2007-11-30

Modeling Transition Probabilities for Loan States Using a Bayesian Hierarchical Model

Rebecca Lee Monson
Brigham Young University - Provo

MODELING TRANSITION PROBABILITIES FOR LOAN STATES USING A BAYESIAN HIERARCHICAL MODEL

by

Rebecca M. Richardson

A project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Statistics

Brigham Young University

December 2007

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a project submitted by

Rebecca M. Richardson

This project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Scott D. Grimshaw, Chair
Date	C. Shane Reese
Date	David G. Whiting

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the project of Rebecca M. Richardson in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Scott D. Grimshaw
Chair, Graduate Committee

Accepted for the Department

Scott D. Grimshaw
Graduate Coordinator

Accepted for the College

Thomas W. Sederberg
Associate Dean, College of Physical and Mathematical Sciences

ABSTRACT

MODELING TRANSITION PROBABILITIES FOR LOAN STATES USING A BAYESIAN HIERARCHICAL MODEL

Rebecca M. Richardson

Department of Statistics

Master of Science

A Markov Chain model can be used to model loan defaults because loans move through delinquency states as the borrower fails to make monthly payments. The transition matrix contains in each location a probability that a borrower in a given state one month moves to the possible delinquency states the next month. In order to use this model, it is necessary to know the transition probabilities, which are unknown quantities. A Bayesian hierarchical model is postulated because there may not be sufficient data for some rare transition probabilities. Using a hierarchical model, similarities between types or families of loans can be taken advantage of to improve estimation, especially for those probabilities with little associated data. The transition probabilities are estimated using MCMC and the Metropolis-Hastings algorithm.

ACKNOWLEDGEMENTS

Thanks to my parents, who encouraged me every step of the way, and my husband who pushed me when I needed it.

CONTENTS

CHAPTER

TABLES

Table

FIGURES

Figure

1. INTRODUCTION

1.1 Markov Chains

A Markov Chain is defined as a sequence of random variables where the probability of being in a given state is dependent only on the state in the previous time period. That is, the future state is completely independent of the past, given a current state. As a more formal definition,

$$P(X_{n+1} = x | X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1, X_0 = x_0) = P(X_{n+1} = x | X_n = x_n)$$

In this paper, consider the discrete case Markov chain, in which there are a fixed number of discrete states, k, to which the random variable may move. The probability of moving from state i to state j at time n is given by

$$p_{ij}^{(n)} = P(X_{n+1} = x_j | X_n = x_i)$$

. Combining these probabilities yields a matrix, called a transition matrix, where each row is a current state for the random variable, and each column represents the next state for the random variable. This matrix appears below:

$$P(n) = \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \cdots & p_{1k}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} & \cdots & p_{2k}^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1}^{(n)} & p_{k2}^{(n)} & \cdots & p_{kk}^{(n)} \end{bmatrix}.$$

Note that each row of this matrix must sum to one. That is, a random variable cannot disappear. If the variable is in a defined state at time n, it must also be in a defined state at time $n+1$. If a state exists such that once a random variable enters that state it remains there, this state is known as an absorbing state. Pre-multiplying a vector containing the current state of a random variable at time n by this matrix yields

a vector containing the probability of being in each state in the next time period. Mathematically, this can be seen as

$$p(n+1) = p(n) \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \dots & p_{1k}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} & \dots & p_{2k}^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1}^{(n)} & p_{k2}^{(n)} & \dots & p_{kk}^{(n)} \end{bmatrix}.$$

In the case where it is only of interest to predict one time period from the future, and the random variable is known to be in state one in the current time period, the prediction for the next time period will be

$$\begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \end{bmatrix}.$$

Expanding this idea, by multiplying the vector of probabilities for the next time period by the transition matrix once again, a vector of probabilities may be created for the state of a random variable in two time periods. If all transition matrices are available, predictions can be made t steps ahead.

1.1.1 Properties of Markov Chains

Certain properties that may be possessed by a Markov chain are desirable for purposes of prediction. Two of these are homogeneity and stationarity.

A population of Markov chains, X_ℓ, may be described as homogeneous if $P_\ell^{(n)} = P^{(n)}$ for all ℓ. That is, all elements in the population have the same transition matrix at time n. This allows the creation of a single transition matrix at each time for any number of objects in the population. If objects in a population have unique transition matrices, it is necessary to create a transition matrix for each member of the population at each time. This makes estimation impossible.

A Markov chain, X_ℓ, can be described as stationary if $P_\ell^{(n)} = P_\ell$ for all n. This condition implies that time does not affect the matrix, but rather the Markov process

can be observed at any time and the same transition matrix will be found. Forecasts are more difficult when a Markov process is nonstationary since $P(n)$, $n = 1, \ldots, T$ must be estimated.

1.2 Business Applications

Markov Chain models are commonly used in business. In a text by Hillier and Lieberman (1995), an example of a Markov chain is given for a department store credit line. In this example, a customer's accounts are classified as fully paid, less than 30 days delinquent, between 30 and 60 days delinquent, or bad debt. A customer can move between these states, and the chance they move from one given state to another is defined by some probability. These values may be used to determine the probability of credit extended becoming bad debt.

Another example of Markov chains is given in a text by Jensen and Bard (2003). In this example, equipment failure is modeled. Two computers have given probabilities of failure. If both fail, all work must be sent out of the office. The office, therefore, has some interest in keeping at least one computer working at all times. The probability of having at least one computer working all day can be modeled using a Markov chain based on the number of working computers at the beginning of the day. This example works well assuming stationarity. The probability of a computer breaking down is the same for any given day.

This text also lists several other situations in which a Markov chain can be used to understand natural phenomena. They can:

(1) predict market shares at specific future points in time,
(2) assess rates of change in market shares over time,
(3) predict market share equilibria (if they exist),
(4) assess the specific effects of marketing strategies when market shares are unfavorable, and
(5) evaluate the process for introducing new products.

This project uses a Markov chain model for mortgage repayment. A certain number of loans are extended, and there is some probability of each loan becoming delinquent or defaulting. More specifically, the loans can be in one of seven states: current, 30 days delinquent, 60 days delinquent, 90 days delinquent, 120 days delinquent, default, or paid off. Those extending the loan are interested in determining whether a loan will become delinquent or default.

To determine whether a loan will enter a delinquent or default state, it is necessary to model the loan's movement from month to month. Considering these movements to be transition probabilities allows for a Markov framework. Notice that this model is very similar to the credit example in a department store listed earlier. Each loan exists in a current state and has some probability of moving to another state in the next time period, dependent on the state they are currently in. Because every loan must be in some state in each time period, the probabilities conditional on the current state must sum to one. A Markov chain model fits this situation, since row probabilities are constrained to sum to one. A loan state such as delinquent or paid off would be considered to be an absorbing state, since loans which move into these states will stay there. A loan that is in one of the delinquent states or current may be current, in one of the delinquent states, default, or be paid off in the next time period. Because default and paid off are absorbing states, there is no need to estimate a transition matrix for these rows. Although these rows are still part of the matrix, they are not of interest in understanding the movement of a loan.

1.3 Transition Matrix Estimation

As observed above, the Markov chain model seems to fit the problem above nicely. It has intrinsically attractive properties for interpretability and problem-solving in a real-world situation. However, one difficulty remains. In order to use the model, it is necessary to know the transition probabilities, which are unknown

quantities.

In order to estimate a transition matrix for a given population, either the assumption of homogeneity or the assumption of stationarity must be made. If the elements within a population are homogeneous, data can be pooled across individual elements of the population. If the transition matrices are stationary, data from different months can be pooled together to estimate a transition matrix for each element of the population. Even with the assumption of homogeneity being satisfied, however, non-stationarity is still a concern. If a population of Markov chains is non-stationary, it is more difficult to justify using an estimated matrix as a prediction for a future transition matrix.

It is possible to construct a transition matrix to be used in estimation without using any data. It can be hypothesized by experts in the field and then used for prediction. This is done in many applications. If someone truly understands the process, he or she may create a matrix representing observations made over years of study in the field. While possible, however, this method of prediction is naive. Any one person can only observe some small proportion of the whole field, and their understanding will not be complete. The understanding held by such a person is valuable, but should not be used for prediction by itself.

A more standard method of estimating the transition probabilities is to simply use proportions from whatever data has been collected. Many advantages exist for estimation in this way. Because actual numbers are being used, the variation in the loans is available and can be used to create intervals of possible values for the probabilities. However, this method takes out any information that might be available outside of the loans. Again, it is unlikely that information about all loans is contained in the data set. A priori information may be available about loan default probabilities, and such information should be included to improve estimation.

A Bayesian solution combines data collected with expert opinion, allowing all

sources of information to be used. The transition probabilities are estimated using a posterior distribution, which is created using the data, but a corrective factor is included based on what experts in the field believe should be observed. In this approach, any information known about the loans is included in a prior distribution. The information from the data is summarized in a likelihood function, and the two are combined to create a posterior distribution. Prior distributions are constructed with the information gathered from experts, allowing the use of information known before data is collected. The likelihood portion allows the inclusion of actual variation within the data. Combining these two yields more specificity in the distribution, but still yields a range of possible values for the probabilities in the transition matrix. With this distribution, forecasts can be made with the knowledge that they are based on a solid statistical foundation.

Consider a Bayesian solution to the problem of estimating the transition probabilities for the loan question. Experts who work in the banking field are consulted to determine an expected transition matrix, and this information is used to create prior distributions. These distributions are combined with the data to create the posterior distribution used for estimation.

Before performing this analysis, the data to be used must be determined. For each month, the state of the loan in the previous month and the state of the loan in the current month is recorded; loans that match are summed and put into tables similar in form to a transition matrix, but using counts instead of probabilities.

One property in this method of estimation that would be desirable is that more data would lead to more accuracy, as in a frequentist model. In a Bayesian model, more data leads to estimates that rely more heavily on the data and less heavily on the prior information. Also, as in a frequentist model, more data leads to tighter distributions around parameters. Since each parameter has a unique distribution, the same data will not be used for each parameter. Only loans actually in a specific state

will be used to estimate the transition probability for that state. Thus, probabilities with more data observed are expected to be more precise — that is, have tighter distributions — than those with few observations.

1.4 Hierarchical Models

As noted previously, one of the assumptions that can be made in order to estimate a transition matrix using data is homogeneity. That is, every loan in the population has the same transition matrix. However, this may not be a practical assumption. It is believed that loans originating at different times may have different transition matrices. There may not be enough data for some of these types of loans, however, to estimate the transition matrix well. In order to improve estimation, perhaps it is reasonable to believe that some transition probabilities differ, but others are the same. A method is then sought to pool the information across loans to take advantage of similarities, but still allow different transition matrices. When information is pooled, the estimates taken from this information will be biased away from the estimates based on non-pooled data. However, if the variance around the estimator decreases enough, or if a quantity that was previously impossible to estimate is now estimable, the tradeoff may be worthwhile. A hierarchical model may be used to pool information and create more accurate, though biased, estimates.

Consider a simple example to demonstrate a Bayesian hierarchical model. The data being analyzed are drawn from a distribution with parameters in turn being drawn from another distribution. Notationally,

$$y_{ij} \sim f(\theta_i)$$

and

$$\theta_i \sim g(\gamma).$$

Finally, to specify the model,

$$\gamma \sim h(a, b)$$

. That is, the j^{th} individual in group i is drawn from the distribution $f(\theta_i)$ and the parameter θ_i is drawn from an overall distribution. In this notation, a and b are numbers developed through research and expert opinion.

For example, all the data may be drawn from normal distributions with the same variance, but each individual i has a unique mean. In this specific case, then, the following distributions would be used:

$$y_i \sim N(\mu_i, \sigma^2)$$

and

$$\mu_i \sim N(\nu, \tau^2).$$

At this point, distributions would be specified for σ^2, ν, and τ^2. Each parameter realization, therefore, is a draw from its distribution, as per the Bayesian framework.

Consider an exam given in class. This exam is designed to test a student's ability and understanding; however, the score is not an exact representation of the student's ability. Illness may drop the score below the student's actual ability, while lucky guessing may yield a score higher than the student's actual ability. Therefore each score is not the student's ability, but an observation from some distribution centered around the student's ability. In order to estimate the student's true ability level, it is necessary to assume that all students' abilities come from some overall distribution. In this case, that constraint makes sense — for students in the same class, some range of ability levels for all the students makes sense. No student should have an ability level less than 40, and no student could have an ability level above 100. These two numbers allow us to construct a distribution, and then each student is a draw from this distribution. Combining the scores with this overall distribution, a distribution can be developed for each student's true ability. Nothing comes without

a penalty, however, and the distribution for each student will be moved towards the overall mean for all students. If another test were to be administered, that score could also be included in the model, and the estimates of true ability would improve.

Consider the following hierarchy for the loan data. Loans originated in a given quarter have a unique transition matrix, but though they are unique, it is believed that within a year, these transition matrices are similar. It is also believed that all transition matrices share certain characteristics. For example, regardless of when a loan is originated, it is not likely to move from 30 days delinquent to 120 days delinquent. However, for a given year, it may be more likely to move from current to 30 days delinquent, and each quarter may have a different, but similar, probability for moving from current to 30 days delinquent. Say loans given in the first year a bank is open were very difficult to qualify for, but as the quarters and years passed, the qualifications became less stringent. Perhaps each year, the stated qualifications were lowered, but in the loan application process, these changes occurred each quarter. Then it might be expected that the probability of becoming 30 days delinquent from current would increase every quarter, but this probability would be more similar for quarters within a year than those from a different year. By combining this information, it is possible to develop distributions for each probability for each transition matrix.

1.4.1 Theoretical Advantages of Hierarchy

Gelman (2006) discusses some of the advantages of hierarchical models, which he terms multilevel models. One advantage he cites is the ability of a multilevel model to take into account the variation in estimating other parameters and correcting for it. Because the estimation has some inherent level of variation, using the estimated values as truth in further analysis is misleading and gives an underestimate of the variance in the parameter estimates. Using a hierarchical model, all variance is openly

acknowledged and adjusted for by "shrinking toward the complete pooling estimate."

Because of the pooling of the data, there is more information available for each estimate and, as expected, our estimates are more accurate. These distributions will not be centered around the proportions calculated directly from the data for each matrix, but will be between those proportions, the proportions suggested by our prior information, and the overall proportion for that portion of the transition matrix. The distribution will be narrower.

1.5 Convergence Assessment

To assess convergence of the MCMC algorithm, mixing plots should be examined. These plots, which consist of the values of one specific parameter plotted against their iteration number, give some idea of whether the plot has reached convergence. Several different features should be examined.

First, consider the overall appearance of the graph. If the overall trend of the line plot appears to be either increasing or decreasing, convergence has probably not been reached. If the beginning of the graph trends either up or down but then angles to become a horizontal pattern, it is possible that convergence has been reached. If the starting values are correct there will be no trend in the graph at all because no iterations were needed for the parameters to be in the proper range.

If a curvilinear pattern is evident in the mixing plot then the jump distance, or candidate sigma, could be too small. If the plot is not allowed to make large jumps relative to the variance in the parameter, it is unlikely that the realizations will span the parameter space. Another sign that the candidate sigma is too small is found in the staying percentage. The Metropolis-Hastings algorithm is designed such that some significant percent of the time, the proposed value will not be accepted. If all or most proposed values are accepted, the candidate sigma may be too small for the realizations to span the parameter space. The candidate sigma could also be too

large, which can be seen if there are large "plateaus" appearing in the graph. If such features are found, the algorithm should be rerun with a smaller candidate sigma.

If the plot does not appear to have converged, the second run should be done with better starting values. This will not change the final distribution, but it may increase the speed of convergence. Beyond these diagnostic, several further methods have been proposed for monitoring convergence. Three such methods will be discussed here.

One method of monitoring convergence was proposed by Gelman and Rubin (1992). The method they propose has two steps: first, an estimate of the target distribution is created, and an overdispersed distribution is created; second, values are generated from this overdispersed distribution as starting values for multiple sequences which are then generated and used to assess the performance of the simulation.

Finding an overdispersed distribution is done in three proposed steps. First, the modes of the target distribution should be found. From these, high-density regions are found, and the full distribution is modeled as a multivariate normal with modes matching the target distribution. Second, values are drawn from the normal mixture and then divided by some scalar random variable. This insures overdispersion. Finally, the approximation is sharpened by downweighting the sections of the distribution that have low density with the target distribution. While it is true that in most cases the target distribution is unknown, the shape of the distribution generally is known up to a constant. This makes the method feasible.

Once an overdispersed distribution is determined, m values are drawn from this distribution and a random sequence is started from each value with length $2n$. The first n iterations are considered burn, and only the last n are used. With these sequences, the between-sequence variance is calculated, as well as the average within-sequence variance. Also calculated are the sample mean, which is the mean over

all sequences, and the target variance, which is a weighed average of the between-sequence and within-sequence variance. From these values, the improvement possible by running the chain longer may be determined with some degree of confidence, and it can be observed whether the chains are converging to the overall mean.

The proposed method is beneficial in that it seeks to use statistical inference to determine whether the chain has converged. However, in practice, finding an overdispersed distribution from which to draw starting values is very difficult, and ultimately, it may not be possible to determine whether such a distribution has been found when dimensionality is high.

Raftery and Lewis (1992) propose a different method for analyzing convergence. Calling the method proposed by Gelman and Rubin inefficient, they propose using a single chain with restarts if the chain does not appear to be converging. If the advantage of multiple starts is simply to avoid poor starting values, Raftery and Lewis suggest that simply choosing new starting values through trial and error is as effective as using the complicated mode-finding method of Gelman and Rubin. Since simply choosing new starting values and restarting the iterations appears to be as effective as running multiple chains in achieving convergence, this simpler method would be preferred.

Polson (1996) proposed a theoretical solution to solve for the upper bound on the number of iterations needed to reach convergence. Given some arbitrary $\epsilon > 0$, Polson posits that it is possible to find an upper bound such that after that number of iterations the Markov chain will have reached stationarity. However, while this is theoretically possible, the math required to find this upper bound is fairly difficult and unique to each problem. Because of this, the implementer must solve for the upper bound of the rate of convergence and show that this constant is polynomially bounded. Thus, although Polson has provided a theoretical method to attain convergence, it is not necessarily feasible to implement.

2. A SIMPLE EXAMPLE

In this simple case, consider the situation where a loan can be in one of only two states — current or default. A loan in each of these states will have some probability of staying in that state, and one minus that probability of moving to the opposite state. Also consider that each loan comes from a "family" of loans based on some characteristic of the loan itself. For example, the loan origination year may be the defining characteristic of a family of loans. In a hierarchical framework, each of these families is also considered to be related to the others via some distribution.

2.1 Sample Data Set

Consider the following example, where a bank offers loans to three types of borrowers. For each borrower type, the monthly transitions between two states, namely current and delinquent, are tallied. These transitions are recorded in the following matrices:

$$\text{Borrower Type I: } \begin{bmatrix} 940 & 10 \\ 1 & 49 \end{bmatrix}$$

$$\text{Borrower Type II: } \begin{bmatrix} 250 & 50 \\ 190 & 10 \end{bmatrix}$$

$$\text{Borrower Type III: } \begin{bmatrix} 200 & 10 \\ 10 & 30 \end{bmatrix}.$$

The first row represents the loans beginning in state one, or current. The second row represents the loans beginning in state two, or default. The first column represents loans that will be current in the next time period, and the second column represents loans that will default in the next time period. For example, in the first family of

loans, 940 loans that are current this time period will be current next period, and 10 loans that were current this time period will be delinquent next period.

It appears that borrower type one is most likely to stay current of the three types. However, if a borrower does default, he or she is likely to stay in that category. This borrower type also has far more data than the other borrower types. The second type of borrower is in default far more regularly, but is also very likely to move back to current in the next time period. Less data are observed for this borrower type. The third borrower type has the least data and shows a similar pattern to type one, with the exception that the probabilities are closer to each other than for type one. Because of the amount of data in each matrix, it is expected that the distribution around estimates of probabilities for borrower type one will be narrower than the distribution around the other probabilities.

Notice that these matrices are in terms of frequencies. The model used will generate probabilities, so instead of the matrices listed above, the following matrices will be generated:

$$\text{Borrower Type I: } \begin{bmatrix} .989 & .011 \\ .020 & .980 \end{bmatrix}$$

$$\text{Borrower Type II: } \begin{bmatrix} .833 & .167 \\ .950 & .050 \end{bmatrix}$$

$$\text{Borrower Type III: } \begin{bmatrix} .952 & .048 \\ .250 & .750 \end{bmatrix}.$$

It must be noted that when displaying the information in this form, the amount of data in each matrix is lost.

2.2 Distributional Choices for the Model

The first and second rows of the transition matrix will be modeled separately since rows are mutually exclusive states the previous month. A loan may be in only one state in a given time period. Consider the first row of each matrix. The probabilities in this row must sum to one, and since a current loan can be considered a "success" and a default loan can be considered a "failure," a binomial model is a natural modeling choice. Consider each y_{iij}, the number of loans in the i^{th} state this month that remain in the ith state next month for borrower type j, to have a binomial distribution, $y_{iij} \sim \text{Bin}(n_i, \pi_i)$, where n_i is the number of loans in the i^{th} state this month and π_{ij} is the probability of staying in the ith state next month for borrower type j. Each $\pi_{ij} \sim \text{Beta}(\alpha_j, \beta_j)$, taking advantage of conjugacy, and $\alpha_j \sim \text{Gam}(a_{1j}, b_{1j})$, $\beta_j \sim \text{Gam}(a_{2j}, b_{2j})$. Each i has a unique hierarchical model. The beta distribution is an obvious choice when modeling probabilities because the range of values of the random variable is [0,1] and the conjugacy results in a closed form solution for the conditional distribution of π_i. The gamma distribution seems to be a good choice for α and β for a few reasons. Most importantly, this distribution maintains the parameter space. The parameters α and β must be greater than zero, and a gamma random variable has range greater than zero. The gamma allows a great deal of flexibility in the shape of the distribution itself. It is possible to specify parameters such that both the mean and the variance can be adjusted, whereas in other distributions where parameter space is maintained, only one parameter can be selected, and both the mean and the variance are determined from this one parameter.

2.3 Gibbs Sampling and Metropolis-Hastings

The posterior distribution of a hierarchical model is given as

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\underline{y}|\pi_i)\pi(\pi_i|\alpha,\beta)\pi(\alpha)\pi(\beta)}{\int f(\underline{y}|\pi_i)\pi(\pi_i|\alpha,\beta)\pi(\alpha)\pi(\beta)d\theta},$$

where $\theta = (\underline{\pi}, \alpha, \beta)$. From the distributional choices given above and noting that the divisor is simply a constant, the following proportional posterior distribution is developed:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \prod_{i=1}^{3} \pi_i^{\sum_j y_{ij}} (1-\pi_i)^{n_i - \sum_j y_{ij}} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi_i^{a-1} (1-\pi_i)^{b-1} a^{\alpha_1 - 1} e^{-a/\beta_1} b^{\alpha_2 - 1} e^{-b/\beta_2}$$

The proportional distribution is used in place of the actual distribution because the divisor, called the constant of integration, is not available in closed form. Because this distribution cannot be written out exactly, the Metropolis-Hastings algorithm will be used to generate the posterior distribution.

