Analysis of Alumni-Giving Behavior With MCMC Method

YUSI LI

Department of Statistics

Iowa State University

June 15, 2021

Abstract

Alumni giving has become a main source of income for many colleges and universities in the United States. In this paper, we impose a dynamic linear model to predict the alumni-giving behavior by capturing the dynamic of the university-alumni interactions and use MCMC method to do estimation. In the simulation studies, we demonstrate that MCMC method can provide accurate estimation for model parameters and latent dynamic linear coefficient. We apply the MCMC method to a real alumni giving data between 2009 and 2016 and show that our model can have high accuracy of prediction using the data without missing values.
Keywords— Alumni, giving, MCMC, Prediction

1 Introduction

Alumni giving has become a main source of income for many colleges and universities in the United States. It is essential to a university's operation funding and can be a decisive factor in the success of colleges and universities. Predictions of alumni-giving behavior is essential in the discovery of prospective donors and thus enable colleges and universities to share their fundraising efforts on the alumni who have a high likelihood to donate to schools.

Researchers have developed models to predict alumni-giving behavior, such as the model of Lindahl [1], the model of Netzer [3] and the model of Sun [2]. However, most of the models are based on static variables, such as demographic variable. They did not consider the dynamics of the effects of interactions between university and alumni. A few models consider the dynamics of the effects of interactions between university and alumni. These models assume an alumni state variable which represent the inner willingness of the alumni to donate and the alumni state will determine by the interactions between university and alumni. However, the alumni states in these models are discrete and limited to some fix values. Specifically, these models may also ignore the potential time dependence in the model parameters.

In this paper, we suggest a dynamic linear model to predict the alumni-giving behavior by capturing the dynamic of the university-alumni interactions and time dependence in the model parameters.

Colleges and universities often engage in activities which have on a short or long period impact on alumni states. These activities aim to the shift the alumni into a different state in which the alumni are more likely to donate. In our model, we assume the alumni states are continuous. The change in alumni states is determined by the proportion of the donor's attended activities. A logit function is used to describes the relationship between the likelihood of the alumni-giving behavior and the alumni states.

In the simulation studies, we examined the accuracy of parameter and latent coefficients estimates to ensure that our MCMC method can indeed identify model parameters and latent coefficients using the simulated data. In the empirical study, we show that our model are not sensitive to the initial values which we chose for the models. We predicted the future alumni-giving behavior using the simulated data and real data, respectively, and showed our model can have high accuracy of prediction using the data without missing values.

The rest of this paper is organized as follows. Section 2 we introduce model, the model parameters, latent coefficients and the prior distribution. Section 3 gives a brief overview to the MCMC method and show how the MCMC method will work in our model. Section 4 show the joint distribution and the posterior distribution of model parameters and latent coefficients. Section 5 presents simulation evidence of the performance of our MCMC method in estimating the model parameters and latent coefficients. Section 6 show the estimation result of the model parameters and latent coefficients using real data. Section 6 discuss the prediction accuracy for our model. Section 7 we concludes the paper with some final remarks.

2 Model

In our model:

$$Y_{it} = \begin{cases} 1, & \text{if the donor } i \text{ donated at year } t \\ 0, & \text{otherwise} \end{cases}$$

Y_{it} can be model as:

$$Y_{it} \sim Bernoulli(\pi_{it}),$$

where π_{it} is the probability the donor i donated at time t. The probability π_{it} is modeled with a logit function:

$$\text{logit}(\pi_{it}) = \beta_{it},$$

or equivalently,

$$\pi_{it} = \frac{\exp(\beta_{it})}{1 + \exp(\beta_{it})}.$$

The alumni state variable β_{it} describes the inner willingness of alumni to donation. In our model, we assume β_{it} is continuous instead of taking a few discrete values. β_{it} has a positive relationship with the probability of the alumni i to donate in the year t. A higher β_{it} means the donor i has a higher probability to donate in the year t while a lower β_{it} means a lower probability to donate in the year t. β_{it} can be further modeled as:

$$\beta_{it} = \beta_0 + \gamma_t * Z_{i(t-1)},$$

where β_0 is the intercept, which describes the state when the alumni did not attend any activities in the university in the year $t-1$. We can see that β_0 will determine the probability to donate π_{it} if the the alumni t do not attend any activities in the year $t-1$. Below we describe how to construct covariate $Z_{i(t-1)}$.

X_{it} is the number of activities the donor i attended in the year t. If we directly use the X_{it} in our model, the π_{it} will be close to 1 since β_{it} will be a very large number. In order to avoid the problem, we standardize X_{it} by dividing it by the number of total activities the university hold in the year t to get Z_{it} which is the proportion of the activities the donor i attended in the year t.

