm:

Define

$$a_0 = \phi_{(\mu_0, \sigma_0^2)}(y_i) \quad (10)$$

$$a_1 = \phi_{(\mu_1, \sigma_1^2)}(y_i) \quad (11)$$

$$\alpha = \frac{a_0 \Phi_{(XB,1)}(0)}{a_0 \Phi_{(XB,1)}(0) + a_1 [1 - \Phi_{(XB,1)}(0)]} \quad (12)$$

$$\beta = \frac{a_1 [1 - \Phi_{(XB,1)}(0)]}{a_0 \Phi_{(XB,1)}(0) + a_1 [1 - \Phi_{(XB,1)}(0)]} \quad (13)$$

Then given the observed data y_i the sampling algorithm for Z is:

- Step 1: With probability α simulate Z from $N(XB, 1)$ truncated at left by 0.
- Step 2: With probability $\beta = 1 - \alpha$ simulate Z from $N(XB, 1)$ truncated at right by 0.

To gauge the quality of fit of this elaborated model, we computed the residual differences between the observed AP values and the corresponding posterior mean. Descriptive statistics of these residuals are reported in Table 2, and a box plot of these residuals is provided in Figure 6.

INSERT FIGURE 6 HERE

Comparison with Table 1 shows that the standard deviation has dropped very substantially from 0.5 to 0.12, a reduction of more than 75%. Thus, the overall fit of the enhanced model is much better. Comparison of Figure 4 with Figure 6 reveals that the residuals have become much more symmetric. This is also evidence by the drastic change in the coefficient of skewness from 3.17 to -1.19 . By allowing the p to vary, the elaborated model has shifted the overall posterior mean upwards the larger prepayment months.

Mean	-0.02
Median	-0.02
1st Quartile	-0.09
3rd Quartile	0.03
Standard Deviation	0.12
Skewness	-1.19
Kurtosis	19.87

Table 2: Descriptive statistics for the distribution of the error means for the probit mixture of regressions model

5 Conclusion

In this paper, we have proposed using two component normal regression mixture models to describe the apparent bimodal distribution of prepayments within mortgage pools over time. Distinct linear regression functions of observed covariates are used to model the means of the two components. We first consider a model where the mixture probabilities are constrained to be fixed over time. We consider an elaboration of this model where the mixture probabilities can vary over time, and are modeled by a separate probit regression of the observed covariates. To fit these multiparameter nonlinear models, we take a Bayesian statistical approach where the uncertainty about all the unknown parameters is described by prior distributions. For this setup, an MCMC algorithm is constructed and used to carry out all the computations. Although the fixed mixture weight model appears to fit the observed data reasonably well, the elaborated model with varying mixtures weights, appears to offer substantial improvement.

We feel that our models are a good start towards the modeling of the prepayment rate process. Naturally, the extent to which these 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 information 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.

6 Appendix

The construction of the data set consists of several steps.

1. Compute the conditional prepayment rate using the formula

$$CPR = PSA(standard).PSA(historical)/100,$$

where $PSA(historical)$ comes from the data set.

2. Compute the single monthly mortality rate, $SMM = 1 - (1 - CPR)^{\frac{1}{12}}$.
3. For $t = 0, \dots, 360$ compute the monthly payment as $MP = \frac{FaceValue \left[\frac{0.085}{12} \right] \left[1 + \frac{0.085}{12} \right]^t}{\left[1 + \frac{0.085}{12} \right]^t - 1}$ and the interest payment as $IP = FaceValue \left[\frac{0.085}{12} \right]$.
4. Compute the scheduled principle, $SP = MP - IP$.
5. Compute the nonscheduled prepayment, $NPP = SMM(FaceValue - SP)$, and the actual payment, $AP = SP + NNP$

References

- [1] N. M. Laird A. P. Dempster and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society*, B 39:1–38, 1977.
- [2] J. H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the american statistical association*, 88(422):669–679, 1993.