In the Metropolis-Hastings Algorithm, the difficulty to be overcome is attempting to generate from a distribution with no closed-form answer. To solve this problem, a random walk algorithm will be used to simulate draws from the distribution. This algorithm proceeds as follows:

(1) Set starting values for each parameter in the distribution. Call these $\boldsymbol{\theta^0}$. Note that these values must be possible values for the distribution.

(2) Choose a candidate density function. This should be a distribution that is easily sampled from. It must also maintain the space of the possible values of the parameter. Using the full Metropolis-Hastings algorithm, it is not necessary to choose a symmetric distribution. Call this $D^*(\theta)$.

(3) Solve for the complete conditionals — those portions of the unnormalized posterior involving the parameter of interest. Call this $c(\theta)$.

(4) For t=1, 2, ..., n,

   (a) generate a value from your candidate density function, θ^*;

   (b) calculate the ratio of densities:

$$\frac{c(\theta^*|y)/D^*(\theta^*|\theta^{t-1})}{c(\theta^{t-1}|y)/D^*(\theta^{t-1}|\theta^*)};$$

(c) generate a Uniform(0,1) random variable, and compare it with the ratio calculated in the previous step. If the ratio is greater than the uniform variable, the move is "accepted," and θ^t is assigned the value θ^*. If the ratio is lower than the uniform random variable, θ^t is assigned the value θ^{t-1}.

(5) Continue this process for some number of iterations and eventually these generated values will be draws from the posterior distribution.

Through this algorithm, likely values of θ will have high ratios and will rarely be less than any draw from a uniform distribution. Unlikely values of θ will have low ratios and will usually be lower than any draw from a uniform distribution. Because of the comparison to a uniform random variable, however, some likely values of θ will be rejected and some unlikely values of θ will be accepted.

Gibbs sampling is a simpler algorithm that can be used when generation directly from the complete conditionals is possible. That is, if the complete conditional for some parameter forms the kernel of some known distribution, then it is possible to sample directly from that distribution given the current values of the other parameters. Each parameter is sampled from each iteration, so the current values of the other parameters will change each step. In this way, a value is generated for each parameter at each iteration. After some number of iterations, the values generated will be realizations from the posterior distribution, just as in the Metropolis-Hastings Algorithm.

2.4 Complete Conditionals

Because the beta is conjugate with the binomial, it can be seen that the complete condition for the π_i's forms the kernel of a Beta $(\sum_j y_{ij} + a, n_i - \sum_j y_{ij} + b)$, and a Gibbs sampler can be used to generate these values from that beta distribution. The

complete conditionals for α and β do not have closed form, but

$$[\alpha] \propto \frac{\Gamma(\alpha+\beta)^3}{\Gamma(\alpha)^3\Gamma(\beta)^3}(\pi_1\pi_2\pi_3)^{\alpha-1}\alpha^{a_1-1}e^{-\alpha/b_1}$$

and

$$[\beta] \propto \frac{\Gamma(\alpha+\beta)^3}{\Gamma(\alpha)^3\Gamma(\beta)^3}((1-\pi_1)(1-\pi_2)(1-\pi_3))^{\beta-1}\beta^{a_2-1}e^{-\beta/b_2}.$$

These functions are the $c(\theta)$ functions described in the algorithm above. Because these functions do not form the kernel of any known distribution, the Metropolis-Hastings algorithm must be used to generate from the posterior distribution of both α and β. Consequently, the posterior distribution will be generated using both methods.

2.4.1 Candidate density function

Because α and β are constrained to be greater than 1, the Truncated Normal distribution will be used as the candidate density function. This distribution cuts the normal distribution off at chosen cut points and rescales the portion that remains to be a valid distribution. In this case, the distribution will be such that only values greater than one are possible. The shape of this distribution will appear identical to a normal distribution, simply higher by some proportion. Figure 2.1 shows a normal distribution centered at 3 with a standard deviation of 1.5, compared to the same distribution truncated at 1. The shape of the distribution appears identical except for a scaling factor.

Using this distribution as a candidate density function in this problem has the advantage of only allowing possible values for α and β to be generated, as it is truncated at 1. Also, generating from this distribution is easy. Because it is only different from the normal distribution by a scaling factor, a simple modification of the already written function in R for generating normal random variables will allow generation from this distribution. Unfortunately, the truncated normal distribution is not a symmetric distribution, so the full Metropolis-Hastings algorithm must be

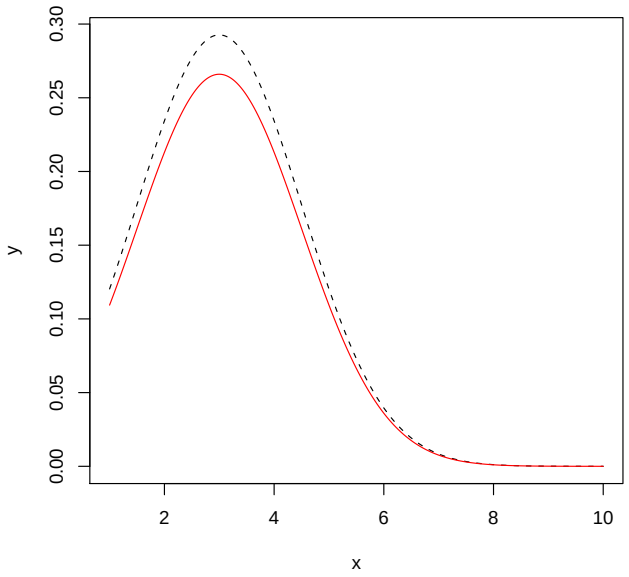

Figure 2.1: Normal distribution compared to a truncated normal distribution

used.

2.4.2 Prior Elicitation

In eliciting a prior for the beta distribution in a simple case, it may be easier to consider an alternate parametrization of the beta distribution. Note that $\frac{a}{a+b}$ is the expected proportion of success, and $a+b$ is the effective sample size representing the expert's level of knowledge. In this example, it has been determined from expert opinion that the probability of moving from current to current is .95 with 100 samples worth of information. That is, the expert consulted felt the estimate of .95 carried with it the same weight as 100 samples would. This leads to a mean for the gamma hyperprior on a of 95. However, this information is only dealing with the overall loan mean for this probability. The purpose for using the hierarchical model is to allow flexibility in the distributions. Using a variance of 25, which yields a standard deviation of 5, allows for more variability in the amount of prior information available for any specific type of loan. The distribution is fairly symmetric, with a mean, median, and mode all around .95. The gamma distribution on b is centered around 5, which will place the overall mean of the probability for the current to current move at .95. The standard deviation of about 1.5 again gives a fairly diffuse distribution, allowing for a great deal of variation in the information and mean for each loan type. To assess whether the prior information is truly representative of a priori information, consider the prior predictive distribution displayed in Figure 2.2.

This graph is a distribution of realizations for the probability of the current to current move. Notice that the mode is .955; the interval (.888, .988) contains approximately 95% of the values. This distribution might be slightly more disperse than is probable, but this will allow the data to have greater influence on the posterior distribution.

2.5 Assessing Metropolis-Hastings

After running the iterative sampler, mixing plots were examined. These plots appear in Figure 2.3. From these mixing plots, which are looking only at a subsection of 1000 draws, it appears that convergence has been reached. There is no systematic pattern either up or down, and the space appears to be spanned. To check a final worrisome possibility, a scatterplot of π_2 and α is examined. If the two parameters are correlated, it is possible that not all of the parameter space will be spanned. However, examining Figure 2.4, this seems to be a needless concern because there is no correlation visible. π_2 was chosen because there is little data associated with it, making it the most vulnerable to correlation.

Figure 2.6 contains the posterior distributions for α and β. Looking at the posterior distributions for the probabilities of staying current for each loan type, which are displayed in Figure 2.5, the patterns observed in the data are seen in the posterior distributions. The first probability, π_{11}, which is associated with the borrower type with the most data and the highest probability of staying current, is displayed in black and has the highest mean and the narrowest distribution. The second borrower type, which had a lower probability as well as less data, is displayed in red and shows a lower mean with a wider distribution. The third borrower type, which is displayed in blue, has a slightly lower mean than the first, but the distribution is much wider.

When comparing the means calculated directly from the data with the observations drawn from the posterior distribution, the "borrowing of strength" between borrower types is apparent. While the posterior realizations are similar to the data, they show a moderated relationship. All values have shrunk somewhat to the overall mean, as can be seen in Table 2.1.

Borrower Type	Estimate from Data	Posterior Mean
I	.98947	.9887
II	.83333	.868
III	.95238	.961

Table 2.1: Table comparing data predictions to posterior predictions.

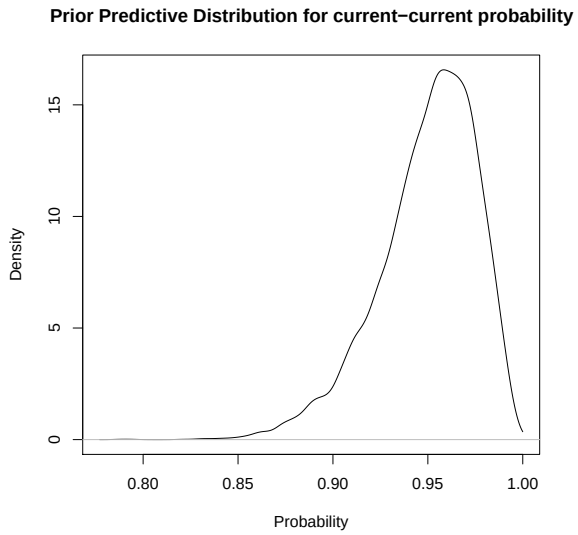

Figure 2.2: Prior Predictive Distribution for the current to current move. Notice that it is centered around .95, but has significant density as low as .85.

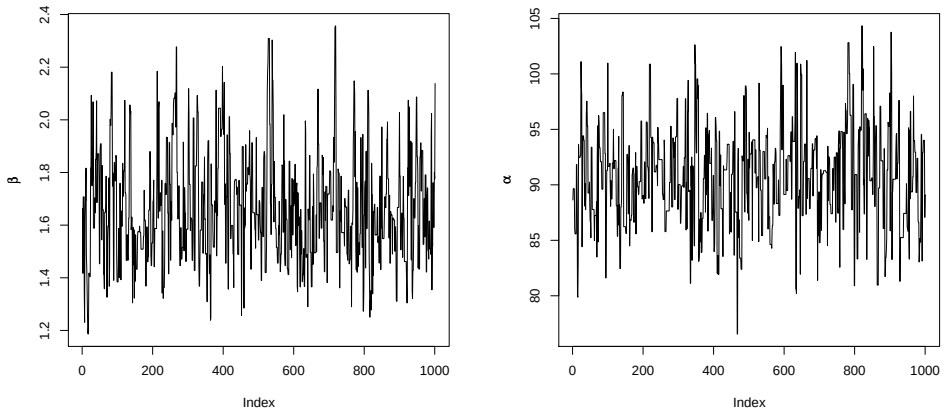

Figure 2.3: Mixing plots for α and β

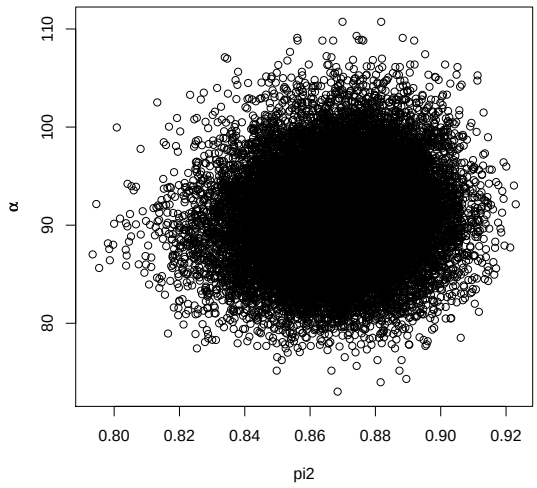

Figure 2.4: Scatterplot of α and π_2. Notice that the alpha parameter is not correlated with the probabilities at all.

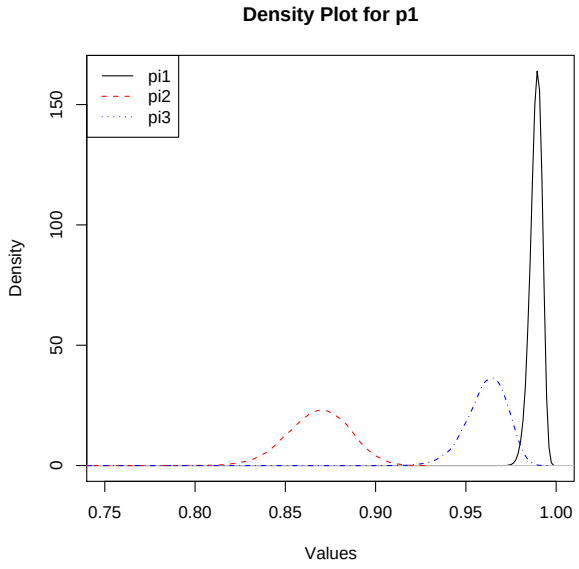

Figure 2.5: Posterior distributions for the probability of staying current for each loan type

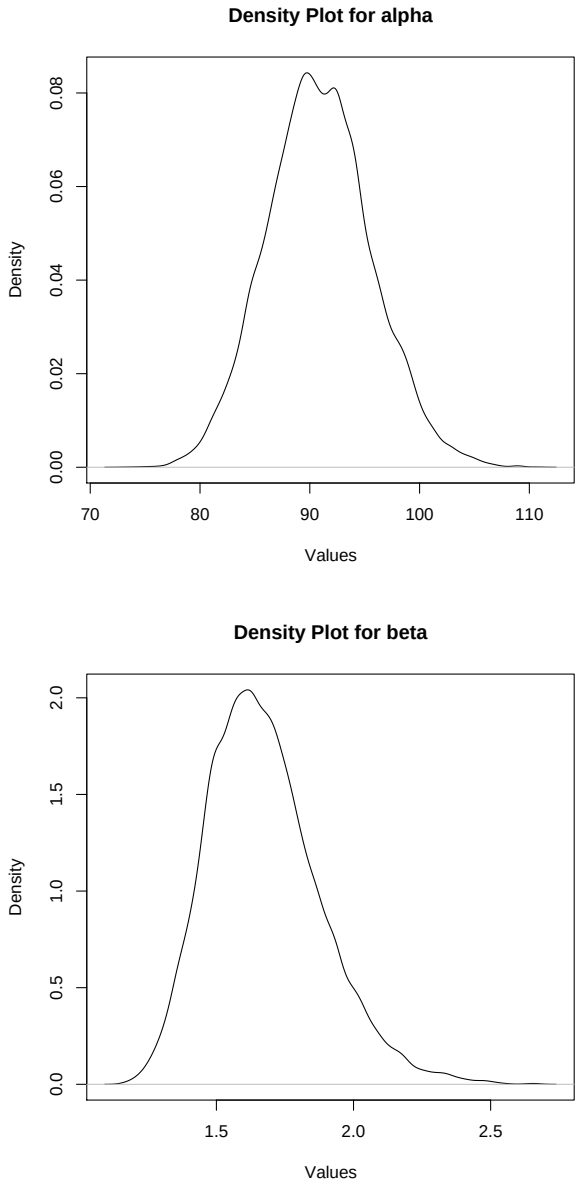

Figure 2.6: Posterior distributions for α and β

3. A SIMULATION STUDY

3.1 Expansion of Question

Because of sparse data in rare transitions, it is of interest to determine whether a hierarchical model may be used in estimating a more complicated transition matrix. This model allows for strength to be borrowed across quarters so that those transition probabilities that correspond to rare transitions can be estimated.

Computationally, this is a far more difficult problem than the simple case considered in Chapter 2. The Metropolis-Hastings Algorithm is still being used, but the number of parameters whose distributions need to be simulated has increased dramatically. As this occurs, the time the simulation needs to run increases exponentially. It also becomes more difficult to assess convergence because more parameters must converge for convergence to be reached.

When expanding from the trivial example discussed previously, the multinomial is an appropriate distribution for the data, as it is an extension of the binomial distribution, and a Dirichlet distribution is appropriate for the parameters of the multinomial, as it is an extension of the beta. The gamma distribution is still appropriate for the hyperparameters of the Dirichlet. These are standard distributions when dealing with multiple states. A multinomial model has the same interpretation for multiple states that a binomial model has for two states. Similarly, a Dirichlet prior is the conjugate prior for the multinomial and is a standard extension for the beta distribution, which was used as the prior for a binomial. The hyperprior distributions themselves do not have to change. Gamma hyperpriors are still appropriate for the same reasons as listed previously. However, in the more complicated example, there will be more hyperprior distributions because there are more parameters in the Dirichlet than in the beta distribution.

In order to determine the effects of minimal data on the Bayesian hierarchical model for a Markov chain, a simulation study will be performed. In this study, the the true hyperprior distributions will be specified, and parameters will be simulated from these distributions. The data will then be generated from the simulated parameters. Because the parameters and the distributions of those parameters are known, it will be possible to compare our estimates to the generated values and determine whether the procedure is valid. Consider the following data. A number of loans are observed having started in four different quarters. There are four possible states for each loan, including: current, delinquent, default, or prepaid. The data can be organized as follows:

$$\begin{bmatrix} y_{111} & y_{112} & y_{113} & y_{114} \\ y_{121} & y_{122} & y_{123} & y_{124} \\ y_{131} & y_{132} & y_{133} & y_{134} \\ y_{141} & y_{142} & y_{143} & y_{144} \end{bmatrix}$$

$$\begin{bmatrix} y_{211} & y_{212} & y_{213} & y_{214} \\ y_{221} & y_{222} & y_{223} & y_{224} \\ y_{231} & y_{232} & y_{233} & y_{234} \\ y_{241} & y_{242} & y_{243} & y_{244} \end{bmatrix}$$

$$\begin{bmatrix} y_{311} & y_{312} & y_{313} & y_{314} \\ y_{321} & y_{322} & y_{323} & y_{324} \\ y_{331} & y_{332} & y_{333} & y_{334} \\ y_{341} & y_{342} & y_{343} & y_{344} \end{bmatrix}$$

$$\begin{bmatrix} y_{411} & y_{412} & y_{413} & y_{414} \\ y_{421} & y_{422} & y_{423} & y_{424} \\ y_{431} & y_{432} & y_{433} & y_{434} \\ y_{441} & y_{442} & y_{443} & y_{444} \end{bmatrix}.$$

Then for y_{ijk}, i represents the quarter in which the loan was originated, j represents the state of the loan in the previous month, and k represents the state of the loan in the current month.

3.2 Hyperprior Selection

Because of the nature of the data being analyzed, special consideration was taken in formulating the hyperprior distributions so generated values would follow observed loan repayment trends. Most people who take out loans make payments on time, so the highest count of loans is expected in this square of the table. Some number, not high, but significant, will have been current in the previous month and delinquent in this month. It should not be possible to move from current to default in one month, but under extreme circumstances this may happen. The probability for this event is very low, so most likely, there will not be any loans here. If the prior gives no probability for that occurrence, however, the posterior will have no probability there, even if data appears in that location. Some number of loans that were current last month are expected to prepay, so a few loans will appear in this location in the transition matrix each month.

For those loans that were delinquent, it is expected that over half will move from delinquent to current in the next month. Many will also stay delinquent, so a large number of loans will appear there as well. Several will move from delinquency to default, and a smaller number will prepay. As noted previously, more loans will start current than will start delinquent. Based on these conditions, the hyperprior distributions were chosen. The hyperparameters are given in Table 3.1.

3.3 Distributional Model for Data

As discussed previously, the data for loan defaults follow a multinomial distribution. Using the same subscripts as above, consider the data in the vector y_j. Notice

Previous to Current Month states	a Parameter	b Parameter
Current to Current	500	5/9
Current to Delinquent	900	15
Current to Default	7	7
Current to Prepayment	80	8
Delinquent to Current	960	4.8
Delinquent to Delinquent	520	5.2
Delinquent to Default	180	3.6
Delinquent to Prepaid	205	8.1

Table 3.1: Parameters for Gamma hyperprior distributions

that this matrix is a combination of the first rows of all four matrices mentioned previously.

$$y_j \begin{bmatrix} y_{1j1} & y_{1j2} & y_{1j3} & y_{1j4} & \cdots & y_{4j1} & y_{4j2} & y_{4j3} & y_{4j4} \end{bmatrix}.$$

That is, for one row of the matrix above, which takes data only from the top row of the four original transition matrices,

$$f(\mathbf{y_{i1}}) = \frac{n_i!}{y_{i11}!y_{i12}!y_{i13}!y_{i14}!} y_{i11}^{p_{i11}} y_{i12}^{p_{i12}} y_{i13}^{p_{i13}} y_{i14}^{p_{i14}},$$

where i represents the quarter in which the data are observed. Then, over all four quarters,

$$f(\mathbf{y_1}) = \frac{n_1!n_2!n_3!n_4!}{y_{111}!y_{112}!y_{113}!y_{114}!y_{211}!y_{212}!y_{213}!y_{214}!y_{311}!y_{312}!y_{313}!y_{314}!y_{411}!y_{412}!y_{413}!y_{414}!} \times$$

$$y_{111}^{p_{111}} y_{112}^{p_{112}} y_{113}^{p_{113}} y_{114}^{p_{114}} y_{211}^{p_{211}} y_{212}^{p_{212}} y_{213}^{p_{213}} y_{214}^{p_{214}} y_{311}^{p_{311}} y_{312}^{p_{312}} y_{313}^{p_{313}} y_{314}^{p_{314}} y_{411}^{p_{411}} y_{412}^{p_{412}} y_{413}^{p_{413}} y_{414}^{p_{414}}.$$

Dropping out constants, this becomes

$$f(\mathbf{y_1}) \propto y_{111}^{p_{111}} y_{112}^{p_{112}} y_{113}^{p_{113}} y_{114}^{p_{114}} y_{211}^{p_{211}} y_{212}^{p_{212}} y_{213}^{p_{213}} y_{214}^{p_{214}} y_{311}^{p_{311}} y_{312}^{p_{312}} y_{313}^{p_{313}} y_{314}^{p_{314}} y_{411}^{p_{411}} y_{412}^{p_{412}} y_{413}^{p_{413}} y_{414}^{p_{414}}$$

Because the second subscript refers to the row of the original matrices, and it is constant across the constructed matrix, it will be dropped in future references. That is, y_{114} will be referred to as y_{14}.