We use $Z_{i(t-1)}$ rather than Z_{it} in our model for prediction purpose. If we use Z_{it} in our model, we need to know the value of $Z_{i(t+1)}$ when we predict alumni-giving behavior at $t+1$. However, the alumni-giving behavior in $t+1$ would have been known if we know the $Z_{i(t+1)}$. In this case, our model will lost its role in predicting the future alumni-giving behavior.

γ_t is the coefficient of $Z_{i(t-1)}$, which represents the effect of the proportion of the donor i's attended activities in year $t-1$. We assume the γ_t is only specific to year t and depend on the γ_{t-1}:

$$\gamma_t = \gamma_{t-1} + \epsilon_r,$$

where the ϵ_r follows a normal distribution. The reason we propose a random walk to model γ_t is because the effect of donors' activity participation should be time-varying but change slowly over time.

$$\epsilon_r \sim Normal(0, \sigma_r^2),$$

We consider the following prior distribution :

$$\beta_0 \sim Uniform(-1, 1),$$

$$\sigma_r^2 \sim IG(2, 1),$$

where IG is the Inverse Gamma Distribution.

3 Introduction to MCMC method

In the section, we give a brief introduction to the MCMC method and show how the MCMC method will be use in our model.

A Markov chain is a stochastic model which experiences transitions from one state to another according to transition probability. The states in Markov chain is either discrete or continuous. The current state in a Markov chain only depend on the recent previous state:

$$P(X_t|X_{t-1}, ...X_0) = P(X_t|X_{t-1}).$$

(MCMC) Markov chain Monte Carlo method is a set of of algorithms which sample from a probability distribution through constructing a Markov Chain. In our model, let $\{\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T\}$

be the vector of all unknown parameters and latent coefficients. $Y = \{y_{it}\}_{i=1,...N}^{t=1,...T}$ and $Z = \{z_{it}\}_{i=1,...N}^{t=1,...T}$ are the observed information. T is the total number of years we considered in the model and is 8 in this case since we considered data from year 2009 to year 2016 when fitting the model, while N is the total number of alumni in the real data, which is 3699. The purpose of the MCMC method is to estimate the model parameters and latent coefficients for the observed information. In other word, we need to estimate the joint posterior distributions $P(\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T|Y)$.

MCMC provides us a conditional simulation approach which can generate random samples from the target distribution. By MCMC, we can sample from the distribution of $\{\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T\}$ given on the observed data $Y = \{y_{it}\}_{i=1,...N}^{t=1,...T}$ by generating a Markov Chain which has desire distribution over $\{\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T\}$.

Based on Cliford-Hammersley Theorem, most of the MCMC method share the idea that the complete conditional distributions can characterize the joint distribution. In our model, it means that the completed conditional distribution $P(\sigma_r^2|\beta_0, \gamma_1, ...\gamma_T, Y)$, $P(\beta_0|\gamma_1, ...\gamma_T, \sigma_r^2, Y)$, $P(\gamma_1|\beta_0, \sigma_r^2, \gamma_2, ...\gamma_T, Y), ..., P(\gamma_t|\beta_0, \sigma_r^2, \gamma_1, ...\gamma_{T-1}, Y)$ completely characterize the joint distribution $P(\beta_0, \sigma_r^2, , \gamma_1, ...\gamma_T|Y)$. In other words, the key point of MCMC method is that it usually easier to sample from the complete conditional distributions $P(\sigma_r^2|\beta_0, \gamma_1, ...\gamma_T, Y)$, $P(\beta_0|\gamma_1, ...\gamma_T, \sigma_r^2, Y)$, $P(\gamma_1|\beta_0, \sigma_r^2, \gamma_2, ...\gamma_T, Y), ..., P(\gamma_t|\beta_0, \sigma_r^2, \gamma_1, ...\gamma_{T-1}, Y)$ than directly sampling from the joint distribution $P(\beta_0, \gamma_1, ...\gamma_T, \sigma_r^2|Y)$.

In MCMC, we need to get a close form for the complete conditional distributions and then sample from it. However, if the close form is not available for the complete conditional distributions, Metropolis-Hastings Algorithm will be used. In the Metropolis-Hasting Algorithm, we will draw a candidate from the proposed density and accept or reject the candidate on the acceptance criteria.

Metropolis-Hastings Algorithm begin with initializing the sample values. These value are usually sampled from the prior distribution. In the loop of the algorithm, the first step is to generate a candidate sample X^{cand} from the proposed distribution $q(X^k|X^{k-1})$. The second step is to compute the acceptance probability via the acceptance function $\alpha(X^k|X^{k-1})$ which base on the proposal distribution and the full joint density $\pi(.)$. In the end, we accept the candidate sample with probability α, the acceptance probability, or reject it with probability $1 - \alpha$.