- [3] L. Devroye. *Non uniform random variate generation*. Springer Verlag, 1986.
- [4] K. Dunn and J. McConnell. Valuation of gnmA mortgage-backed securities. *The Journal of Finance*, XXXVI(3):599–616, 1981.
- [5] J. E. Ingersoll J. C. Cox and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385 – 407, 1985.
- [6] P. Lecuyer. <http://www.iro.umontreal.ca/~lecuyer>.
- [7] C. P. Robert M. Hurn, A. Justel. Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*, 12(1):1–25, 2003.
- [8] R. Richard and R. Roll. Modeling prepayments on fixed-rate mortgage-backed securities. *Journal of Portfolio Management*, 15(3):73–82, 1989.
- [9] J. Diebolt C. P. Robert. Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society*, B 56(2):363–375, 1994.
- [10] E. Schwartz and W. Torous. Prepayment and the valuation of mortgage-backed securities. *Journal of Finance*, 44(2):375–392, 1989.
- [11] R. Stanton. Rational prepayment and the valuation of mortgage-backed securities. *Review of Financial Studies*, 8(3):677–708, 1996.
- [12] R. Stanton. Unobservable heterogeneity and rational learning: Pool-specific versus generic mortgage-backed security prices. *Journal of Real Estate and Economics*, 12:243–263, 1996.
- [13] Suresh M. Sundaresan. *Fixed income markets and their derivatives*. South-Western College Publishing, 1997.

- [14] S. Richardson W. R. Gilks and D. J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall, 1996.

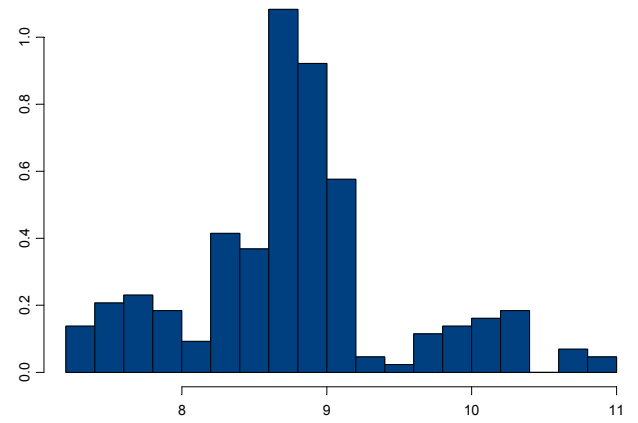
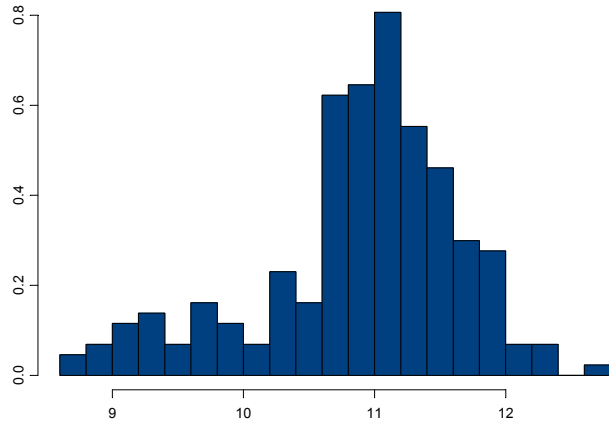
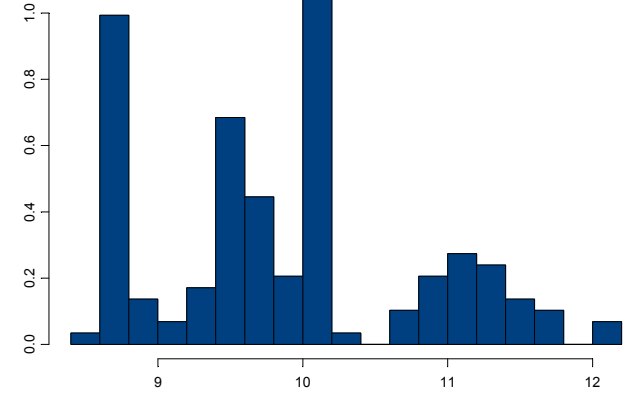
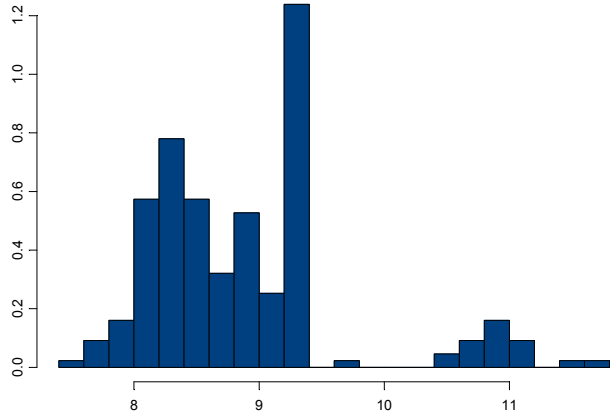