3.4 Distributional Model for Probabilities

Because the data are distributed according to a multinomial distribution, and this distribution is a member of the regular exponential class, a conjugate prior exists. This prior distribution, a Dirichlet, is used. The probabilities associated with a single quarter for a given row of the transition matrix are distributed according to a Dirichlet distribution with parameters α_{j1}, α_{j2}, α_{j3}, and α_{j4}, where j represents the row of the transition matrix. Mathematically,

$$(p_{ij1}, p_{ij2}, p_{ij3}, p_{ij4}) \sim Dir(\alpha_{j1}, \alpha_{j2}, \alpha_{j3}, \alpha_{j4}).$$

This is true for all i, or all quarters. Considering only the first row of the transition matrix and dropping that subscript,

$$f(\underline{p_i}) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)\Gamma(\alpha_4)} p_{i1}^{\alpha_1 - 1} p_{i2}^{\alpha_2 - 1} p_{i3}^{\alpha_3 - 1} p_{i4}^{\alpha_4 - 1}.$$

3.5 Distributional Model for Prior Parameters

In order to maintain the positive nature of the α parameters in the Dirichlet distribution, a gamma distribution seems appropriate. This model allows more flexibility because the mean and variance are not dependent upon each other as they are in other continuous positive distributions, such as the χ^2 distribution or the exponential distribution, which also maintain the parameter space. For a given row of the transition matrix,

$$f(\alpha_i) = \frac{1}{\Gamma(a_1) b_1^{a_1}} \alpha_1^{a_1 - 1} e^{-\alpha_1 / b_1}.$$

3.6 Proportional Posterior Distribution

Combining all the previous distributions together, a function proportional to the posterior distribution is constructed.

$$f(\theta) \propto y_{111}^{p_{111}} y_{112}^{p_{112}} y_{113}^{p_{113}} y_{114}^{p_{114}} y_{211}^{p_{211}} y_{212}^{p_{212}} y_{213}^{p_{213}} y_{214}^{p_{214}} y_{311}^{p_{311}} y_{312}^{p_{312}} y_{313}^{p_{313}} y_{314}^{p_{314}} y_{411}^{p_{411}} y_{412}^{p_{412}} y_{413}^{p_{413}} y_{414}^{p_{414}} \times$$

$$\frac{\Gamma(\alpha_1+\alpha_2+\alpha_3+\alpha_4)^4}{(\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)\Gamma(\alpha_4))^4}(p_{11}p_{21}p_{31}p_{41})^{\alpha_1-1}(p_{12}p_{22}p_{32}p_{42})^{\alpha_2-1}(p_{13}p_{23}p_{33}p_{43})^{\alpha_3-1}\times$$

$$(p_{14}p_{24}p_{34}p_{44})^{\alpha_4-1}\alpha_1^{a_1-1}e^{-\alpha_1/b_1}\alpha_2^{a_2-1}e^{-\alpha_2/b_2}\alpha_3^{a_3-1}e^{-\alpha_3/b_3}\alpha_4^{a_4-1}e^{-\alpha_4/b_4}.$$

From these, complete conditionals for each of the parameters are constructed by dropping all constants with respect to each parameter. Then the complete conditional for α_1 is

$$[\alpha_1] \propto \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3+\alpha_4)^4}{\Gamma(\alpha_1)^4}(p_{11}p_{21}p_{31}p_{41})^{\alpha_1-1}\alpha_1^{a_1-1}e^{-\alpha_1/b_1}.$$

Similarly,

$$[\alpha_2] \propto \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3+\alpha_4)^4}{\Gamma(\alpha_2)^4}(p_{12}p_{22}p_{32}p_{42})^{\alpha_2-1}\alpha_2^{a_2-1}e^{-\alpha_2/b_2},$$

$$[\alpha_3] \propto \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3+\alpha_4)^4}{\Gamma(\alpha_3)^4}(p_{13}p_{23}p_{33}p_{43})^{\alpha_3-1}\alpha_3^{a_3-1}e^{-\alpha_3/b_3}, \text{ and}$$

$$[\alpha_4] \propto \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3+\alpha_4)^4}{\Gamma(\alpha_4)^4}(p_{14}p_{24}p_{34}p_{44})^{\alpha_4-1}\alpha_4^{a_4-1}e^{-\alpha_4/b_4}.$$

Remembering that the probabilities are generated jointly, complete conditionals are also created for each π_i.

$$[\pi_1] \propto p_{11}^{\alpha_1-1+y_{11}}p_{12}^{\alpha_2-1+y_{12}}p_{13}^{\alpha_3-1+y_{13}}p_{14}^{\alpha_4-1+y_{14}},$$

$$[\pi_2] \propto p_{21}^{\alpha_1-1+y_{21}}p_{22}^{\alpha_2-1+y_{22}}p_{23}^{\alpha_3-1+y_{23}}p_{24}^{\alpha_4-1+y_{24}},$$

$$[\pi_3] \propto p_{31}^{\alpha_1-1+y_{31}} p_{32}^{\alpha_2-1+y_{32}} p_{33}^{\alpha_3-1+y_{33}} p_{34}^{\alpha_4-1+y_{34}}, \text{ and}$$

$$[\pi_4] \propto p_{41}^{\alpha_1-1+y_{41}} p_{42}^{\alpha_2-1+y_{42}} p_{43}^{\alpha_3-1+y_{43}} p_{44}^{\alpha_4-1+y_{44}}.$$

Notice that these are proportional to the Dirichlet distribution. Thus, it may be stated that

$$[\pi_i] \propto Dir\left((\alpha_1 + y_{i1}), (\alpha_2 + y_{i2}), (\alpha_3 + y_{i3}), (\alpha_4 + y_{i4})\right)$$

This proportionality is very useful. Because the complete conditionals for the probabilities are proportional to a known distribution, it is unnecessary to use the Metropolis-Hastings algorithm. Rather, a simple Gibbs sampler will work. This was the result expected because the Dirichlet distribution is conjugate to the multinomial distribution, and the benefit of conjugacy is a known distribution for the complete conditional from which values may be drawn directly. The complete conditionals for the αs, unfortunately, are not proportional to any known distribution, and therefore a Gibbs sampler will not work. To sample from these distributions, it is necessary to use the Metropolis-Hastings algorithm.

3.7 Generated Data

Alpha parameters are now generated from the previously chosen hyperparameters. The generated values for loans that were current the time period before — that is, the first row of the transition matrix — are displayed in Table 3.2.

Using these parameters, four sets of probabilities are generated from the Dirichlet distribution — one for each quarter loans originated. These probabilities appear in Table 3.3. The probabilities follow the pattern that was expected when hyperpriors were chosen for this row of the transition matrix. Unlike the simple case where the data were chosen to be differentiable, the probabilities generated in this simulation

α	Generated α values:
α_1	874.15
α_2	60.87
α_3	.832
α_4	9.34

Table 3.2: α values generated for simulation for the first row of the transition matrix

are very similar to one another across quarters, which may make differentiation very difficult. The ability to differentiate will depend on the variance of the parameter and the amount of data associated with it.

Because these are simulated values, it is more difficult to specify differences in probabilities for each quarter than with human-created data. The most differentiable quarter is the fourth. This quarter has the largest probability of staying current and a much smaller probability of becoming delinquent than any of the other quarters. This quarter seems to be the best credit risk. The second quarter has the highest probability of prepaying and the highest probability of defaulting. The only noticeable feature of the first quarter is that the probability of delinquency is slightly higher than that of the other three quarters. The third quarter has a higher default rate that the other quarters.

The α parameters are also generated for the second row of the transition matrix, as the second row of the transition matrix acts independently of the first. These parameters are seen in Table 3.4. These values for α are much smaller than those

Quarter	Current	Delinquent	Default	Prepay
First Quarter	.926	.063	.000185	.011
Second Quarter	.920	.062	.00268	.016
Third Quarter	.936	.056	.00120	.0071
Fourth Quarter	.942	.047	.000180	.011

Table 3.3: Generated probabilities for all four quarters for the first row of the transition matrix

α	Generated α values:
α_1	210.68
α_2	103.81
α_3	45.86
α_4	24.78

Table 3.4: α values generated for simulation for the second row of the transition matrix

for the first row of the transition matrix. This drop in size is caused by the smaller amount of data in second row as compared to the first. The pattern chosen in the hyperparameters is still visible.

From these αs, probabilities are generated for the second row. These probabilities can be seen in Table 3.5. Because the αs are so different between the first and second rows of the transition matrix, the probabilities should be as well. Differences among quarters should also be less extreme because less data will be generated.

For this row of the transition matrix, the first and third quarters have the highest probability of returning to current. The fourth quarter has the smallest probability of returning to current and the highest probability of remaining delinquent. The second quarter has the highest probability of prepaying. Those in quarter one are most likely to default.

From each of these sets of probabilities, data are generated. The data for the first row of the transition matrix are found in Table 3.6, and the data for the second row of the transition matrix are found in Table 3.7.

Quarter	Current	Delinquent	Default	Prepay
First Quarter	.571	.250	.121	.059
Second Quarter	.549	.279	.107	.065
Third Quarter	.568	.267	.106	.058
Fourth Quarter	.515	.318	.117	.050

Table 3.5: Generated probabilities for all four quarters for the second row of the transition matrix

Quarter	Current	Delinquent	Default	Prepay
First Quarter	9244	635	1	120
Second Quarter	9200	597	27	176
Third Quarter	9330	586	9	75
Fourth Quarter	9457	447	1	95

Table 3.6: Generated data for all four quarters for the first row of the transition matrix

Quarter	Current	Delinquent	Default	Prepay
First Quarter	380	169	105	46
Second Quarter	408	183	61	48
Third Quarter	414	173	64	49
Fourth Quarter	382	212	73	33

Table 3.7: Generated data for all four quarters for the second row of the transition matrix

It is of interest to see whether the data exhibit the same patterns the generated probabilities do. If the data do not exhibit the same patterns, it will be impossible for the generated distributions to each probability will show the trends noticed above.

For the data from the first row of the transition matrix, the extremes noticed in quarter four seem to have translated perfectly into the data. The probability of staying current is the highest of the four quarters, and the probability of becoming delinquent is the lowest. For the second quarter, the number of prepaid loans is much higher, just as it was in the probabilities, and the number of defaults is much higher that the other quarters and is higher than expected. The probability for defaulting for the second quarter was only twice the next highest, but the number of loan is three times higher than the next closest. Quarter one does not show anything. A slightly higher delinquency rate is observed, but it is very close to that of other quarters.

For the data from the second row of the transition matrix, the third quarter shows a higher probability of becoming current, as expected, but the first quarter moves from one of the highest probabilities of moving to current to the lowest number of loans. This is an unfortunate result of less data, and is not necessarily surprising. The fourth quarter still has a small number of loans staying current, but this is now very close to the number of loans staying current in the first quarter. Delinquency is still highest for the fourth quarter, however. Although the second quarter had the highest rate of prepay, this does not appear in the loan data, as all numbers of prepay are very similar. The first quarter shows a much higher rate of default than the other three, as is seen in the probabilities.

All in all, most of the trends observed in the probability tables are still evident in the generated loan tables. Those differences still noticeable should be visible in the distribution of each probability. Those differences that were not visible in the data will not be seen in the same distributions.

3.8 Convergence of MCMC

To assess convergence, mixing plots are examined. Figure 3.1 shows the mixing plot for α_1 for the first row of the transition matrix. The graph shows no trends either up or down, and there is no reason to believe that the simulation has yet to reach the parameter space. It appears that the simulation does not need to run longer to reach the appropriate parameter space.

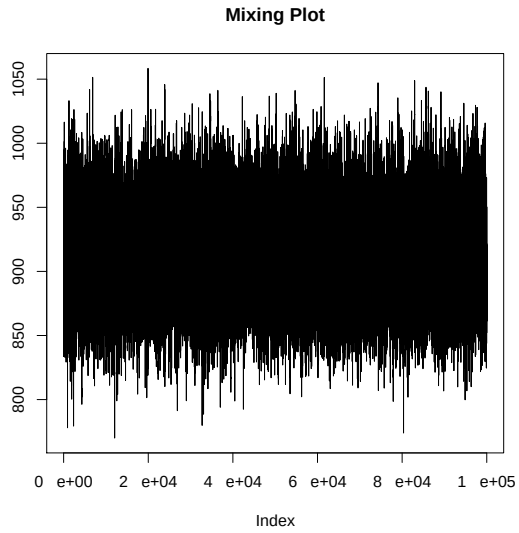

Figure 3.1: Mixing plot for α_1 for the first row

In order to make sure the parameter has converged, that the whole parameter space has been spanned, and that the candidate sigma has been chosen appropriately to span the parameter space, the mixing plots are more closely examined. This is shown in Figure 3.2. Periods where no change exists can be seen, indicating that the candidate sigma chosen is large enough, and multiple jumps are also apparent. The choice of candidate sigma seems appropriate, and the space seems to have been spanned. The parameter α_1 appears to have converged.

Similar graphs for α_2, α_3, and α_4, the same patterns are noted. All appear to have converged. This may be seen in Appendix C.

α	π_1	π_2	π_3	π_4
α_1	.069	.071	.068	.082
α_2	.053	.053	.052	.043
α_3	.256	.257	.167	.041
α_4	.077	.069	.084	.091

Table 3.8: Correlation Coefficients for αs and associated πs

Correlation between parameters may create problems with convergence even when mixing plots show no problems. If two parameters are highly correlated, a portion of the parameter space may not have been reached, and the convergence seen will be false. To make sure this is not an issue, correlations of interest will be looked at as a final check of convergence. A correlation of interest is a correlation involving the probability directly associated with each α and the α itself. These correlations are contained in Table 3.8

Looking at these correlations, it is easy to see that there is no concern developing from correlation. There are a few expected patterns. Probabilities associated with α_3 tend to be more highly correlated with α_3 than other probabilities are with the other αs. This is expected and follows the same pattern as the simple case. That is, the probabilities associated with very rare occurrences and thus, less data, are more dependent upon the α parameter than are those with more data. Though this pattern is visible, it is not of concern, as no correlation is higher than .257, which is not a strong enough correlation to cause any problems.

Figure 3.3 contains the density for α_1. Notice the vertical line drawn on the graph for the generated value. Although not directly in the middle, the line falls in the main portion of the graph, showing that the method has succeeded. The value used to generate data is contained in the main portion of the density.

The same pattern is observed in Figures C.10, C.11, and C.12 for the other αs. Although not necessarily in the center, the generated αs are always close to the mode

of the distribution.

The αs have converged satisfactorily, so it is meaningful to examine the probabilities. Since these were generated using a Gibbs Sampler, it is not necessary to examine convergence. The important factor here is whether or not the generated probability falls within the generated range. Examining Figures C.13 and the graphs for probabilities associated with α_2, α_3, and α_4 in Appendix C, it is apparent that the generated probabilities, marked with the vertical lines, are within the normal range of the simulated draws. Additionally, the patterns observed in the generated probabilities and the data seem to have translated into the graphs. For example, consider the probability of defaulting. It was previously noted that the second quarter had the highest probability, the fourth quarter had the lowest probability, and the third quarter had a high probability, but not the highest. The graphs show this relationship, as well as all other relationships noted previously. For this row, the methodology chosen appears to have worked very well.

The first row appears to have converged, but that is the row with more data. It is important to know whether the method will work with less data. To determine this, consider the second row of the original transition matrix. Because this row deals with an initial state of delinquent, it will contain much less data. Less data always makes a statistical procedure more volatile, so to determine whether the method is truly working, it is necessary to consider the second row as well as the first.

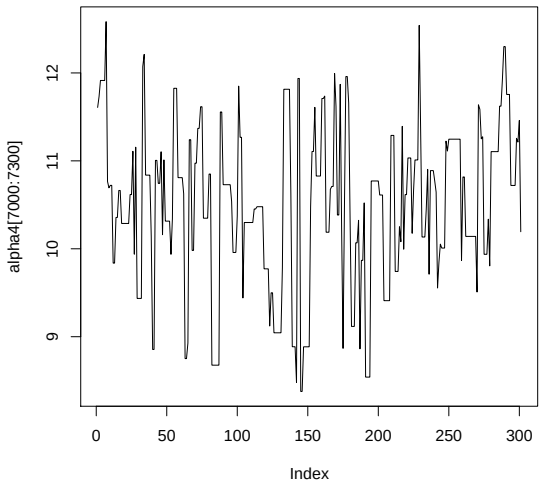

Figure 3.2: Mixing plot for α_1 over 300 observations

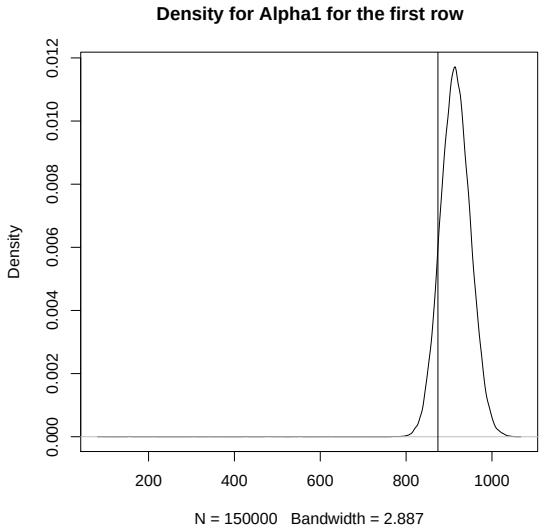

Figure 3.3: Density for α_1 for the first row

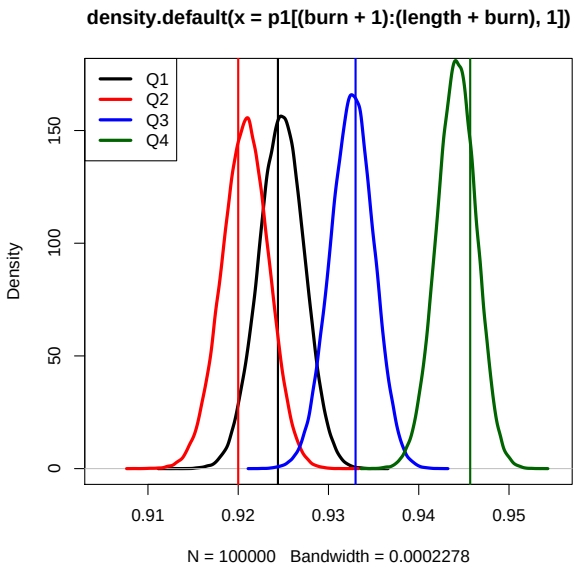

Figure 3.4: Densities for probabilities in the first row, first column (current to current) of the transition matrix

Again, consider α_1, but now examine mixing for the second row. Notice the stationarity demonstrated in Figure 3.5. There does not appear to be any systematic change in the mean of the graph, so there is no reason to assume convergence has not been reached. Looking at Figure 3.6, the choice for the candidate sigma appears to be appropriate, as there are periods where the simulated value doesn't change, but they are not large enough to think the candidate sigma is too large.

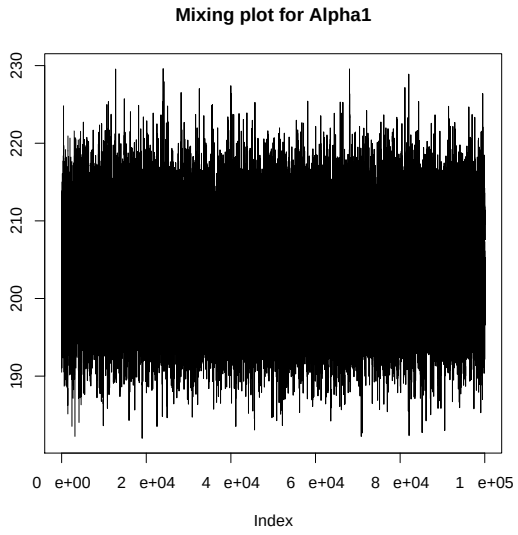

Figure 3.5: Mixing plot for α_1 for the second row

The plots for α_2, α_3, and α_4 show the same thing in Appendix C. Even though there is less data, it appears that convergence has still occurred. In the simulation of the first row, correlation was inversely proportional to the amount of data. Since this row has less data, it is necessary to examine the correlations here. The correlations for αs and associated probabilities for the second row are found in Table 3.9. As noticed for the first row, the same relationships that existed for the first row of the transition matrix exist here. No correlation is greater than .226, which is not large enough to worry about.

Although this table follows the same pattern that the previous table of correla-

α	π_1	π_2	π_3	π_4
α_1	.125	.122	.134	.106
α_2	.164	.170	.154	.176
α_3	.225	.213	.214	.226
α_4	.154	.171	.160	.192

Table 3.9: Correlation Coefficients for αs and associated πs for the second row of the transition matrix

tions did, that less data leads to a higher correlation; still, none of these correlations are high enough to be worrisome. It appears that the methodology has succeeded with the second row of the transition matrix. Because the method has been successful on the simulated data, the real data will be considered.

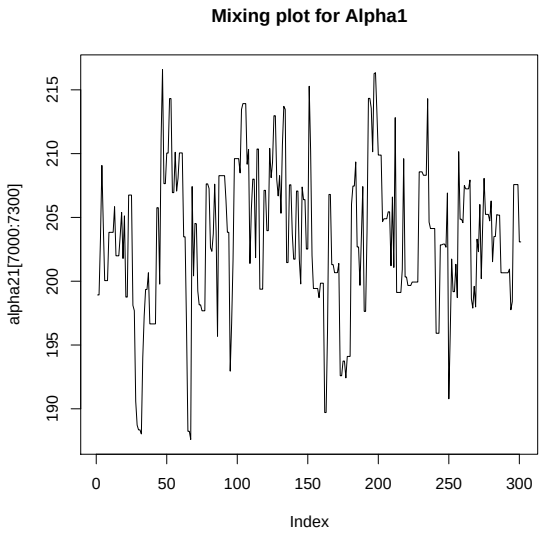

Figure 3.6: Mixing plot for 300 realizations of α_1 for the second row

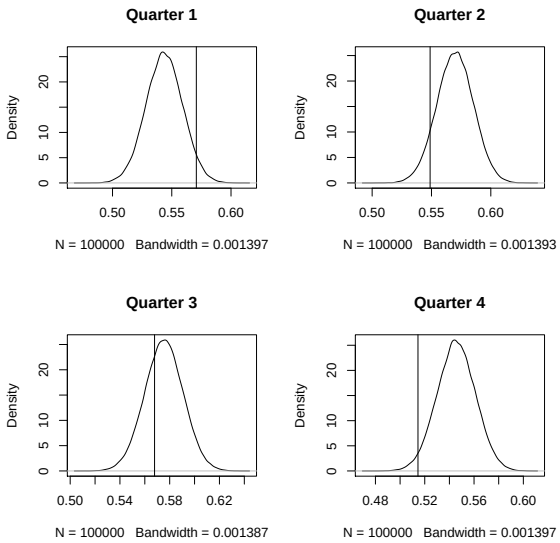

Figure 3.7: Density of probabilities associated with α_1 for the second row

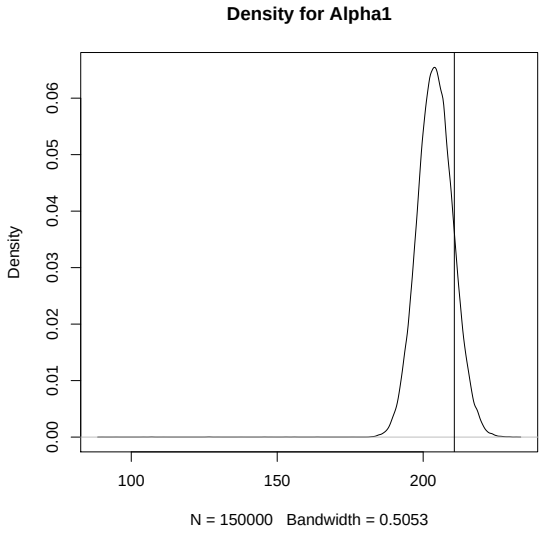

Figure 3.8: Density for α_1 for the second row

4. ANALYSIS OF REAL DATA

4.1 Data Description

Because of the success of the methodology on the simulated hierarchical data, actual data will now be considered. Consider an $8 billion portfolio in the subprime home equity market. These borrowers generally have weaker or damaged credit, which prevents them from qualifying for loans in the prime market. It should not be surprising, then, that loss rates in the subprime sector are greater than those in the prime market. These loans are fixed rate with first liens secured by residential real estate originated between 1979 and 2002, with 90% originated after 1995. Table 3 contains summary statistics on these 108,646 loans.