In the iteration k=1,2,3....

- Propose: $X^{Cand} \sim q(X^k|X^{k-1})$
- Acceptance Probability: $\alpha(X^{Cand}|X^{k-1}) = min\left(1, \frac{q(X^{k-1}|X^{Cand})\cdot\pi(X^{Cand})}{q(X^{Cand}|X^{k-1})\cdot\pi(X^{k-1})}\right)$
- We draw $\mu \sim Uniform(0, 1)$
- If $\mu < \alpha$, then we accept the proposal: $X^k = X^{Cand}$
- If $\mu > \alpha$, then we reject the proposal: $X^k = X^{k-1}$

Specifically, if the proposals distribution is symmetric, such as normal distribution, the acceptance function can be simplified as:

$$\alpha(X^{Cand}|X^{k-1}) = min\left(1, \frac{\pi(X^{Cand})}{\pi(X^{k-1})}\right)$$

The sample $\{\beta_0^{(k)}, \sigma_r^{2(k)}, \gamma_1^{(k)}, ...\gamma_t^{(k)}\}_t^T$ can be used to estimate the model parameters via Monte Carlo methods. The estimate of the $\{\beta_0^{(k)}, \sigma_r^{2(k)}, \gamma_1^{(k)}, ...\gamma_t^{(k)}\}$ is the posterior mean of $P(\beta_0^{(k)}, \sigma_r^{2(k)}, \gamma_1^{(k)}, ...\gamma_t^{(k)}|Y)$.

In our model, the joint distribution density $P(\beta_0, \gamma_1, ...\gamma_T, \sigma_r^2)$ can be written in the form which is proportional to the likelihood function, joint distribution of $\gamma_1, ...\gamma_T$, and the prior distribution of model parameters.

$$P(\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T|Y) \propto P(Y|\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T) \cdot P(\gamma_1, ...\gamma_T|\beta_0, \sigma_r^2) \cdot \pi(\beta_0) \cdot \pi(\sigma_r^2).$$

We can see that the latent coefficients $\gamma_1, ...\gamma_T$ do not depend on parameter β_0, so the joint distribution can be further written as:

$$P(\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T|Y) \propto P(Y|\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T) \cdot P(\gamma_1, ...\gamma_T|\sigma_r^2) \cdot \pi(\beta_0) \cdot \pi(\sigma_r^2).$$

$\{\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T\}$ is the vector of all parameters of our model. $Y = \{y_{it}\}_{i=1,...N}^{t=1,...T}$ is a vector of the observed data, respectively. Below we give an iterative procedure to draw posterior samples:

- Set the initial values for $\beta_0^{(0)}, \sigma_r^{2(0)}, \gamma_1^{(0)}, ...\gamma_T^{(0)}$ and $Y = \{y_{it}\}_{i=1,...N}^{t=1,...T}$
- we can draw $\sigma_r^{2(1)} \sim P(\sigma_r^2|\beta_0^{(0)}, \gamma_1^{(0)}, ...\gamma_T^{(0)}, Y)$
- then $\beta_0^{(1)} \sim P(\beta_0|\sigma_r^{2(1)}, \gamma_1^{(0)}, ...\gamma_T^{(0)}, Y)$

- $\gamma_1^{(1)} \sim P(\gamma_1|\beta_0^{(1)}, \sigma_r^{2(1)}, \gamma_2^{(0)}, ...\gamma_T^{(0)}, Y)$

- ...

- $\gamma_T^{(1)} \sim P(\gamma_T|\beta_0^{(1)}, \sigma_r^{2(1)}, \gamma_1^{(1)}, ...\gamma_{T-1}^{(1)}, Y)$

We can continue in this fashion. In the kth iteration, the algorithm will generate a sequence of random variables$\{\beta_0^{(k)}, \sigma_r^{2(k)}, \gamma_1^{(k)}, ...\gamma_T^{(k)}\}_{k=1}^{M}$. This sequence forms a Markov Chain whose distribution converges to the target joint distribution $P(\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T|Y)$. Then we can obtain the model parameter estimate by averaging the posterior draw after burn-in.

4 Posterior Distribution

In this section, we show the joint distribution for $\{\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T\}$ and the posterior distribution for each model parameters and latent coefficients. We also show the Metropolis-Hasting algorithm will work when the close form are not available.