Figure 1: Histograms of the logarithm of the total money paid at the end of the month for four pools of mortgages.

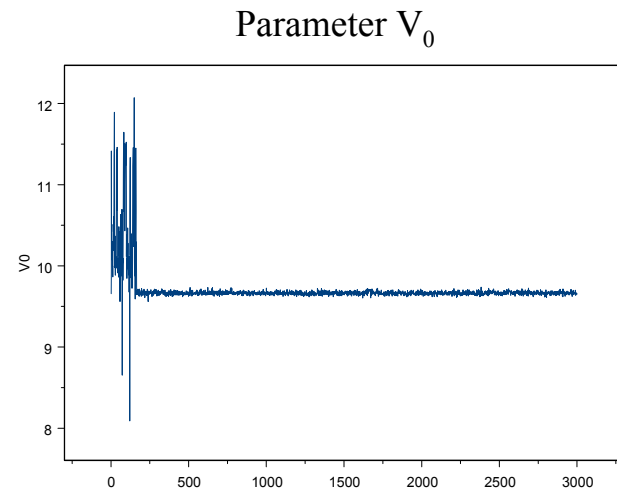
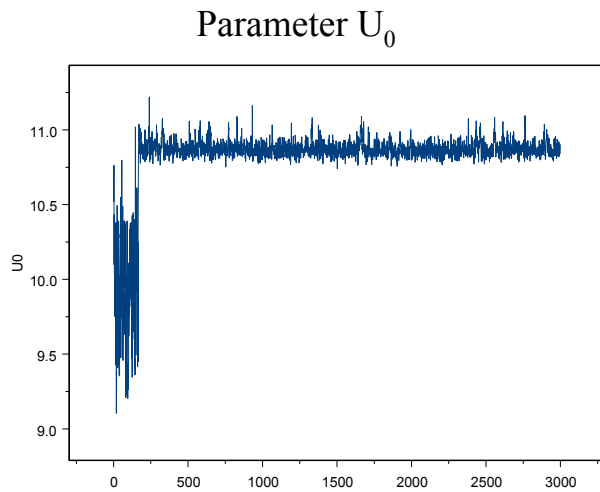
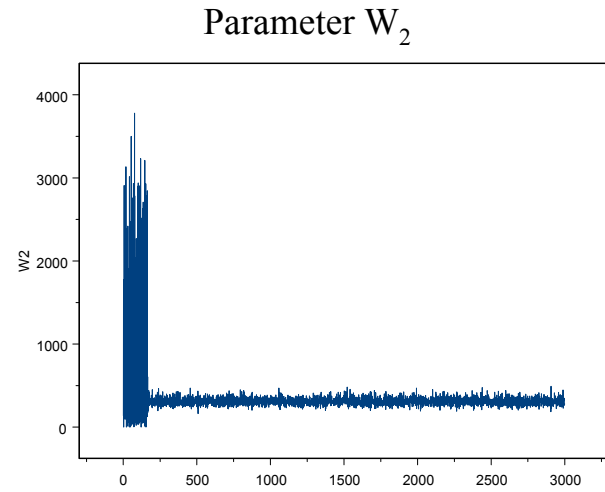
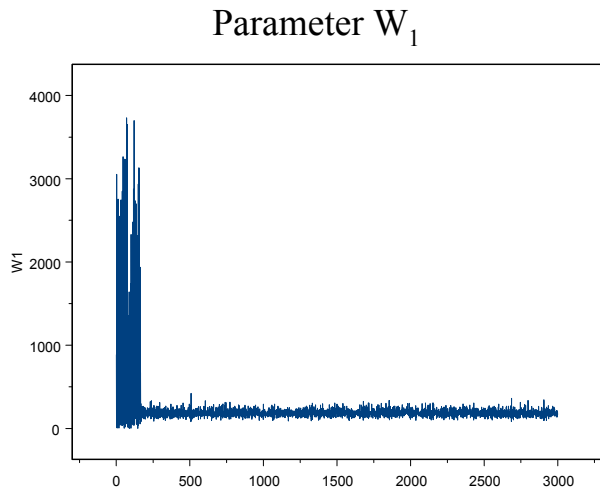