There are 8 states of the Markov chain model: current, loss, paid, and 1 to 30, 31 to 60, 61 to 90, 91 to 120, and 121+ days past due. The five delinquency states are determined by comparing the due date to the last day of the month. For example, a borrower whose payment is due on the 5th of the month and whose last payment was 1 Jan would be 23 days past due on 28 Feb, placing them in the 1 to 30 days past due state. This definition of delinquency obviously treats borrowers whose payments are due on the 5th and 25th the same even though the borrower whose payment is due on the 5th is more delinquent, but it is an accepted accounting practice. The 121 or more days past due state reflects seriously delinquent borrowers who may remain in this state while loss mitigation or foreclosure proceedings are considered. Loss and paid are absorbing states. For this example, all states of delinquency, from 1 to 121+, are considered delinquent. This compresses the eight by eight transition matrix into a four by four matrix with states current, delinquent, loss, and paid.

One approach to modeling the changing subprime lending environment is to estimate a different transition matrix for loans originating in each quarter. That

Table 4.1: Summary Statistics of an $8 Billion Portfolio of Subprime Home Equity Loans Secured by Residential Real Estate.

	Minimum	Q1	Median	Q3	Maximum
Interest Rate (%)	6.00	7.74	8.75	9.50	18.50
Loan Amount ($)	24,936	55,151	70,874	88,234	331,015
Loan-to-Value Ratio (%)	15.52	90.21	94.85	97.64	100.00

Percentage of Credit Report Derogatories	
12.95	Filed for Bankruptcy
11.59	At Least One NSF Check
59.31	At Least One Major Derogatory

Percentage of Mortgage Repayment Delinquency	
65.17	Never Delinquent
34.83	At Least One 30+
10.53	At Least One 60+
4.45	At Least One 90+

Quarter	Current	Delinquent	Default	Prepay
First Quarter	51682	2846	11	1443
Second Quarter	73835	4069	16	2187
Third Quarter	82838	4492	18	2303
Fourth Quarter	69240	3939	11	1971

Table 4.2: Data for borrowers whose loans originated in 1991 Q1, Q2, Q3, Q4 for the first row of the transition matrix

is, choose quarterly originations as the segmentation variable. This is somewhat problematic since older loans have been observed longer than new loans and the rare transition probabilities must have a sufficient sample size for every segment.

In order to determine whether this methodology is effective for real data, only the data from a single origination year, 1991, is considered. These data are contained in Tables 4.2 and 4.3. Table 4.2 contains data from the first row of the transition matrix, while Table 4.3 contains data from the second row.

These tables contain far more data than was simulated in the study. Because each quarter has a different number of loans, numbers do not compare directly. Therefore, to try to understand the data, two new tables are created. These tables have the probabilities for each row, and a total number of loans in the row.

In Table 4.4, the third quarter is shown to have the most loans. It also has a slightly higher probability of staying current than the other quarters, and none of the other probabilities associated with this quarter are extreme in any way. The first quarter has the fewest loans, but all the probabilities associated with it are

Quarter	Current	Delinquent	Default	Prepay
First Quarter	2532	7738	88	328
Second Quarter	3703	10496	110	546
Third Quarter	3941	11327	126	551
Fourth Quarter	3491	9822	100	467

Table 4.3: Data for the second row of the transition matrix

Number of Loans	Current	Delinquent	Default	Prepay
55982	.9232	.0508	.000196	.0258
80107	.9217	.0508	.0002	.0273
89651	.9240	.0501	.000201	.0257
75161	.9212	.0524	.000146	.0262

Table 4.4: Transition proportions computed from borrower repayment for the first row of the transition matrix by origination quarter

less extreme than those in the other quarters. The second quarter has the highest probability of prepaying as its only unique feature, while the fourth quarter has a much lower probability of the loan defaulting. It also has the largest delinquency rate, but this difference is not large.

In Table 4.5, the third quarter again has the most data, with the first quarter having the least. The first quarter has the lowest probability of returning to current, while the fourth quarter has the highest. While the fourth quarter is fairly close to the other quarters for this probability, however, the first quarter is much lower than the other quarters. The first quarter is the most likely to stay delinquent and the most likely to default. The second quarter is most likely to prepay, while the fourth quarter is the least likely to default, and is second least likely to prepay, with the first quarter being the least likely.

Number of Loans	Current	Delinquent	Default	Prepay
10686	.2369	.7241	.00824	.0307
14855	.2493	.7066	.00741	.0368
15945	.2472	.7104	.00790	.0346
13880	.2515	.7076	.00721	.0336

Table 4.5: Transition proportions computed from borrower repayment for the second row of the transition matrix by origination quarter

4.2 Assessment of Convergence

As in Chapter 3, the mixing plots are examined to assess convergence. First, the mixing plot for all observations is examined to determine if the simulation is still trending in one direction. Consider Figure 4.1. These realizations appear stationary, so it appears the process does not need to continue longer if a closer examination of the mixing plot shows appropriate mixing.

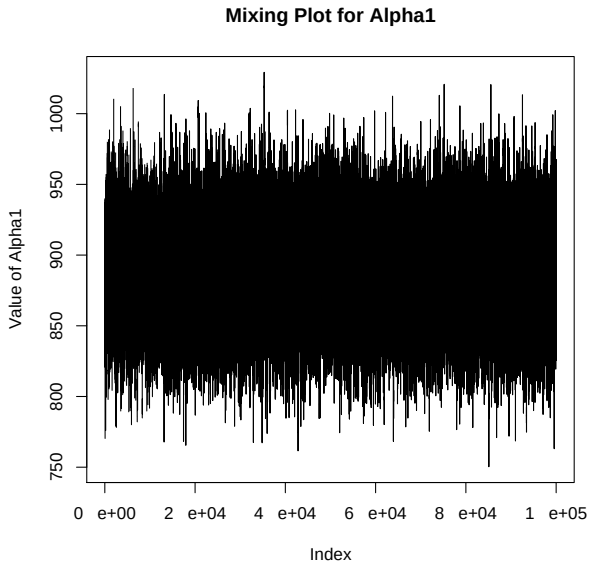

Figure 4.1: Mixing plot for α_1 for the first row of real data

For the same parameter, now consider a subset of three hundred of the realizations. If this plot shows long time periods where the value of the parameter does not change, the candidate sigma is too large and should be adjusted down before convergence is reached. If the plot shows the parameter changing on every iteration, the candidate sigma is too small. In Figure 4.2, it appears that the choice of candidate sigma is appropriate. This plot appears to have converged, as there is no apparent trend and the candidate sigma is appropriate. All other α parameters for both rows of the transition matrix appear to have converged, as can be seen in Appendix D.

α	Hyperprior μ	Simulated μ	Hyperprior σ^2	Simulated σ^2
α_1	900.5	887.4	1642.7	1165.1
α_2	60	57.6	4.0	3.4
α_3	1	.73	.14	.044
α_4	10	14.2	1.27	1.5

Table 4.6: Comparison between hyperprior means and variances and simulated means and variances for αs for the first row of the transition matrix

4.3 Examination of Results

Convergence has been reached, so now it is of interest to examine the distribution of each parameter. It is possible that though convergence has been reached, the analysis is valueless. If the posterior distributions for the parameters do not show the patterns that were seen in the data, the analysis has not helped in understanding the data.

Figures 4.3, 4.4, 4.5, and 4.6 show the distributions of the alpha parameters. These values all look about as expected. For a more concise look at these distributions and how they differ from prior knowledge, Table 4.6 contains the hyperprior means and variances compared to the means of the generated αs and their variances. Notice that the means are close in value, but not the same. The data has shifted the means and, in most cases, has decreased the variance of the parameter. Only α_4 does not have a decreasing variance, and the increase is slight. Most of the hyperprior distributions were more variable to allow for uncertainty. The prepay probability seems to simply be more variable than expected.

Now, consider the probabilities associated with each α. Previously, associations between these probabilities were noted. Figure 4.7 shows the density plots for the probability associated with α_1 for each quarter. In this figure, the fourth quarter and second quarter probabilities seem lower than the first quarter and third quarter probabilities. And, in fact, the table showing the probabilities indicates that the first

and third quarter probabilities are higher than the second and fourth quarter. It can also be seen that the first quarter has a wider, more spread distribution. This seems appropriate, since the first quarter has the least data in this cell. The difference between the second and fourth quarters does not appear large, but the difference between the two clusters, first and third, second and fourth, does appear quite large.

Figure 4.8 also shows expected patterns. The first and second quarters have the same probability, but the first quarter has less data. This pattern is again visible, with the distribution of the first quarter probability wider and less specific. The fourth quarter has the highest mean, by far, which is appropriate because the probability is so high in the table; additionally, because the third quarter has the most data, this curve is the narrowest.

The third column, the probability of defaulting, shows little differentiation. This is as expected because the probabilities were extremely close and there were not many loans here. These distributions are found in Figure 4.9. The fourth quarter has a slightly lower probability, and the other curves are virtually on top of one another, with the only difference being the amount of data. Though the fourth quarter's probability is twenty-five percent lower than the other probabilities, which is larger than any other relative difference, very little difference is visible in the graph due to the lack of data.

In Figure 4.10, the second quarter probability is the most differentiated from the other posterior densities. And although the first and third quarters have virtually the same mean, the difference in the amount of data differentiates the curves. The fourth quarter is in between, but is closer to the first and third quarters than the second quarter.

Now consider the second row. This row has much less data and so it is of interest to notice whether results, which are so apparent in the data associated with the first row of the transition matrix, are visible for this row. Figures 4.11, 4.12, 4.13,

α	Hyperprior μ	Simulated μ	Hyperprior σ^2	Simulated σ^2
α_1	200	139.9	42.0	16.7
α_2	100	149.0	18.7	30.5
α_3	50	16.9	13.9	1.2
α_4	25	19.9	3.1	1.4

Table 4.7: Comparison between hyperprior means and variances and simulated means and variances for αs for the second row of the transition matrix

and 4.14 contain the densities for αs associated with the second row of the transition matrix.

Table 4.7 shows that the same pattern occurs with these αs as with the αs associated with the first row of the transition matrix. Each α may be shrunk back, but the variances of those αs are also smaller, except for α_2. It seems that the prior means chosen had the wrong relationship. Instead of most people returning to current from a delinquent status, it seems that most loans in a delinquent status will stay there. The probability of staying delinquent is also more variable than expected in the hyperpriors. All other αs for this row were shrunk back and given a smaller variance. The shrinking of all the αs together indicates that the hierarchical relationship is not as strong as the hyperpriors made it seem. Smaller αs lead to a larger variance for each probability generated from the Dirichlet distribution and less influence of the hierarchy.

Now, consider the probabilities associated with these αs. Figure 4.15 shows the first quarter with a much lower probability as well as a wider distribution due to less data. The second, third, and fourth quarters have very similar probabilities, and the means increase in almost perfect intervals, with the fourth quarter being the highest, as expected.

Figure 4.16 shows the distributions for the probabilities associated with staying delinquent. As noticed from the αs, these are the largest probabilities. Although this was not the case in the hyperpriors, it can be seen in the data. The first quarter is by

far the largest probability. This density is almost completely unique from the other densities. The third quarter is slightly higher than the second and fourth quarter, which appear almost identical.

Now, considering loans that move to default, the first quarter is again the largest probability here, but it is less significant than in other states due to the fact that very few loans default. There is a lot of overlapping area between these curves, making it difficult to differentiate one of them. Since this column only has 424 loans between all the quarters, the dispersed distributions make sense.

For the probability of prepaying, the first quarter has a much lower probability than any of the others, and the second quarter is much higher. The third and fourth quarters appear very much the same, except that the third quarter has the narrowest distribution because it has the most data.

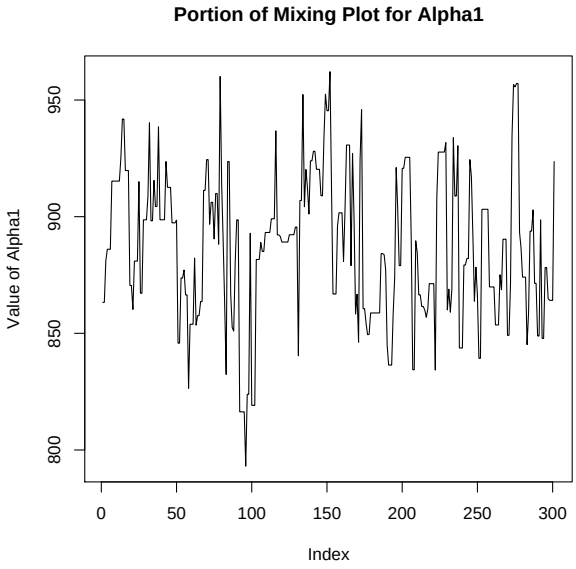

Figure 4.2: Closer Examination of mixing plot for α_1 for the first row of real data

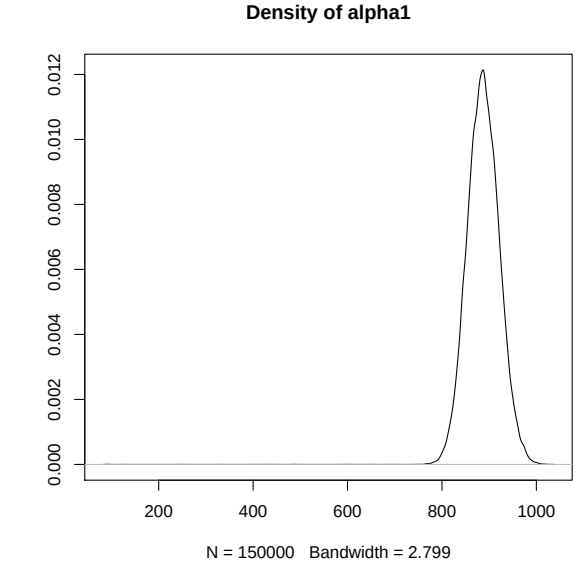

Figure 4.3: Density for α_1 for the first row

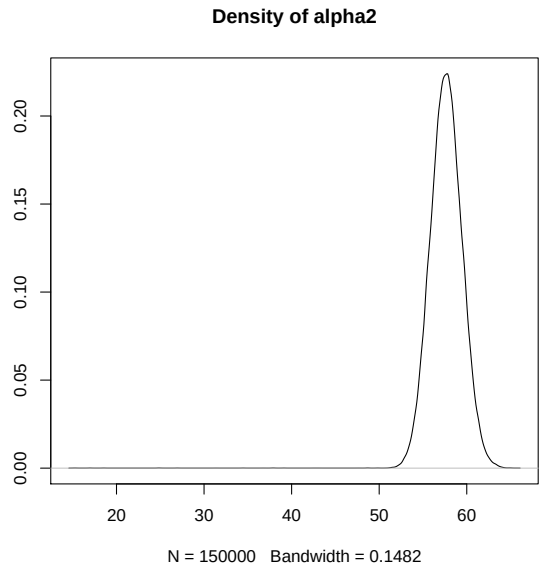

Figure 4.4: Density for α_2 for the first row

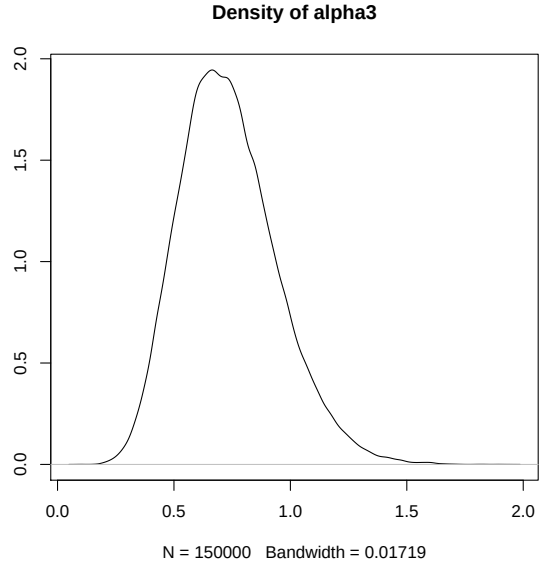

Figure 4.5: Density for α_3 for the first row

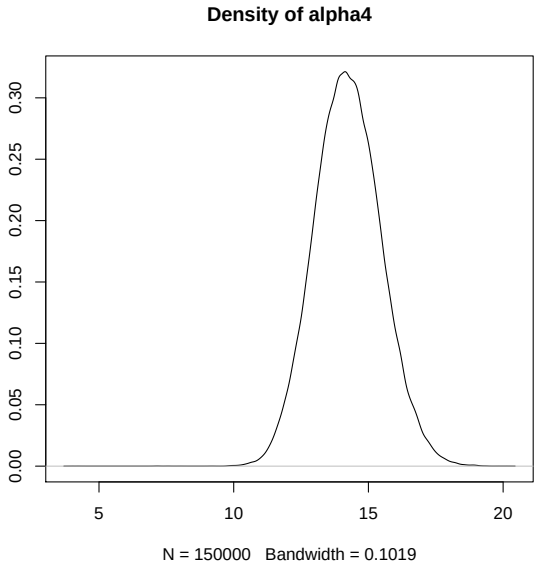

Figure 4.6: Density for α_4 for the first row

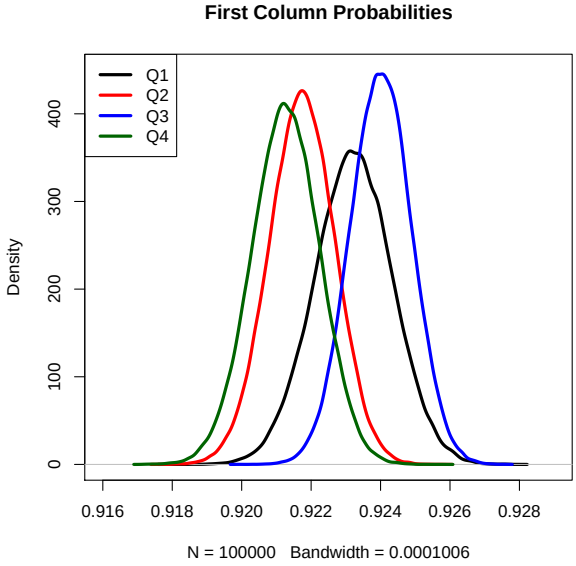

Figure 4.7: Density for probabilities by quarter associated with α_1 for the first row

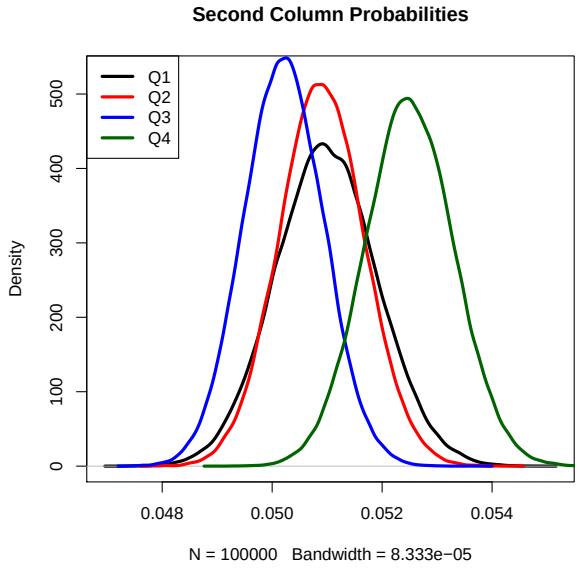

Figure 4.8: Density for probabilities by quarter associated with α_2 for the first row

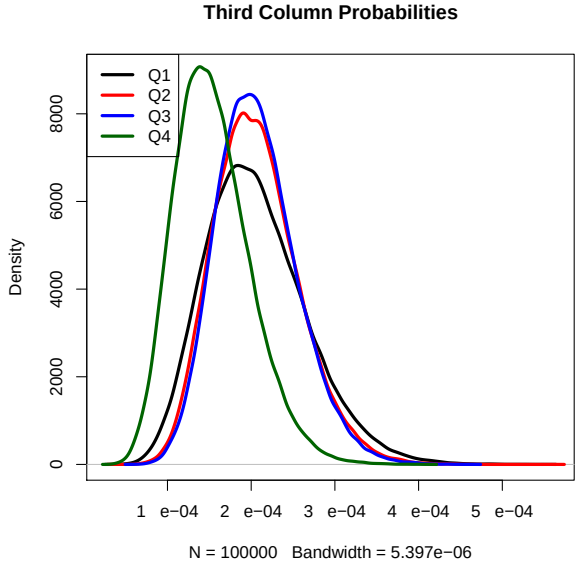

Figure 4.9: Density for probabilities by quarter associated with α_3 for the first row

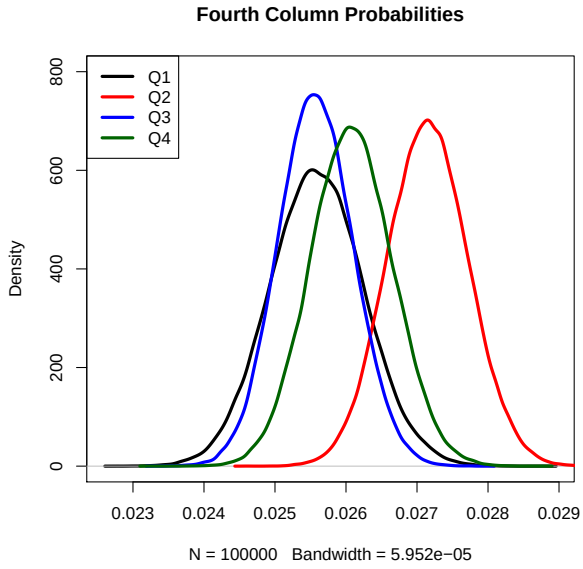

Figure 4.10: Density for probabilities by quarter associated with α_4 for the first row