4.1 The joint distribution for $\{\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T\}$

With the specified priors, the joint distribution for $\{\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T\}$ can be written as

$$
\begin{aligned}
P(\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T|Y) &\propto P(Y|\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T) \cdot P(\gamma_1, ...\gamma_T|\beta_0, \sigma_r^2) \cdot \pi(\beta_0) \cdot \pi(\sigma_r^2)) \\
&\propto P(Y|\beta_0, \sigma_r^2, \gamma) \cdot P(\gamma_1, ...\gamma_T|\sigma_r^2) \cdot \pi(\beta_0) \cdot \pi(\sigma_r^2) \\
&\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \cdot \pi(\beta_0) \cdot \pi(\sigma_r^2) \\
&\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} (1-\pi_{it})^{1-y_{it}} \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2\sigma_r^2}\right) (\sigma_r^2)^{-3} \exp\left(-\frac{1}{\sigma_r^2}\right)
\end{aligned}
$$

4.2 The posterior distribution for σ_r^2

The posterior distribution of σ_r^2 conditioned on $\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T$ and Y is

$$\begin{aligned}
P(\sigma_r^2|\beta_0, \gamma_1, ...\gamma_T, Y) &\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \cdot \pi(\beta_0) \cdot \pi(\sigma_r^2) \\
&\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} (1-\pi_{it})^{1-y_{it}} \prod_{t=2}^T \frac{1}{\sqrt{2\pi}\sigma_r^2} \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2\sigma_r^2}\right) (\sigma_r^2)^{-3} \exp\left(-\frac{1}{\sigma_r^2}\right) \\
&\propto \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \cdot (\sigma_r^2)^{-3} \exp\left(-\frac{1}{\sigma_r^2}\right) \\
&\propto \left(\sigma_r^2\right)^{-\left(\frac{1}{2}(T-1)+2\right)-1} \cdot \exp\left(-\frac{\frac{1}{2}\sum_{t=2}^T (\gamma_t - \gamma_{t-1})^2 + 1}{\sigma_r^2}\right)
\end{aligned}$$

This is kernel of $IG\left(\frac{1}{2}(T-1)+2, \frac{1}{2}\sum_{t=2}^T (\gamma_t - \gamma_{t-1})^2 + 1\right)$. So, The posterior of σ_r^2 follows a Inverse Gamma distribution;

$$\sigma_r^2 \sim IG\left(\frac{1}{2}(T-1)+2, \frac{1}{2}\sum_{t=2}^T (\gamma_t - \gamma_{t-1})^2 + 1\right)$$

4.3 The posterior distribution for β_0

The posterior distribution of β_0 conditioned on $\sigma_r^2, \gamma_1, ...\gamma_T$ and Y is

$$\begin{aligned}
P(\beta_0|\sigma_r^2, \gamma_1, ...\gamma_T, Y) &\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \cdot \pi(\beta_0) \cdot \pi(\sigma_r^2) \\
&\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \\
&\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}}
\end{aligned}$$

As it is shown, there isn't any closed form for the conditional distribution $P(\beta_0|\sigma_r^2, \gamma_1, ...\gamma_T, Y)$, So the Metropolis-Hasting algorithm has to be used.

In the iteration $k = 1, 2, 3...$

- Propose: $\beta_0^{Cand} \sim Normal(\beta_0^{k-1}, 0.5)$
- Acceptance Probability: $\alpha(\beta_0^{Cand}|\beta_0^{k-1}) = min\left(1, \frac{\pi(\beta_0^{Cand})}{\pi(\beta_0^{k-1})}\right)$
- We draw $\mu \sim Uniform(0, 1)$

- If $\mu < \alpha$, then we accept the proposal: $\beta_0^{(k)} = \beta_0^{Cand}$
- If $\mu > \alpha$, then we reject the proposal: $\beta_0^{(k)} = \beta_0^{(k-1)}$

4.4 The posterior distribution for γ_t

The posterior distribution of γ_1 conditioned on $\beta_0, \sigma_r^2, \gamma_2, ...\gamma_T$ and Y is

$$\begin{aligned}
P(\gamma_1|\sigma_r^2, \beta_0, \gamma_2, ...\gamma_T, Y) &\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \cdot \pi(\beta_0) \cdot \pi(\sigma_r^2) \\
&\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \\
&\propto \prod_i^N \pi_{i1}^{y_{it}} \cdot (1-\pi_{i1})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \\
&\propto \prod_i^N \pi_{i1}^{y_{i1}} \cdot (1-\pi_{it})^{1-y_{i1}} \cdot \exp\left(-\frac{(\gamma_2 - \gamma_1)^2}{2 \cdot \sigma_r^2}\right) \\
&\propto \prod_i^N \pi_{i1}^{y_{i1}} \cdot (1-\pi_{i1})^{1-y_{i1}} \cdot \exp\left(-\frac{(\gamma_2 - \gamma_1)^2}{2 \cdot \sigma_r^2}\right)
\end{aligned}$$