Figure 2: 3000 simulated points for four of the parameters of one of the pools of mortgages.

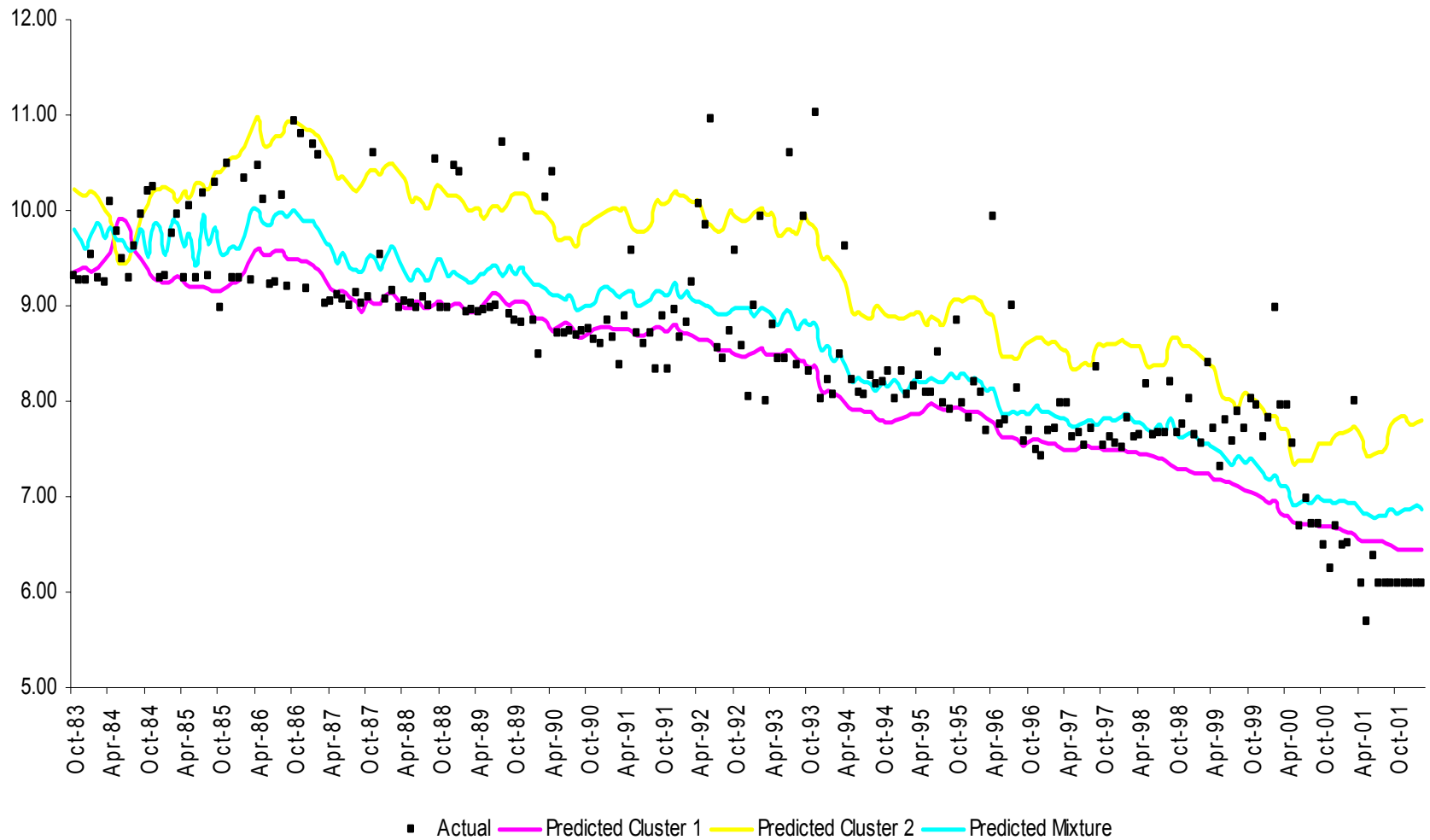


Figure 3: Compares the actual prepayment with the predicted prepayment for one pool of mortgages.