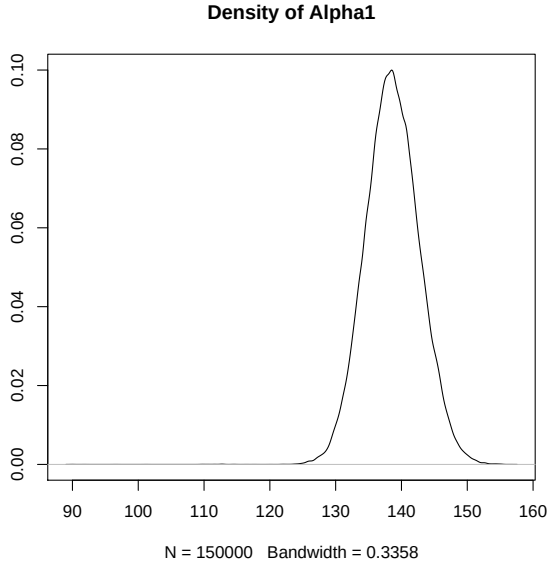

Figure 4.11: Density for α_1 for the second row

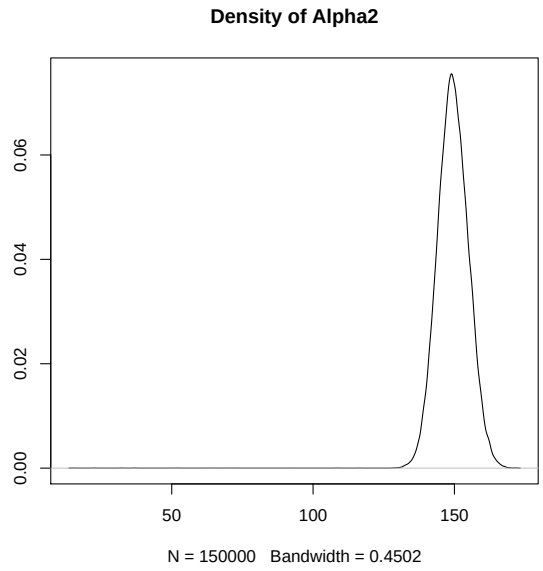

Figure 4.12: Density for α_2 for the second row

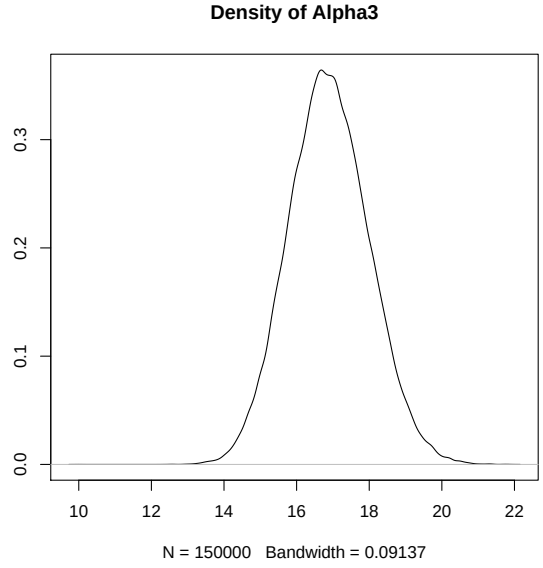

Figure 4.13: Density for α_3 for the second row

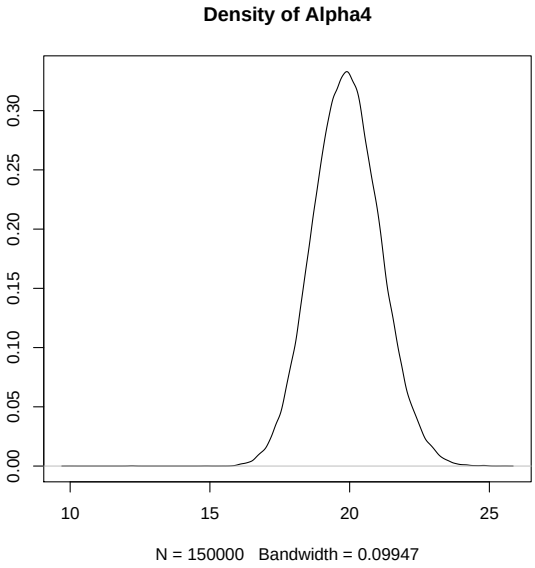

Figure 4.14: Density for α_4 for the second row

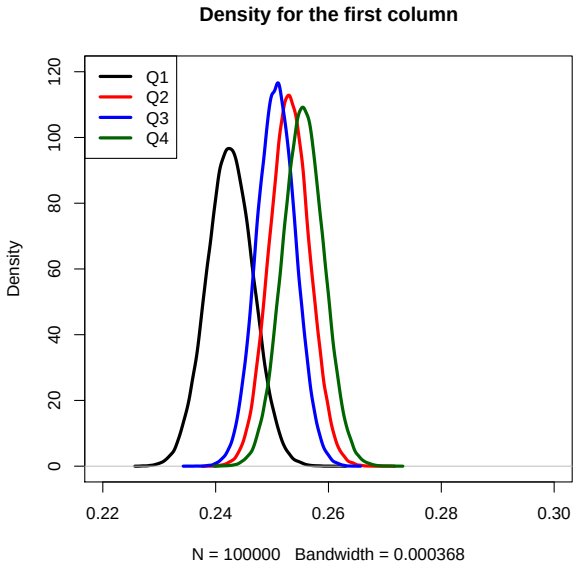

Figure 4.15: Density for probabilities by quarter associated with α_1 for the second row

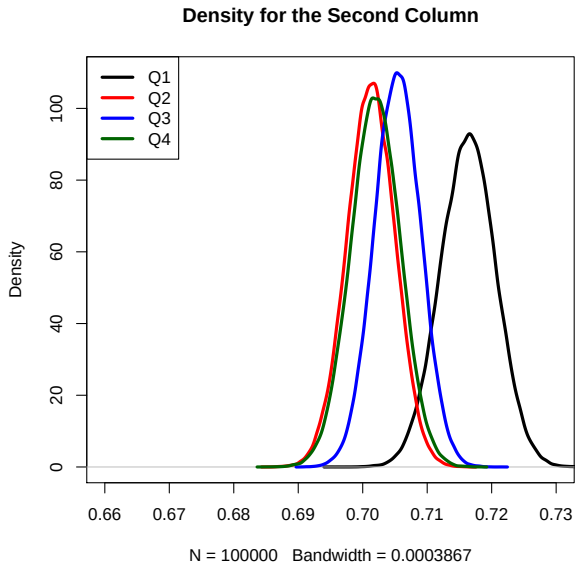

Figure 4.16: Density for probabilities by quarter associated with α_2 for the second row

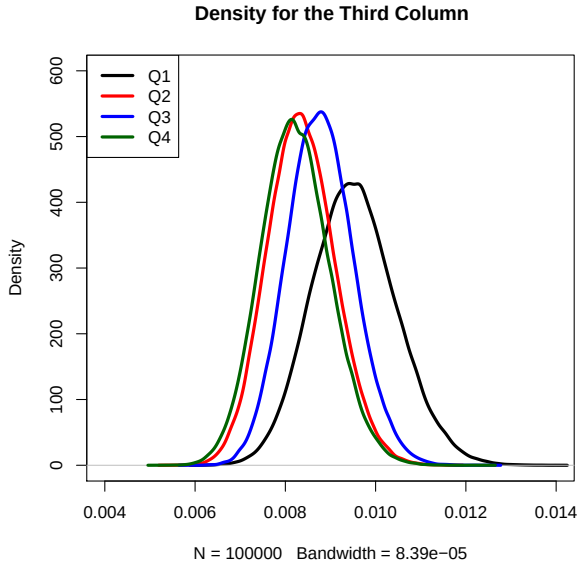

Figure 4.17: Density for probabilities by quarter associated with α_3 for the second row

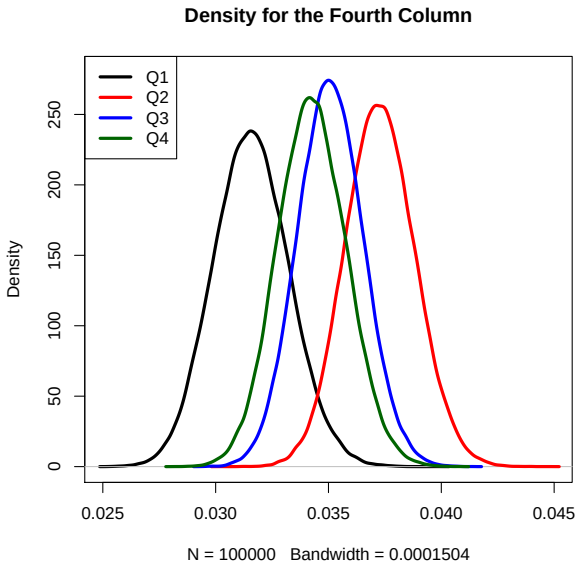

Figure 4.18: Density for probabilities by quarter associated with α_4 for the second row

5. CONCLUSIONS

Having considered both a simulated data set as well as real data, a Bayesian hierarchical model appears to perform well in the analysis of a transition matrix. All parameters that were estimated using the Metropolis-Hastings algorithm appear to have converged, even when poor starting positions were chosen. The fact that some rows have little data did not appear to cause any problems with convergence.

When examining the posterior distributions of the same probability across different quarters, differences emerged. This was true even when examining probabilities that did not have much data associated with them, like the probability of defaulting given a current status. Therefore, this methodology is sensitive enough to pick up differences between quarters even with less data. Had this not been the case, the methodology would have been useless, even if the parameters did converge. This differentiation is important, as it might allow a lending institution to set different levels of reserves for loans that had different origination dates, and thus minimize both the level of reserves as well as the risk associated with a bad loan.

One step that was not taken in this project but ought to be considered is the performance of this methodology in a situation where a location in the transition matrix has no data. Whether this lack of data would cause problems with convergence is unknown, but this factor could become important when the data being analyzed come from more recent years, where even more sparse data is expected.

Finally, it would be interesting to determine whether this methodology would work if applied over multiple years. Only one year was considered here, with the α parameters being generated only once for each row. If each year were considered a draw from the α distributions, it is of interest to know whether the estimation of the αs improve, or whether the addition of years would simply make convergence for these parameters more challenging. The effect of analyzing multiple years at once is

something that should be explored further.

BIBLIOGRAPHY

Chen, M.-H. and Schmeiser, B. (1993), "Performance of the Gibbs, Hit-and-Run, and Metropolis Samplers," *Journal of Computational and Graphical Statistics*, 2, 251–272.

Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49, 327–335.

Cowles, M. K. and Carlin, B. P. (1996), "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, 91, 883–904.

Gelman, A. (2006), "Multilevel (Hierarchical Modeling: What It Can and Cannot Do," *Technometrics*, 48, 432–435.

Gelman, A. and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472.

Gilks, W. R., Roberts, G. O., and George, E. I. (1994), "Adaptive Direction Sampling," *The Statistician*, 43, 179–189.

Hillier, F. S. and Lieberman, G. J. (1995), *Introduction to Operations Research*, McGraw-Hill Inc., sixth ed.

Jensen, P. A. and Bard, J. F. (2003), *Operations Research: Models and Methods*, Wiley, seventh ed.

Polson, N. (1996), "Convergence of Markov Chain Monte Carlo Algorithms," *Bayesian Statistics 5*, 6, 297–321.

Raftery, A. E. and Lewis, S. M. (1992), "[Practical Markov Chain Monte Carlo]:

Comment: One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo," *Statistical Science*, 7, 493–497.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701–1728.

Western, B. (1998), "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach," *American Journal of Political Science*, 42, 1233–1259.

A. DATA GENERATION CODE

```
###  Simulation of Hierarchical Model ###

library(MCMCpack) library(multinomRob)

a1<-c(500,900,7,80) a2<-c(960,520,180,205) b1<-c(5/9,15,7,8)
b2<-c(4.8,5.2,3.6,8.1)

prealpha1<-rgamma(4,a1,b1) prealpha2<-rgamma(4,a2,b2) .
prob1<-rdirichlet(4,prealpha1) prob2<-rdirichlet(4,prealpha2)

x1<-matrix(data=NA,ncol=4,nrow=4) x2<-matrix(data=NA,ncol=4,nrow=4)

for (i in 1:4){ x1[i,]<-rmultinomial(10000,prob1[i,])
x2[i,]<-rmultinomial(700,prob2[i,]) }

### Data is contained in x1 and x2.  True probabilities are
contained in prob1 and prob2 ###
```

```
### Skeleton Code for MCMC on simulated data set ###

dtn<-function(x,mu,sigma,lowcutpoint){
dnorm(x,mu,sigma)/(1-pnorm(lowcutpoint,mu,sigma))}

rtn<-function(x,mu,sigma,lowcutpoint){
vec<-matrix(NA,x,1)
    for(i in 1:x){
    check<-0
    while(check==0){
    try<-rnorm(1,mu,sigma)
    if (try>lowcutpoint){check<-1}
    }
vec[i,1]<-try
}
vec
}

ga1<-function(alpha1,alpha2,alpha3,alpha4,p11,p21,p31,p41,a1,b1){
4*(lgamma(alpha1 + alpha2 + alpha3 + alpha4)-lgamma(alpha1)) +
(alpha1-1)*(log(p11)+log(p21)+log(p31)+log(p41)) +
(a1-1)*log(alpha1) - (alpha1*b1)}

ga2<-function(alpha1,alpha2,alpha3,alpha4,p12,p22,p32,p42,a2,b2){
4*(lgamma(alpha1 + alpha2 + alpha3 + alpha4)-lgamma(alpha2)) +
(alpha2-1)*(log(p12)+log(p22)+log(p32)+log(p42)) +
(a2-1)*log(alpha2) - (alpha2*b2)}

ga3<-function(alpha1,alpha2,alpha3,alpha4,p13,p23,p33,p43,a3,b3){
4*(lgamma(alpha1 + alpha2 + alpha3 + alpha4)-lgamma(alpha3)) +
(alpha3-1)*(log(p13)+log(p23)+log(p33)+log(p43)) +
(a3-1)*log(alpha3) - (alpha3*b3)}

ga4<-function(alpha1,alpha2,alpha3,alpha4,p14,p24,p34,p44,a4,b4){
4*(lgamma(alpha1 + alpha2 + alpha3 + alpha4)-lgamma(alpha4)) +
(alpha4-1)*(log(p14)+log(p24)+log(p34)+log(p44)) +
(a4-1)*log(alpha4) - (alpha4*b4)}

a1<-500
a2<-900
```

```
a3<-7
a4<-80
b1<-(5/9)
b2<-15
b3<-7
b4<-8

y11<-51682
y12<-2846
y13<-11
y14<-1443
y21<-73835
y22<-4069
y23<-16
y24<-2187
y31<-82838
y32<-4492
y33<-18
y34<-2303
y41<-69240
y42<-3939
y43<-11
y44<-1971


length <- 100000
burn <-   50000
ralpha1 <- numeric(length+burn)
ralpha2 <- numeric(length+burn)
ralpha3 <- numeric(length+burn)
ralpha4 <- numeric(length+burn)
rp1 <- matrix(nrow=length+burn,ncol=4,NA)
rp2 <- matrix(nrow=length+burn,ncol=4,NA)
rp3 <- matrix(nrow=length+burn,ncol=4,NA)
rp4 <- matrix(nrow=length+burn,ncol=4,NA)


ralpha1[1]<-90
ralpha2[1]<-15
ralpha3[1]<-.1
ralpha4[1]<-4

rp1[1,]<-c(.93,.06,.001,.009)
rp2[1,]<-c(.93,.06,.001,.009)
rp3[1,]<-c(.93,.06,.001,.009)
```

```
rp4[1,]<-c(.93,.06,.001,.009)

candsig.a1 <- 50
candsig.a2 <- 3
candsig.a3 <- .3
candsig.a4 <- 2


for(i in 2:(length+burn)){
#Update alpha1
    ralpha1[i] <-ralpha1[i-1]
    new<-rtn(1,ralpha1[i-1],candsig.a1,0)
    old <-ralpha1[i-1]
    llo<-ga1(old,ralpha2[i-1],ralpha3[i-1],ralpha4[i-1],rp1[i-1,1]
    ,rp2[i-1,1],rp3[i-1,1],rp4[i-1,1],a1,b1)
    +log(dtn(new,old,candsig.a1,0))
    lln<-ga1(new,ralpha2[i-1],ralpha3[i-1],ralpha4[i-1],rp1[i-1,1]
    ,rp2[i-1,1],rp3[i-1,1],rp4[i-1,1],a1,b1)
    +log(dtn(old,new,candsig.a1,0))
    uu<-runif(1,0,1)
    if(log(uu)<(lln-llo)){ralpha1[i]<-new}

# Update alpha2
    ralpha2[i] <-ralpha2[i-1]
    new<-rtn(1,ralpha2[i-1],candsig.a2,0)
    old <-ralpha2[i-1]
    llo<-ga2(ralpha1[i],old,ralpha3[i-1],ralpha4[i-1],rp1[i-1,2]
    ,rp2[i-1,2],rp3[i-1,2],rp4[i-1,2],a2,b2)+
    log(dtn(new,old,candsig.a2,0))
    lln<-ga2(ralpha1[i],new,ralpha3[i-1],ralpha4[i-1],rp1[i-1,2]
    ,rp2[i-1,2],rp3[i-1,2],rp4[i-1,2],a2,b2)+
    log(dtn(old,new,candsig.a2,0))
    uu<-runif(1,0,1)
    if(log(uu)<(lln-llo)){ralpha2[i]<-new}

# Update a3
    ralpha3[i] <-ralpha3[i-1]
    new<-rtn(1,ralpha3[i-1],candsig.a3,0)
    old <-ralpha3[i-1]
    llo<-ga3(ralpha1[i],ralpha2[i],old,ralpha4[i-1],rp1[i-1,3],
    rp2[i-1,3],rp3[i-1,3],rp4[i-1,3],a3,b3)+
    log(dtn(new,old,candsig.a3,0))
    lln<-ga3(ralpha1[i],ralpha2[i],new,ralpha4[i-1],rp1[i-1,3],
    rp2[i-1,3],rp3[i-1,3],rp4[i-1,3],a3,b3)+
    log(dtn(old,new,candsig.a3,0))
```

```
    uu<-runif(1,0,1)
    if(log(uu)<(lln-llo)){ralpha3[i]<-new}

# Update a4
    ralpha4[i] <-ralpha4[i-1]
    new<-rtn(1,ralpha4[i-1],candsig.a4,0)
    old <-ralpha4[i-1]
    llo<-ga4(ralpha1[i],ralpha2[i],ralpha3[i],old,rp1[i-1,4],
    rp2[i-1,4],rp3[i-1,4],rp4[i-1,4],a4,b4)+
    log(dtn(new,old,candsig.a4,0))
    lln<-ga4(ralpha1[i],ralpha2[i],ralpha3[i],new,rp1[i-1,4],
    rp2[i-1,4],rp3[i-1,4],rp4[i-1,4],a4,b4)+
    log(dtn(old,new,candsig.a4,0))
    uu<-runif(1,0,1)
    if(log(uu)<(lln-llo)){ralpha4[i]<-new}

# Update p1 - p4

rp1[i,]<-rdirichlet(1,c(y11+ralpha1[i],y12+ralpha2[i],
y13+ralpha3[i],y14+ralpha4[i]))

rp2[i,]<-rdirichlet(1,c(y21+ralpha1[i],y22+ralpha2[i],
y23+ralpha3[i],y24+ralpha4[i]))

rp3[i,]<-rdirichlet(1,c(y31+ralpha1[i],y32+ralpha2[i],
y33+ralpha3[i],y34+ralpha4[i]))

rp4[i,]<-rdirichlet(1,c(y41+ralpha1[i],y42+ralpha2[i],
y43+ralpha3[i],y44+ralpha4[i]))
}

pdf("C://Documents and Settings//Rebecca//My Documents//Statistics
//Master's Project//ralpha1row1density.pdf")
plot(density(ralpha1), main = 'Density of alpha1',ylab = ' ')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents//Statistics
//Master's Project//ralpha2row1density.pdf")
plot(density(ralpha2), main = 'Density of alpha2',ylab = ' ')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents//Statistics
//Master's Project//ralpha3row1density.pdf")
plot(density(ralpha3), main = 'Density of alpha3',ylab = ' ')
dev.off()
```

```
pdf("C://Documents and Settings//Rebecca//My Documents//Statistics
//Master's Project//ralpha4row1density.pdf")
plot(density(ralpha4), main = 'Density of alpha4',ylab = ' ')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//ralpha1row1line.pdf")
plot(ralpha1[(burn+1):(length+burn)],type='l',
main = 'Mixing Plot for Alpha1',ylab = 'Value of Alpha1')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//ralpha2row1line.pdf")
plot(ralpha2[(burn+1):(length+burn)],type='l',
main = 'Mixing Plot for Alpha2',ylab = 'Value of Alpha2')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//ralpha3row1line.pdf")
plot(ralpha3[(burn+1):(length+burn)],type='l',
main = 'Mixing Plot for Alpha3',ylab = 'Value of Alpha3')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//ralpha4row1line.pdf")
plot(ralpha4[(burn+1):(length+burn)],type='l',
main = 'Mixing plot for alpha4',ylab = 'Value of Alpha4')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//ralpha1row1check.pdf")
plot(ralpha1[7000:7300],type='l',
main = 'Portion of Mixing Plot for Alpha1',
ylab='Value of Alpha1')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//ralpha2row1check.pdf")
plot(ralpha2[7000:7300],type='l',
main = 'Portion of Mixing Plot for Alpha2',
ylab='Value of Alpha2')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
```

```
//Statistics//Master's Project//ralpha3row1check.pdf")
plot(ralpha3[7000:7300],type='l',
main = 'Portion of Mixing Plot for Alpha3',
ylab='Value of Alpha3')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//ralpha4row1check.pdf")
plot(ralpha4[7000:7300],type='l',
main = 'Portion of Mixing Plot for Alpha4',
ylab='Value of Alpha4')
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//rpalpha1row1density.pdf")
plot(density(rp1[(burn+1):(length+burn),1]),xlim=c(.916,.929),
ylim=c(0,450),lwd=3,main='First Column Probabilities')
lines(density(rp2[(burn+1):(length+burn),1]),col='red',lwd=3)
lines(density(rp3[(burn+1):(length+burn),1]),col='blue',lwd=3)
lines(density(rp4[(burn+1):(length+burn),1]),col='darkgreen',lwd=3)
legend("topleft",legend=c("Q1","Q2","Q3","Q4"),
col=c("black","red","blue","darkgreen"),lty=c(1,1,1,1),lwd=c(3,3,3,3))
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//rpalpha2row1density.pdf")
plot(density(rp1[(burn+1):(length+burn),2]),lwd=3,
 ylim=c(0,530),main='Second Column Probabilities')
lines(density(rp2[(burn+1):(length+burn),2]),col='red',lwd=3)
lines(density(rp3[(burn+1):(length+burn),2]),col='blue',lwd=3)
lines(density(rp4[(burn+1):(length+burn),2]),col='darkgreen',lwd=3)
legend("topleft",legend=c("Q1","Q2","Q3","Q4"),
col=c("black","red","blue","darkgreen"),lty=c(1,1,1,1),lwd=c(3,3,3,3))
dev.off()

pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//rpalpha3row1density.pdf")
plot(density(rp1[(burn+1):(length+burn),3]),
ylim=c(0,9000),lwd=3,main='Third Column Probabilities')
lines(density(rp2[(burn+1):(length+burn),3]),col='red',lwd=3)
lines(density(rp3[(burn+1):(length+burn),3]),col='blue',lwd=3)
lines(density(rp4[(burn+1):(length+burn),3]),col='darkgreen',lwd=3)
legend("topleft",legend=c("Q1","Q2","Q3","Q4"),
col=c("black","red","blue","darkgreen"),lty=c(1,1,1,1),lwd=c(3,3,3,3))
dev.off()
```

```
pdf("C://Documents and Settings//Rebecca//My Documents
//Statistics//Master's Project//rpalpha4row1density.pdf")
plot(density(rp1[(burn+1):(length+burn),4]),
ylim=c(0,800),lwd=3,main='Fourth Column Probabilities')
lines(density(rp2[(burn+1):(length+burn),4]),col='red',lwd=3)
lines(density(rp3[(burn+1):(length+burn),4]),col='blue',lwd=3)
lines(density(rp4[(burn+1):(length+burn),4]),col='darkgreen',lwd=3)
legend("topleft",legend=c("Q1","Q2","Q3","Q4"),
col=c("black","red","blue","darkgreen"),lty=c(1,1,1,1),lwd=c(3,3,3,3))
dev.off()
```

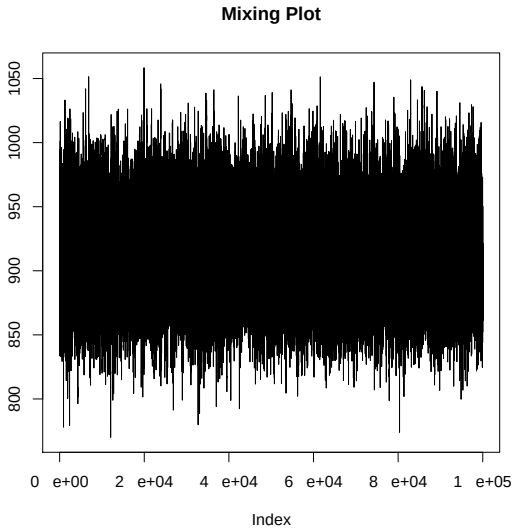

Figure C.1: Time plot for α_1 for the first row

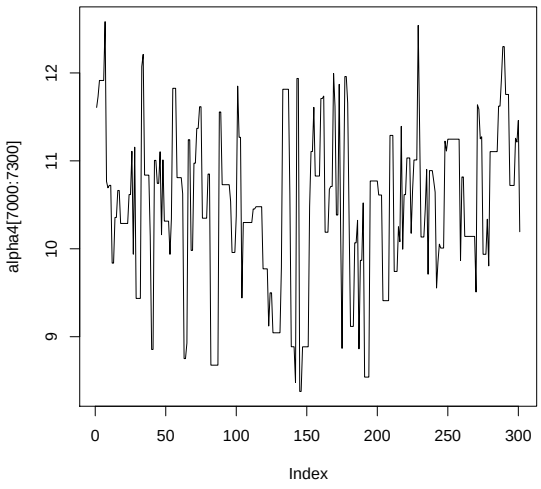

Figure C.2: Time plot for α_1 over 300 observations

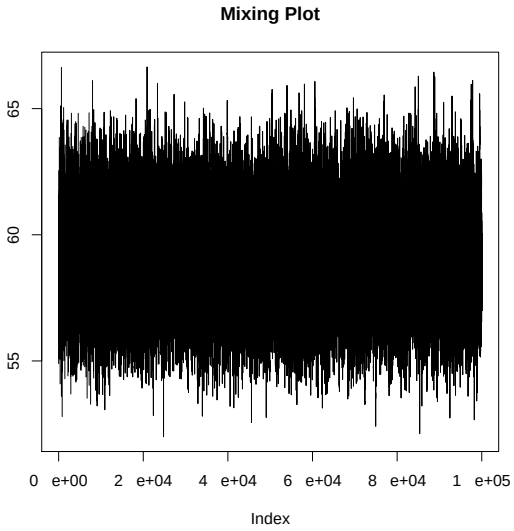

Figure C.3: Time plot for α_2 for the first row

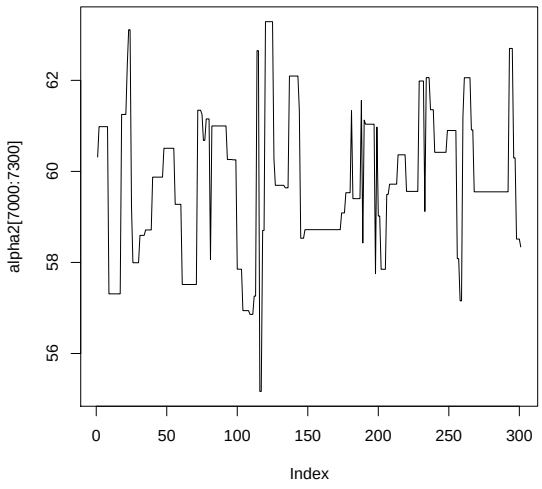

Figure C.4: Time plot for α_2 over 300 observations

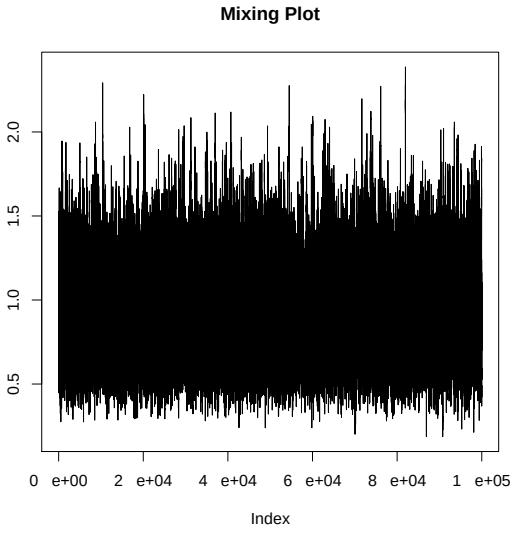

Figure C.5: Time plot for α_3 for the first row

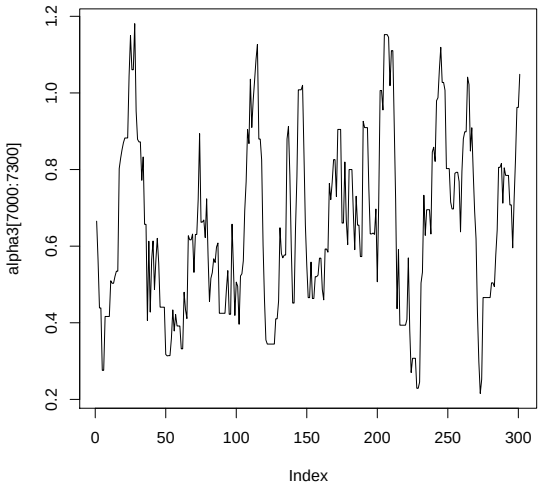

Figure C.6: Time plot for α_3 over 300 observations

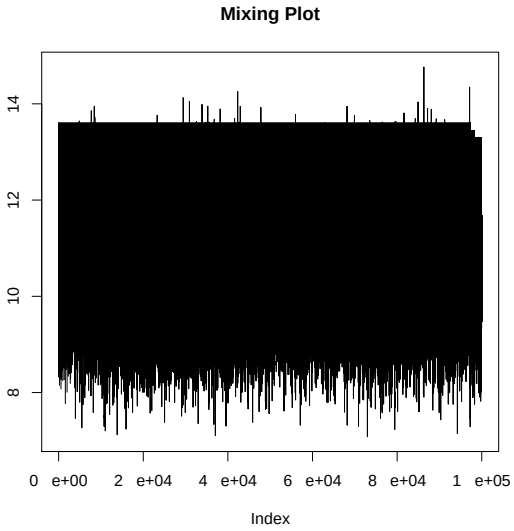

Figure C.7: Time plot for α_4 for the first row

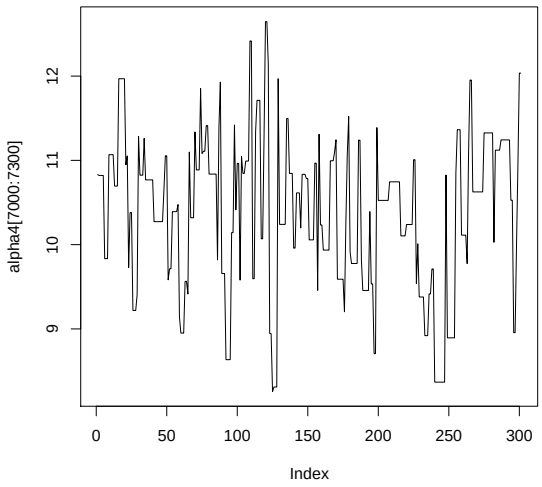

Figure C.8: Time plot for α_4 over 300 observations

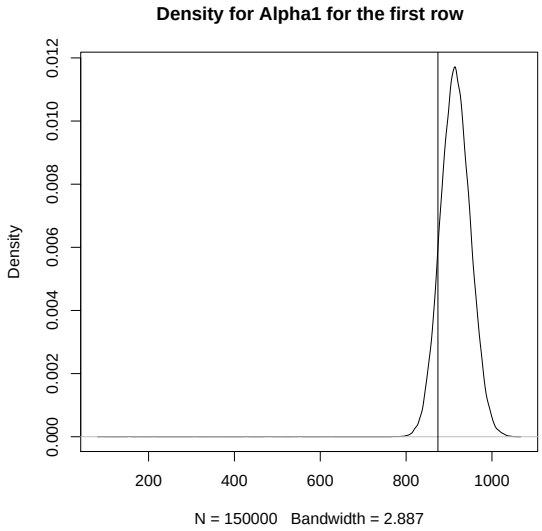

Figure C.9: Density for α_1 for the first row

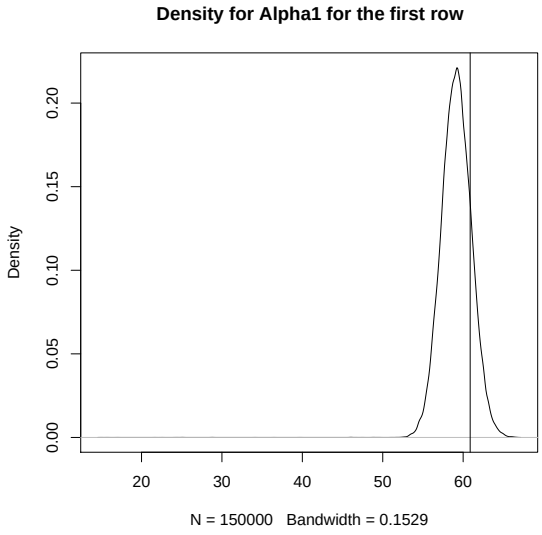

Figure C.10: Density for α_2 for the first row

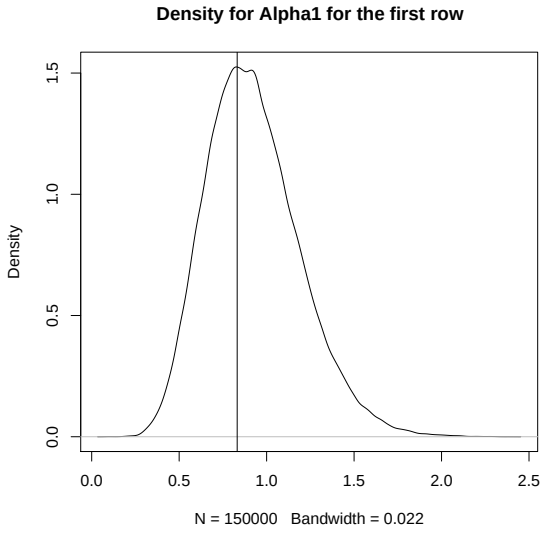

Figure C.11: Density for α_3 for the first row

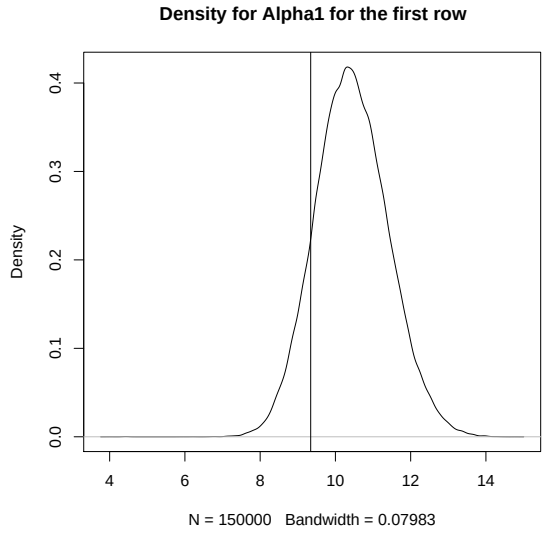

Figure C.12: Density for α_4 for the first row

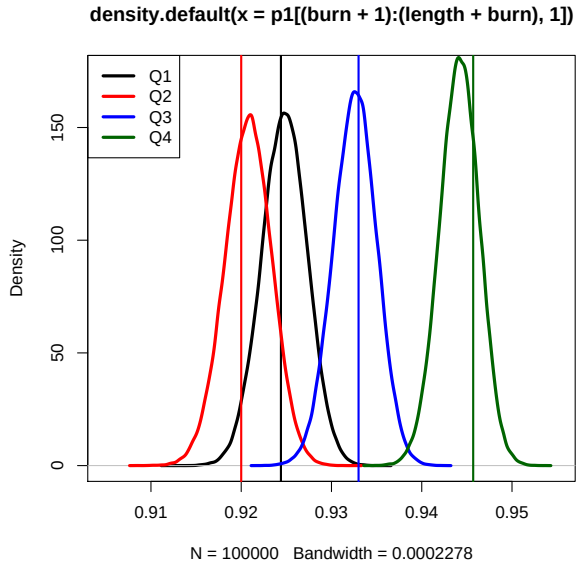

Figure C.13: Densities for probabilities in the first row, first column (current to current) of the transition matrix

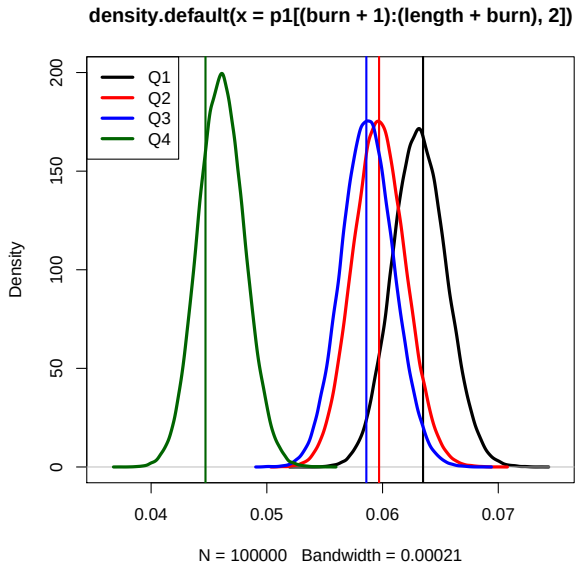

Figure C.14: Density for probabilities in the first row, second column of the transition matrix

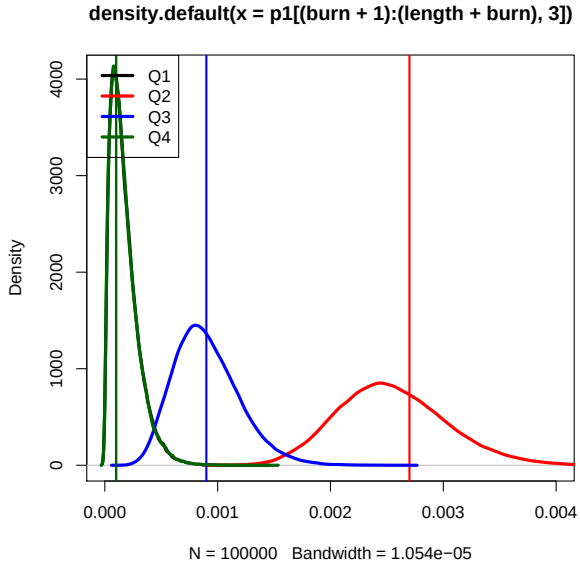

Figure C.15: Density for probabilities in the first row, third column of the transition matrix

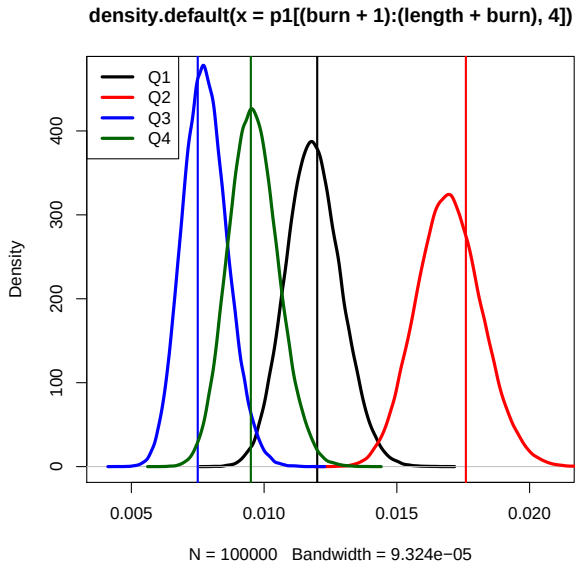

Figure C.16: Density for probabilities in the first row, fourth column of the transition matrix