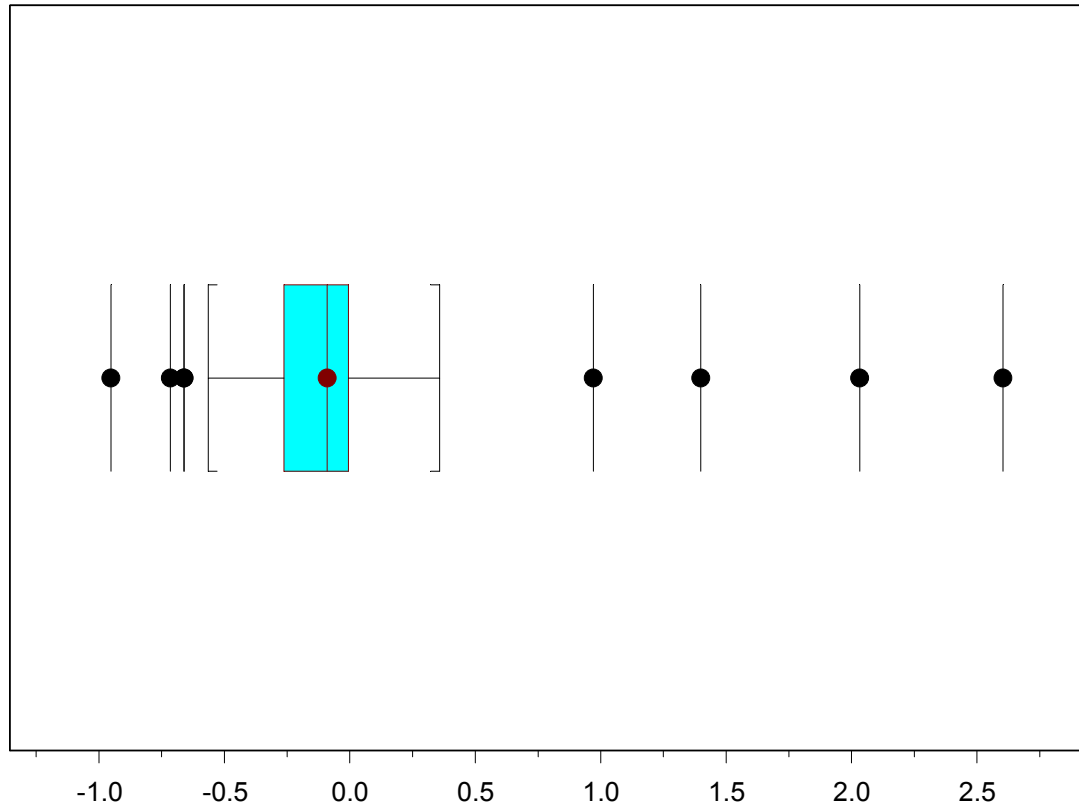


Figure 4: Box plot of the mean of the residuals (measured as the actual minus the predicted) for Bayesian mixture of regressions model.