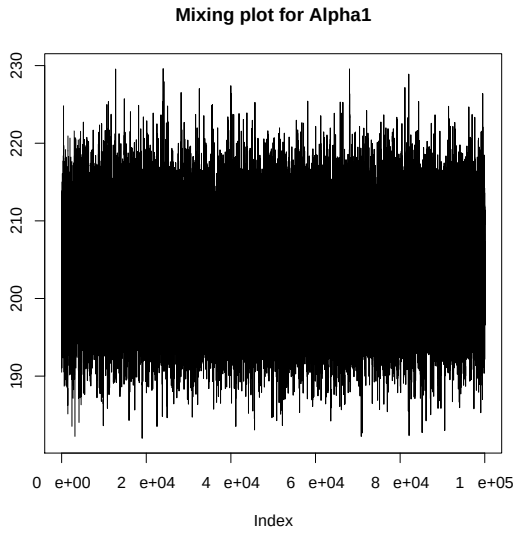

Figure C.17: Mixing plot for α_1 for the second row

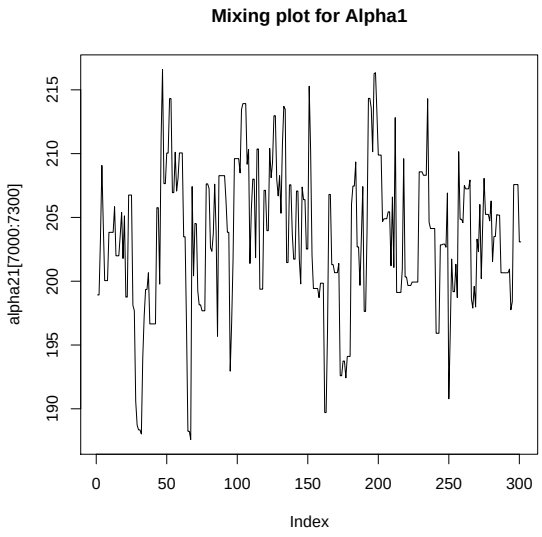

Figure C.18: Mixing plot for 300 realizations of α_1 for the second row

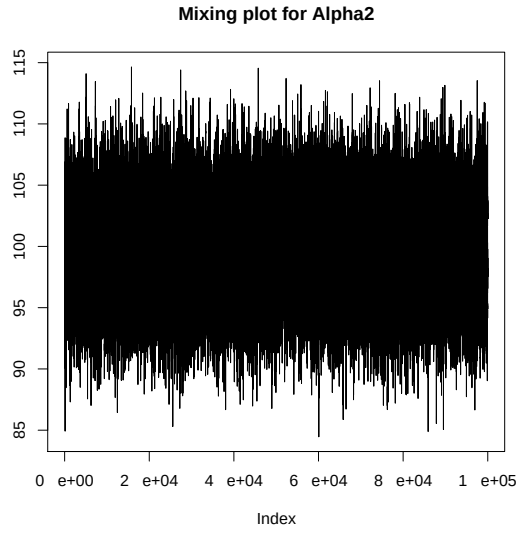

Figure C.19: Mixing plot for α_2 for the second row

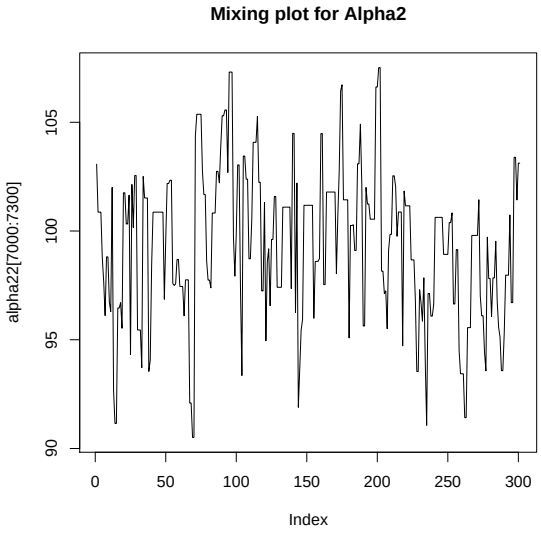

Figure C.20: Mixing plot for 300 realizations of α_2 for the second row

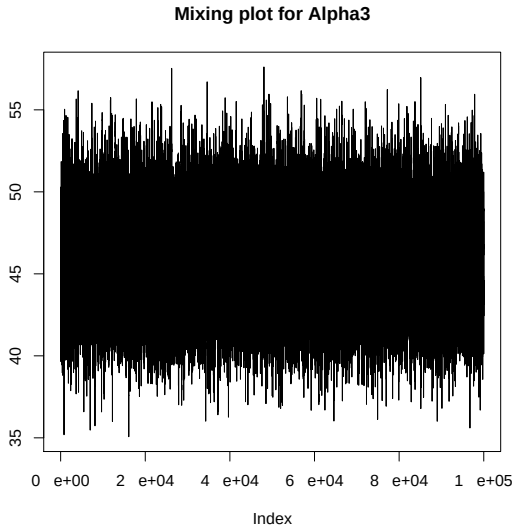

Figure C.21: Mixing plot for α_3 for the second row

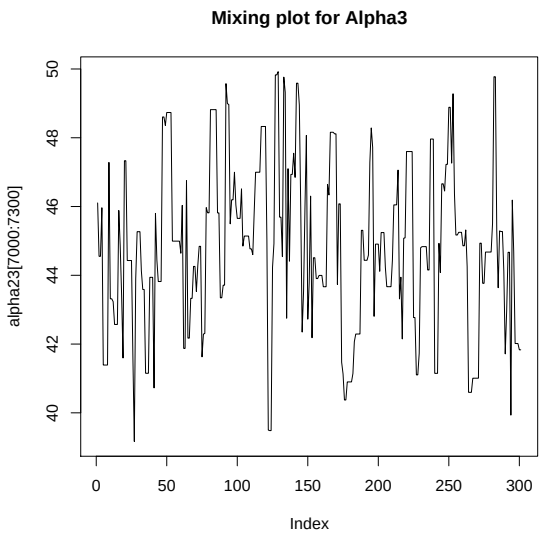

Figure C.22: Mixing plot for 300 realizations of α_3 for the second row

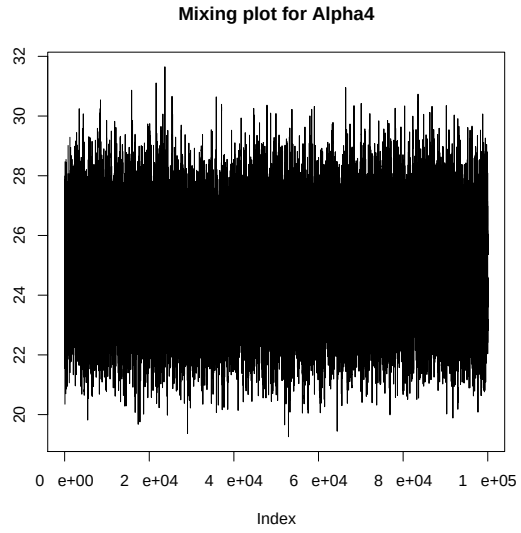

Figure C.23: Mixing plot for α_4 for the second row

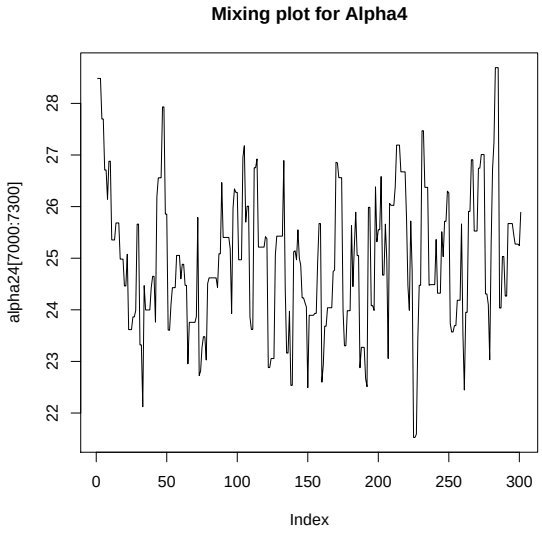

Figure C.24: Mixing plot for 300 realizations of α_4 for the second row

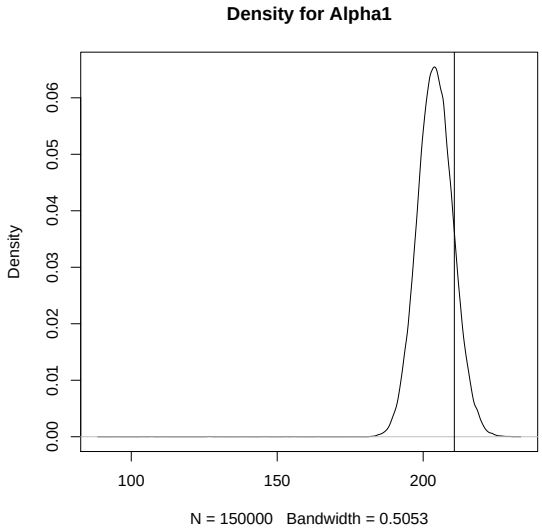

Figure C.25: Density for α_1 for the second row

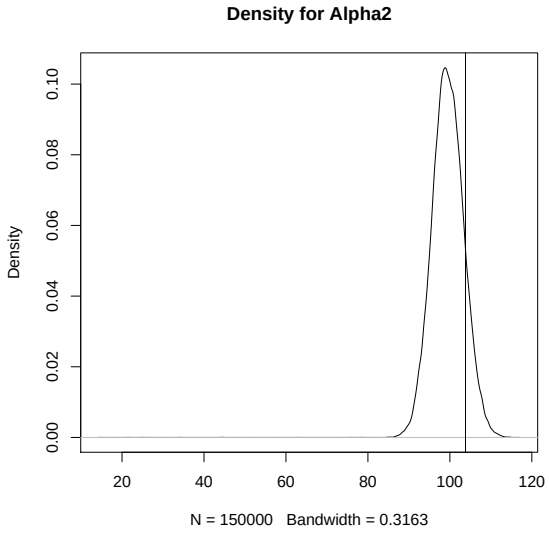

Figure C.26: Density for α_2 for the second row

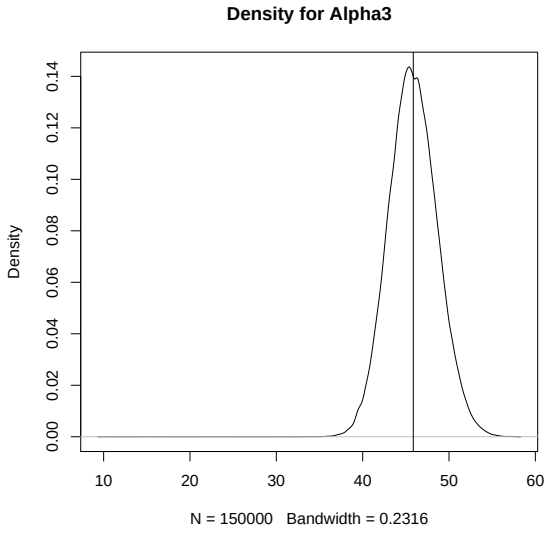

Figure C.27: Density for α_3 for the second row

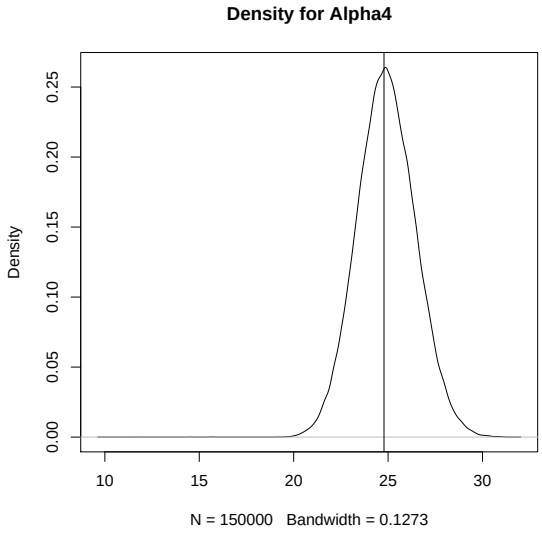

Figure C.28: Density for α_4 for the second row

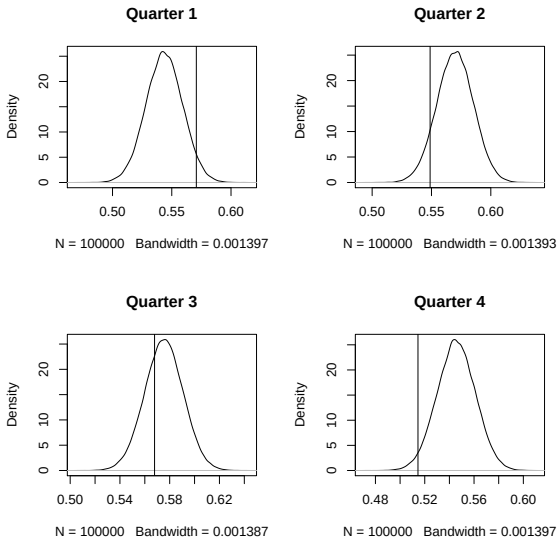

Figure C.29: Density of probabilities associated with α_1 for the second row

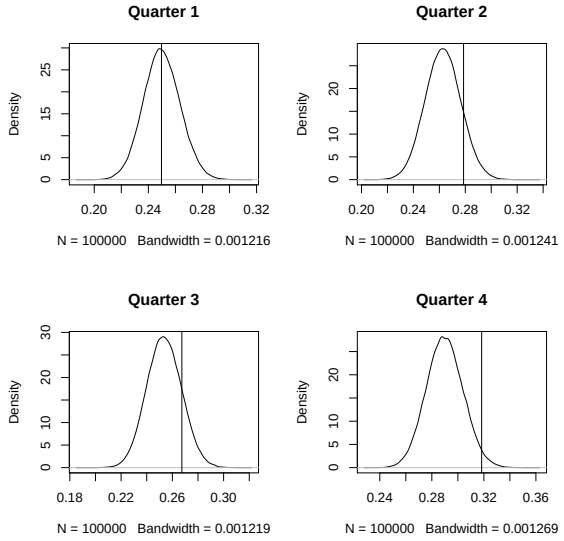

Figure C.30: Density of probabilities associated with α_2 for the second row

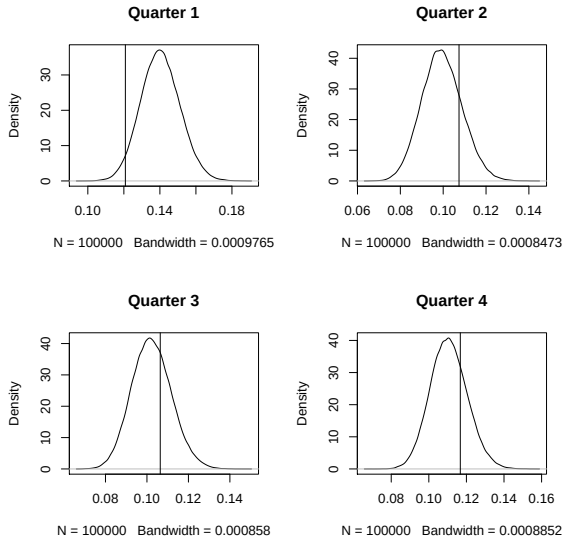

Figure C.31: Density of probabilities associated with α_3 for the second row

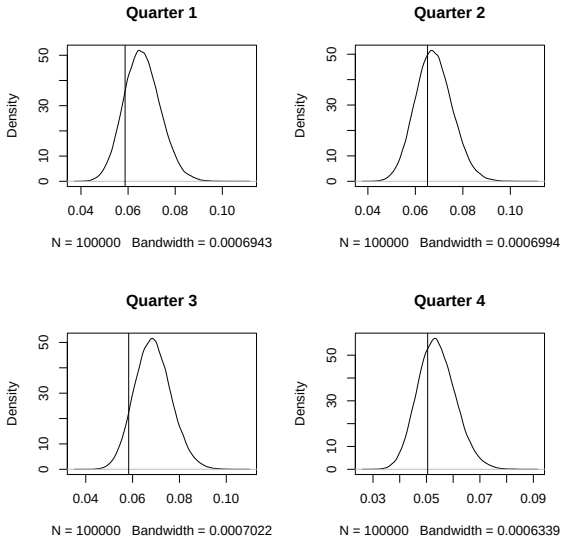

Figure C.32: Density of probabilities associated with α_4 for the second row

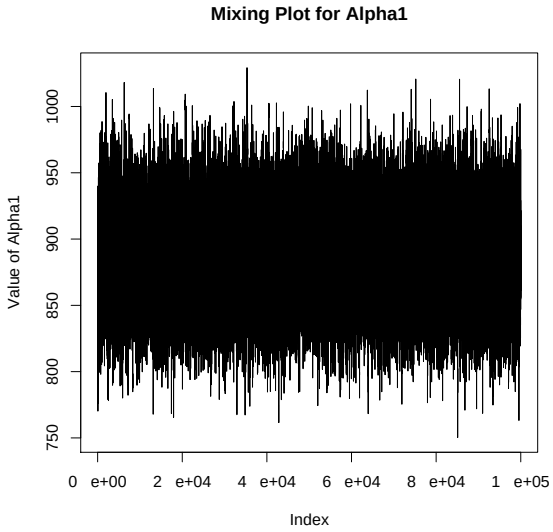

Figure D.1: Mixing plot for α_1 for the first row

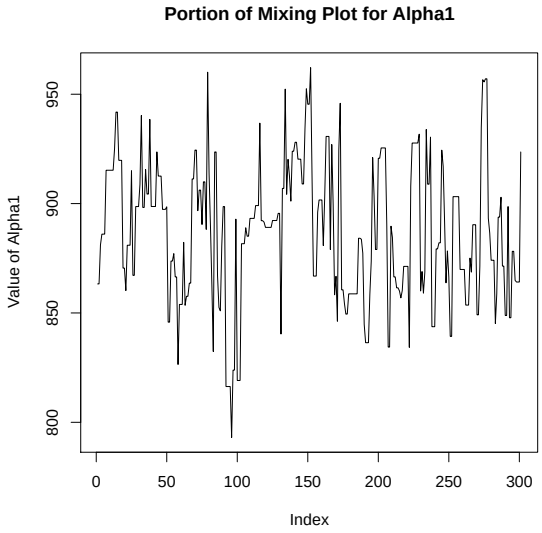

Figure D.2: Closer examination of the mixing plot for α_1 for the first row

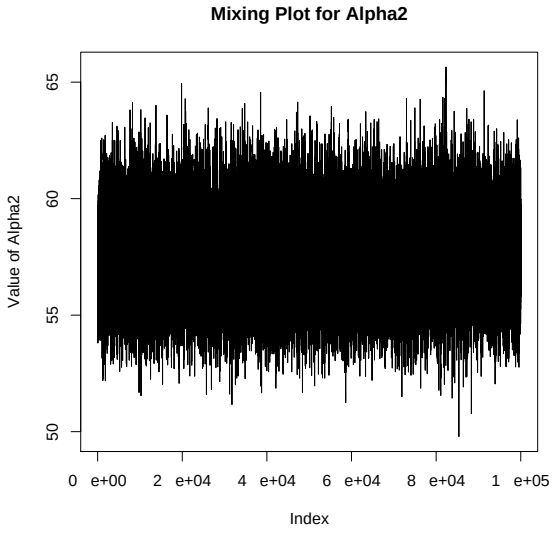

Figure D.3: Mixing plot for α_2 for the first row

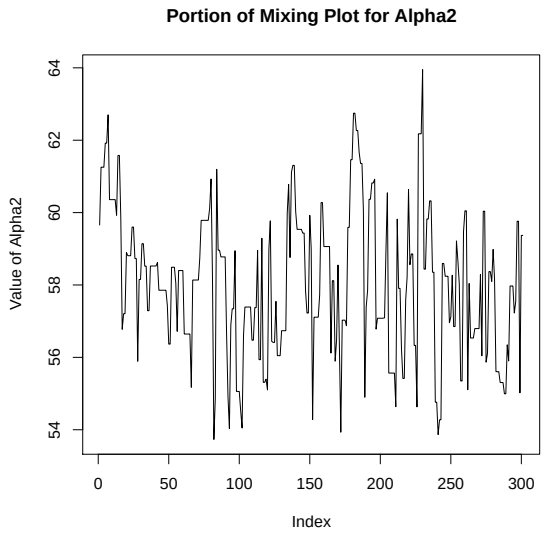

Figure D.4: Closer examination of the mixing plot for α_2 for the first row

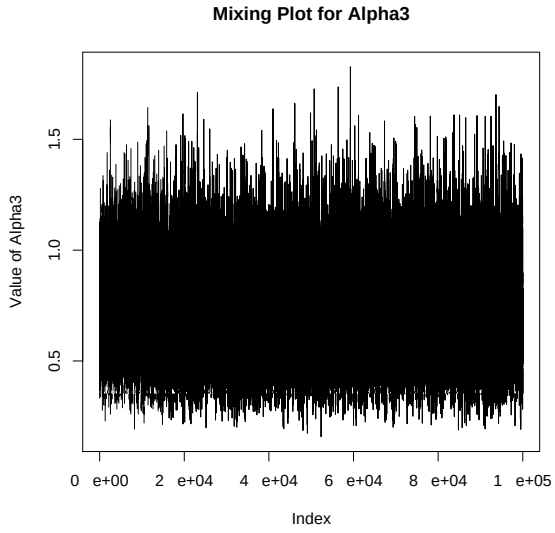

Figure D.5: Mixing plot for α_3 for the first row

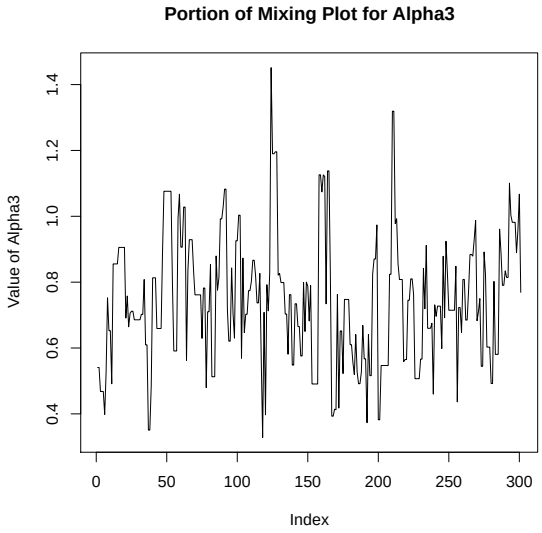

Figure D.6: Closer examination of the mixing plot for α_3 for the first row

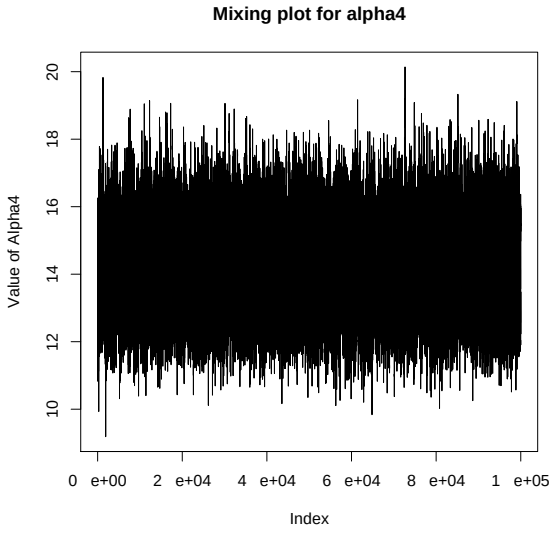

Figure D.7: Mixing plot for α_4 for the first row

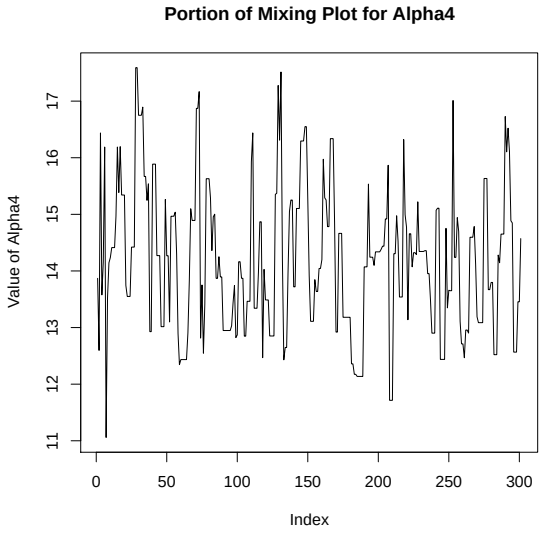

Figure D.8: Closer examination of the mixing plot for α_4 for the first row

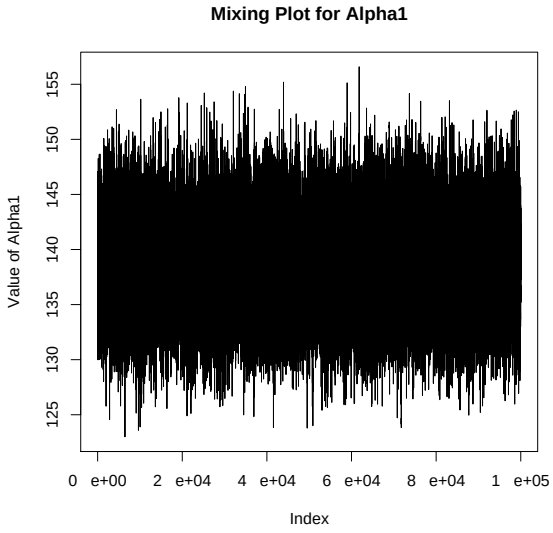

Figure D.9: Mixing plot for α_1 for the second row of real data

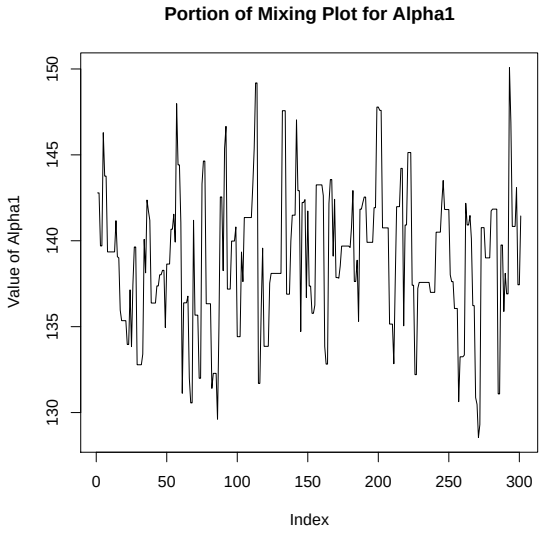

Figure D.10: Closer examination of the mixing plot for α_1 for the second row

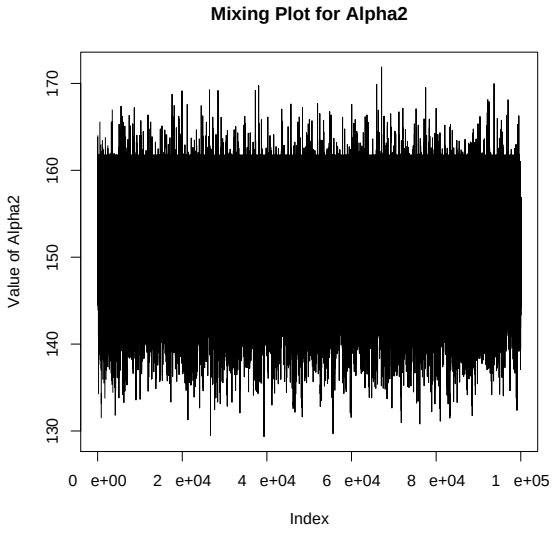

Figure D.11: Mixing plot for α_2 for the second row

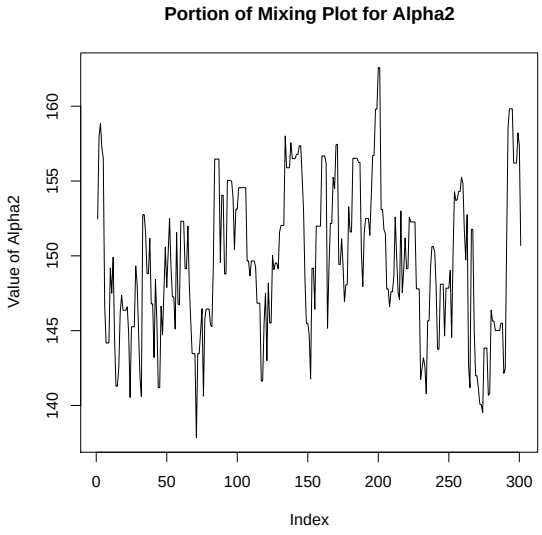

Figure D.12: Closer examination of the mixing plot for α_2 for the second row

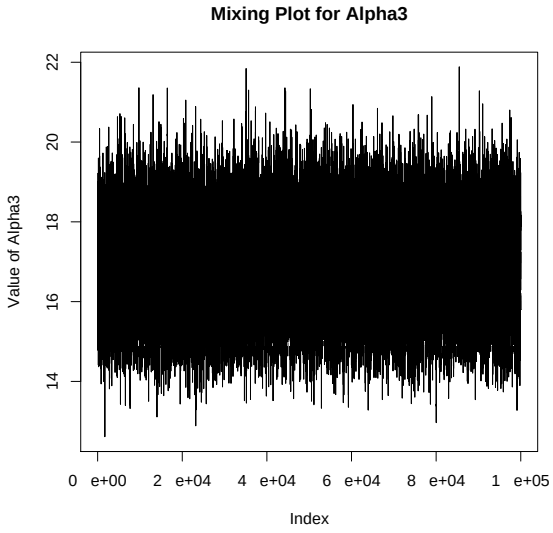

Figure D.13: Mixing plot for α_3 for the second row

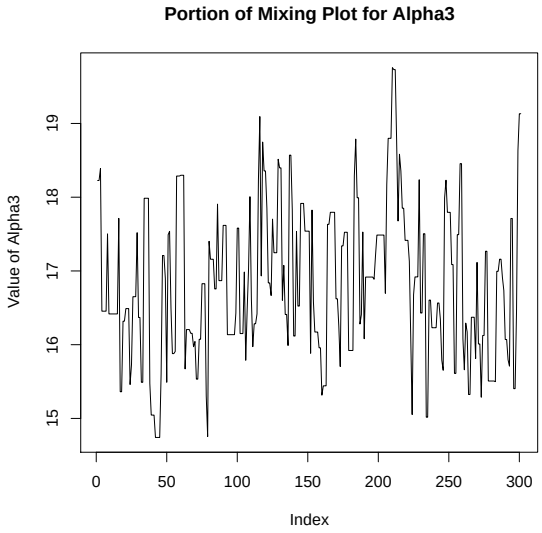

Figure D.14: Closer examination of the mixing plot for α_3 for the second row

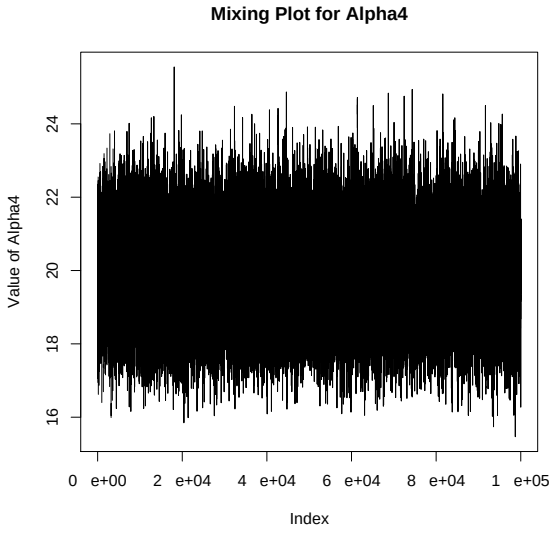

Figure D.15: Mixing plot for α_4 for the second row

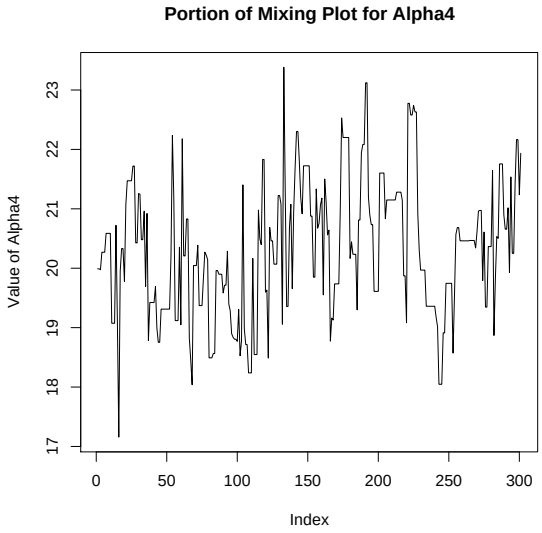

Figure D.16: Closer examination of the mixing plot for α_4 for the second row

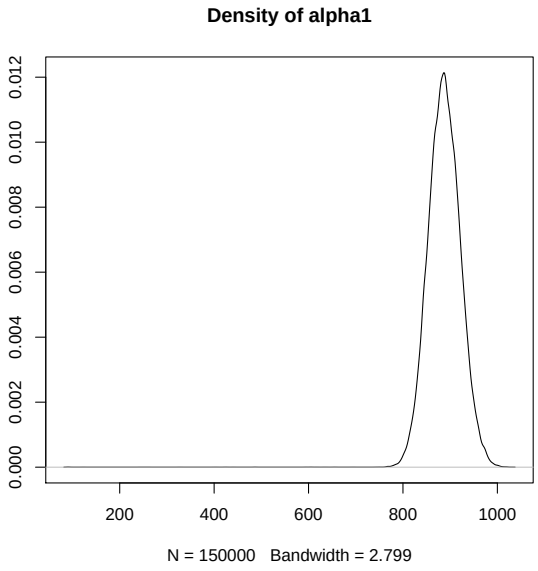

Figure D.17: Density for α_1 for the first row

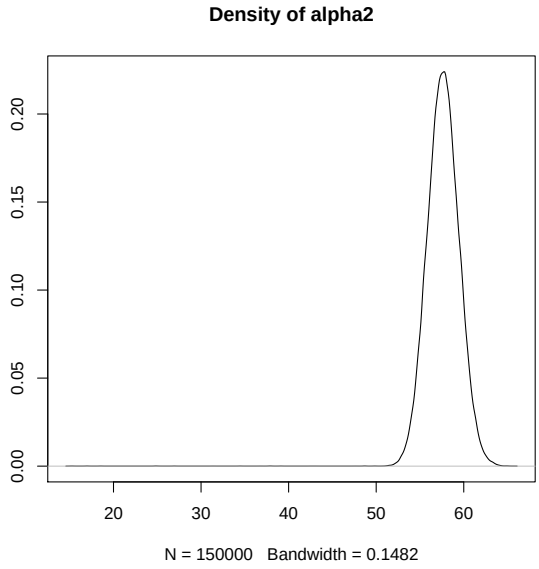

Figure D.18: Density for α_2 for the first row

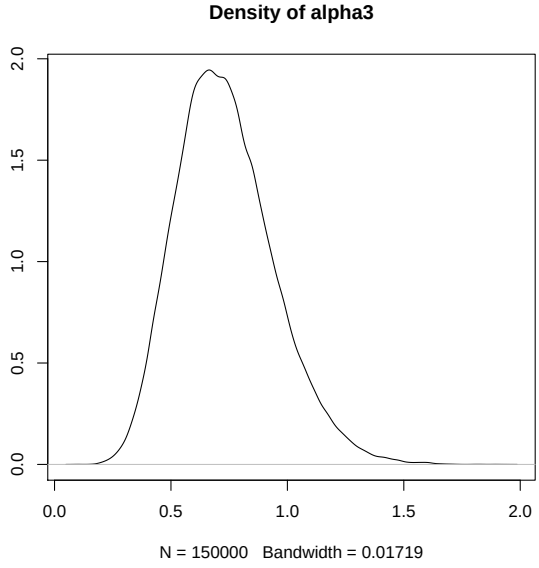

Figure D.19: Density for α_3 for the first row

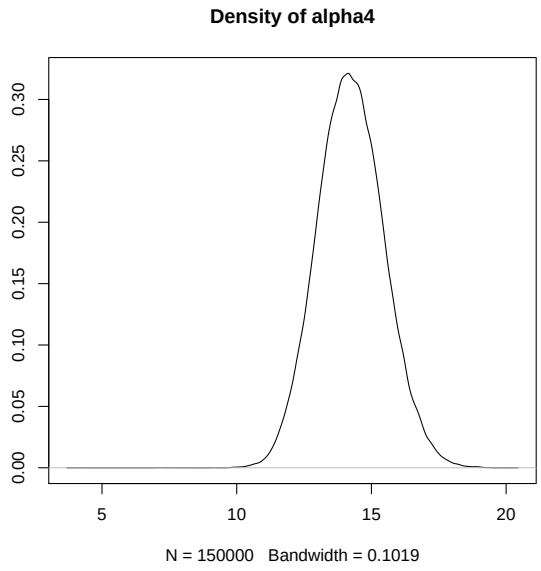

Figure D.20: Density for α_4 for the first row

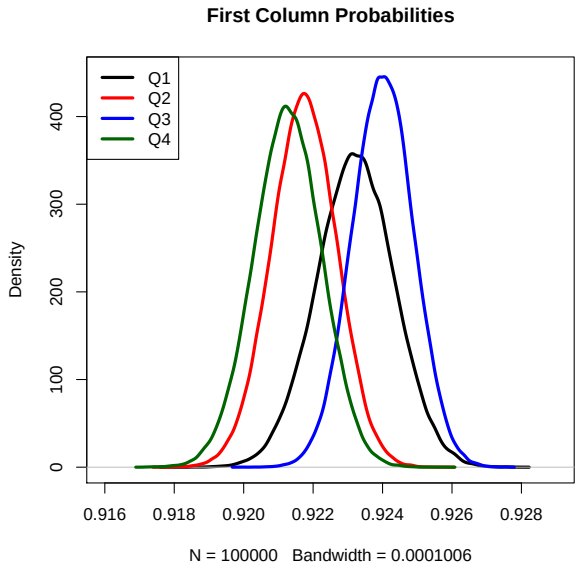

Figure D.21: Density for probabilities by quarter associated with α_1 for the first row

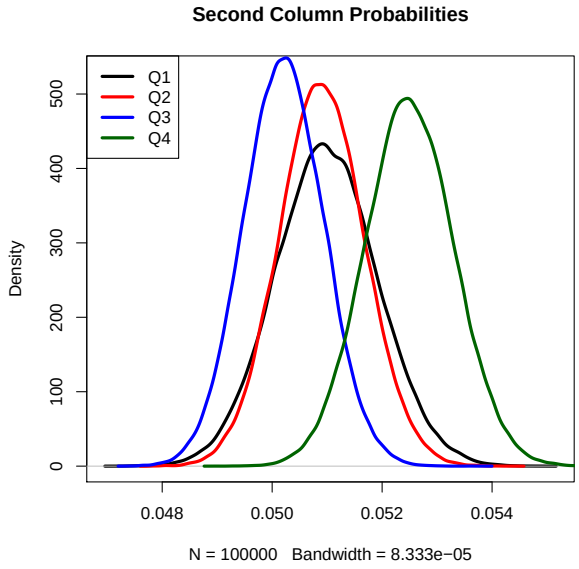

Figure D.22: Density for probabilities by quarter associated with α_2 for the first row

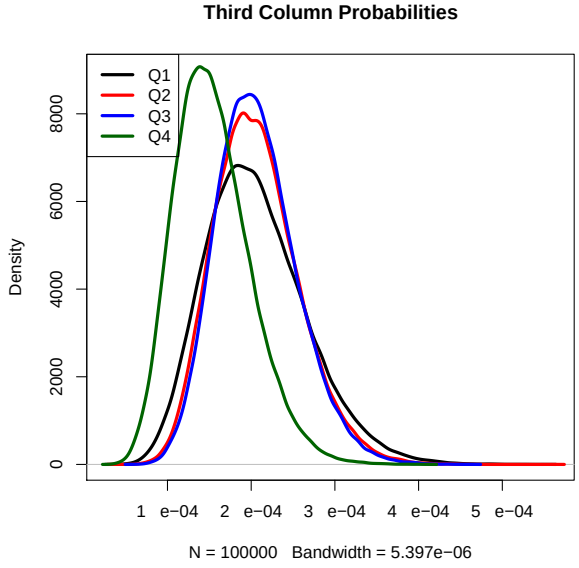

Figure D.23: Density for probabilities by quarter associated with α_3 for the first row

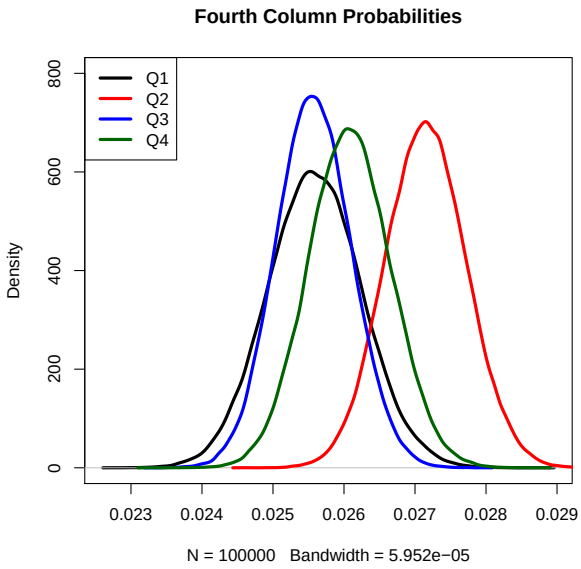

Figure D.24: Density for probabilities by quarter associated with α_4 for the first row

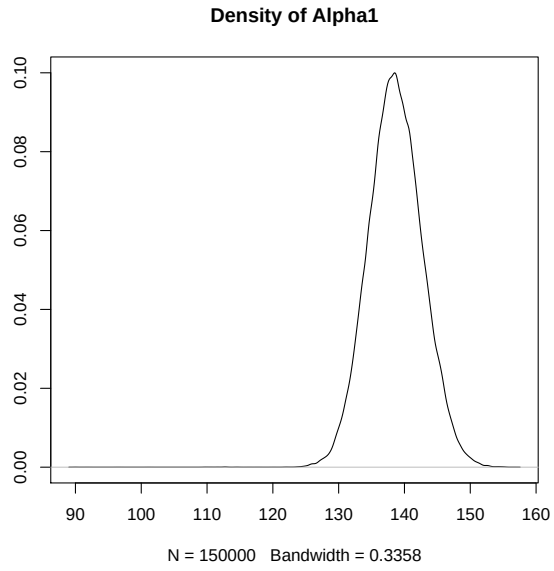

Figure D.25: Density for α_1 for the second row

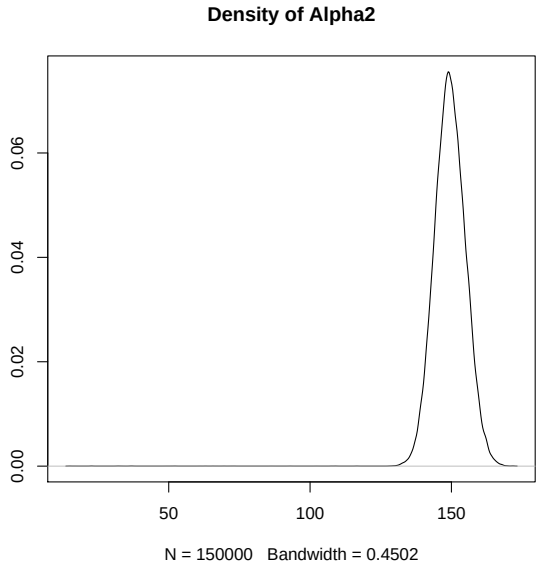

Figure D.26: Density for α_2 for the second row

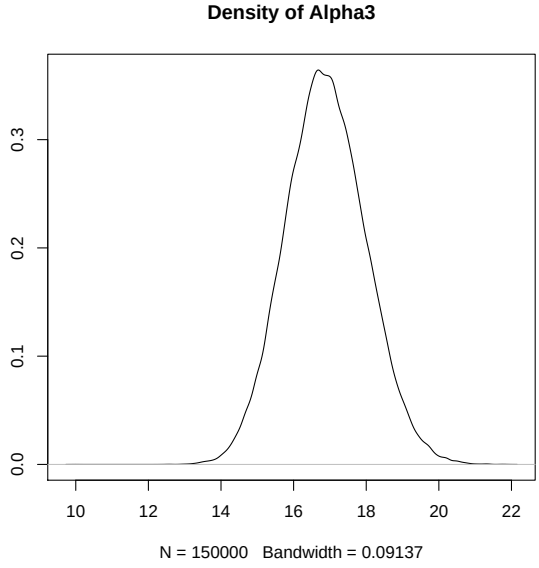

Figure D.27: Density for α_3 for the second row

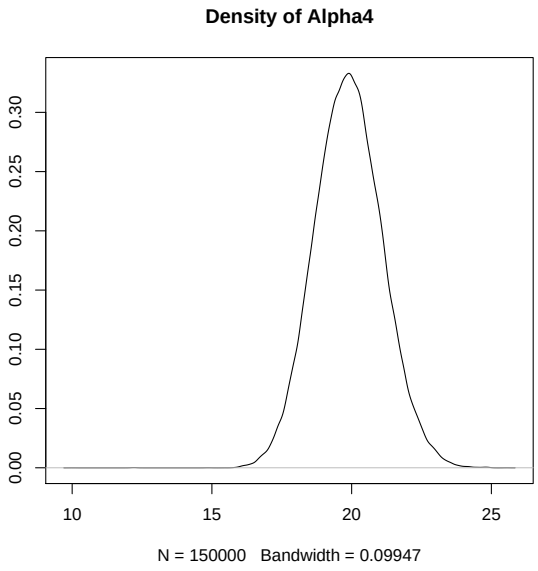

Figure D.28: Density for α_4 for the second row

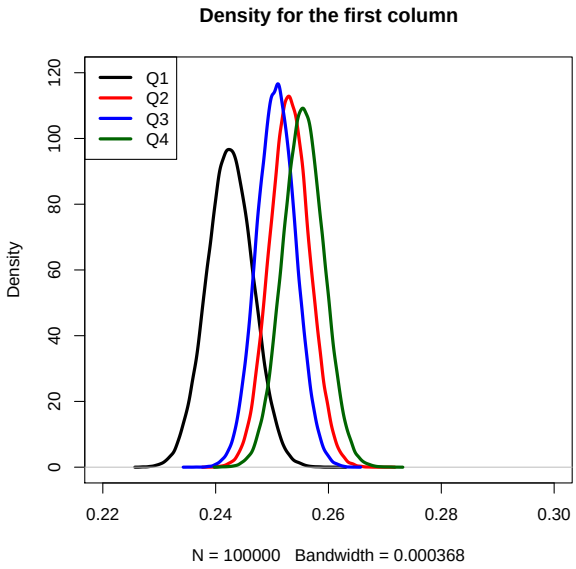

Figure D.29: Density for probabilities by quarter associated with α_1 for the second row

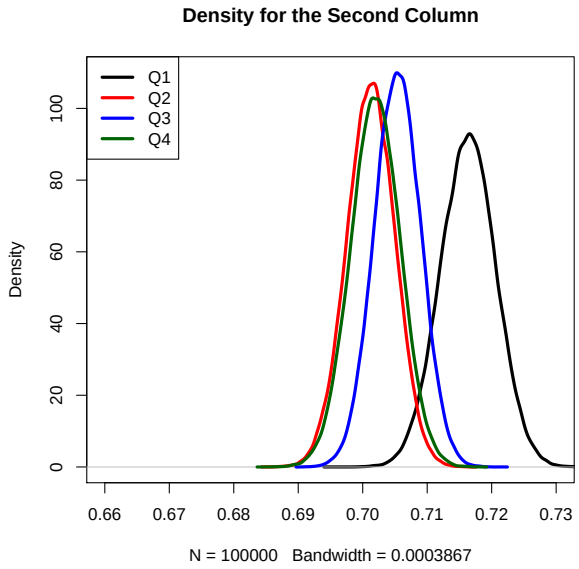

Figure D.30: Density for probabilities by quarter associated with α_2 for the second row

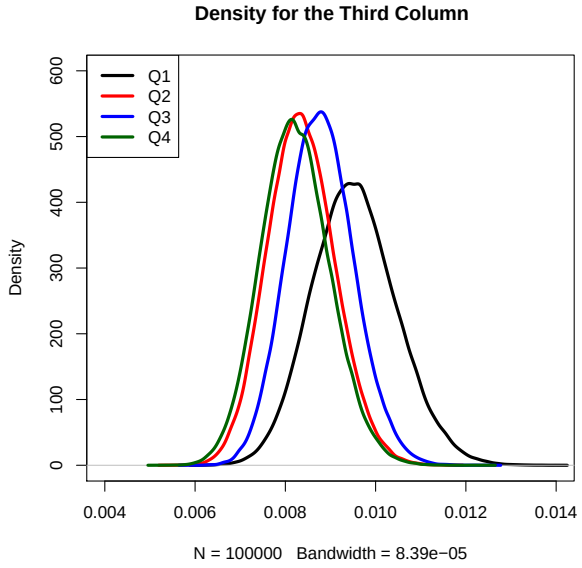

Figure D.31: Density for probabilities by quarter associated with α_3 for the second row

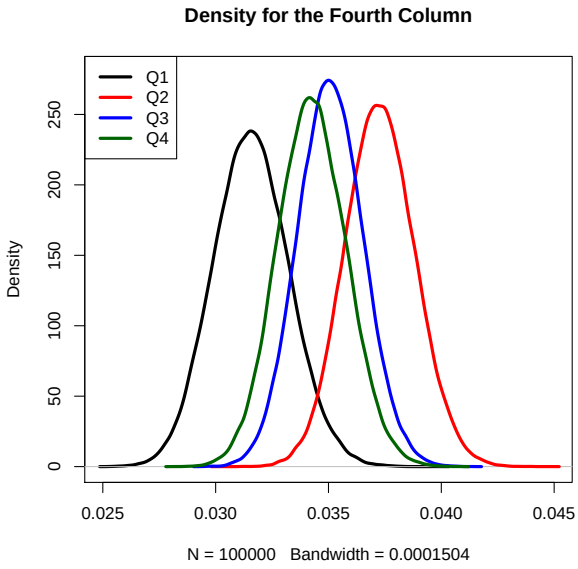

Figure D.32: Density for probabilities by quarter associated with α_4 for the second row