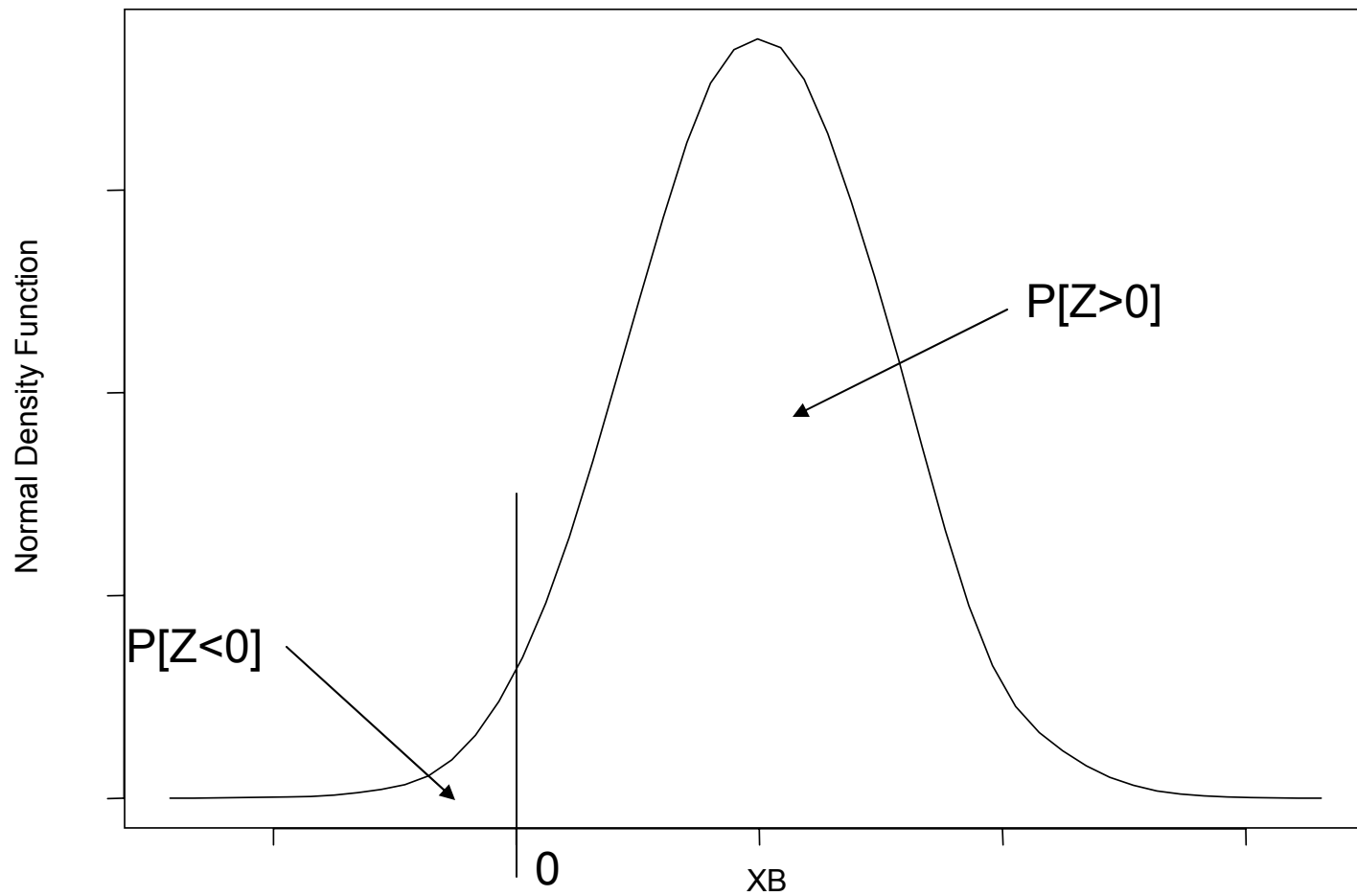


Figure 5

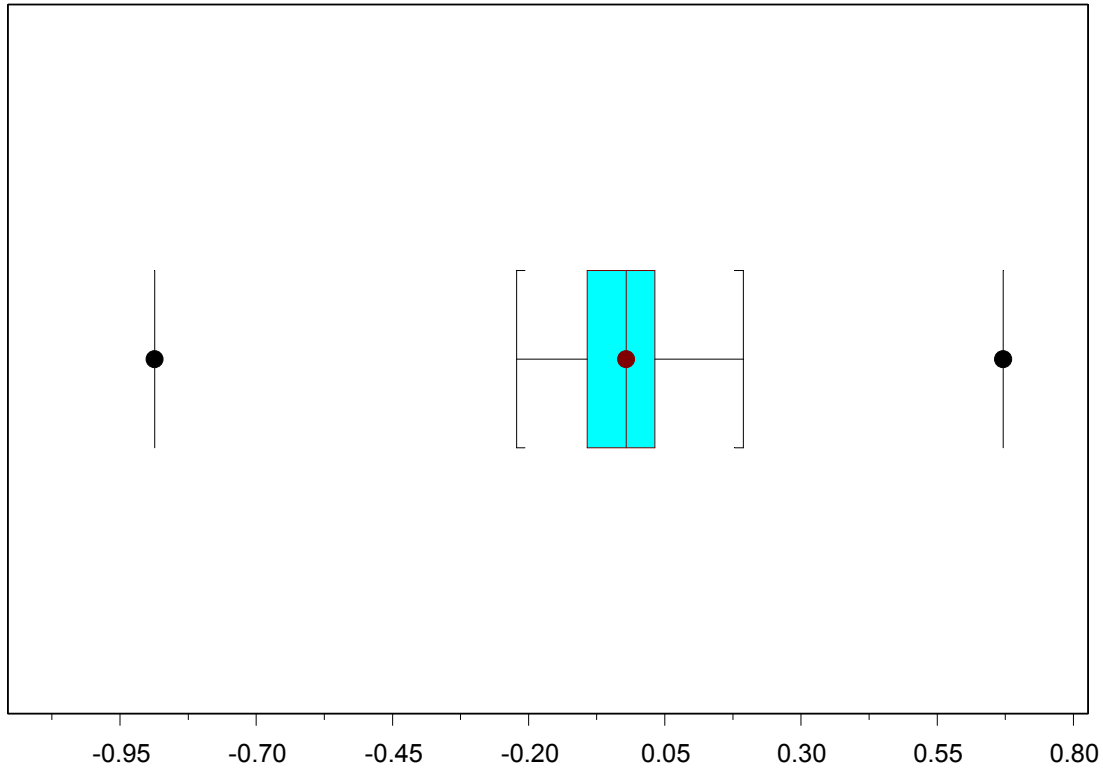


Figure 6: Box plot of the mean of the residuals (measured as the actual minus the predicted) for Probit mixture of regressions model.