The posterior distribution of γ_T conditioned on $\beta_0, \sigma_r^2, \gamma_2, ...\gamma_{T-1}$ and Y is

$$\begin{aligned}
P(\gamma_T|\sigma_r^2, \beta_0, \gamma_2, ...\gamma_{T-1}, Y) &\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \cdot \pi(\beta_0) \cdot \pi(\sigma_r^2) \\
&\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \\
&\propto \prod_i^N \pi_{iT}^{y_{iT}} \cdot (1-\pi_{iT})^{1-y_{iT}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \\
&\propto \prod_i^N \pi_{iT}^{y_{iT}} \cdot (1-\pi_{iT})^{1-y_{iT}} \cdot \exp\left(-\frac{(\gamma_T - \gamma_{T-1})^2}{2 \cdot \sigma_r^2}\right) \\
&\propto \prod_i^N \pi_{iT}^{y_{iT}} \cdot (1-\pi_{iT})^{1-y_{iT}} \cdot \exp\left(-\frac{(\gamma_T - \gamma_{T-1})^2}{2 \cdot \sigma_r^2}\right)
\end{aligned}$$

The posterior distribution for γ_1 and γ_T are proportional to the similar form. However, the posterior distribution for γ_t will be proportional to different form if $1 < t < T$.

The posterior distribution of γ_t conditioned on $\beta_0, \sigma_r^2, \gamma_1, ...\gamma_T$ and Y is

$$\begin{aligned}
P(\gamma_t|\sigma_r^2, \beta_0, \gamma_1, ...\gamma_T, Y) &\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \cdot \pi(\beta_0) \cdot \pi(\sigma_r^2) \\
&\propto \prod_i^N \prod_t^T \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \\
&\propto \prod_i^N \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \prod_{t=2}^T \cdot \frac{1}{\sqrt{2\pi\sigma_r^2}} \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \\
&\propto \prod_i^N \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \exp\left(-\frac{(\gamma_{t+1} - \gamma_t)^2}{2 \cdot \sigma_r^2}\right) \cdot \exp\left(-\frac{(\gamma_t - \gamma_{t-1})^2}{2 \cdot \sigma_r^2}\right) \\
&\propto \prod_i^N \pi_{it}^{y_{it}} \cdot (1-\pi_{it})^{1-y_{it}} \cdot \exp\left(-\frac{2\gamma_{t+1}^2 - 2 \cdot \gamma_t \cdot \gamma_{t+1} - 2 \cdot \gamma_t \cdot \gamma_{t-1}}{2 \cdot \sigma_r^2}\right)
\end{aligned}$$

As it is shown, there isn't also any closed form for the conditional distribution $P(\gamma_t|\beta_0, \sigma_r^2, \gamma_1, ...\gamma_t, Y)$, So the Metropolis-Hasting algorithm has to be used.
In the iteration $k = 1, 2, 3...$
In the iteration $t = 1, 2, 3, ...T$

- Propose: $\gamma_t^{Cand} \sim Normal(\gamma_t^{k-1}, 0.5)$
- Acceptance Probability: $\alpha(\gamma_t^k|\gamma_t^{k-1}) = min\left(1, \frac{\pi(\gamma_t^k)}{\pi(\gamma_t^{k-1})}\right)$
- We draw $\mu \sim Uniform(0, 1)$
- If $\mu < \alpha$, then we accept the proposal: $\gamma_t^{(k)} = \gamma_t^{Cand}$
- If $\mu > \alpha$, then we reject the proposal:$\gamma_t^{(k)} = \gamma_t^{(k-1)}$

5 Simulation

In this section, simulation studies are shown using the MCMC method. We examined the accuracy of parameters and latent coefficients estimates to ensure that our methods can indeed identify model parameters and latent coefficients using the simulated data. For our simulations we used the software R.

5.1 Generating simulated data

We set the true value to the parameters and latent coefficients:

- $\beta_0^{(true)} = 0.5$
- $\sigma_r^{2(true)} = 20$
- $\gamma_0^{(true)} = 4$

With $\beta_0^{(true)}$, $\sigma_r^{2(true)}$, and $\gamma_0^{(true)}$, we can generate the $\gamma_1^{(true)}$ to $\gamma_8^{(true)}$ by the below iterative procedure:

In the iteration $t = 1, 2, 3, ...T$

- We draw $\epsilon_{r(t)} \sim Normal(0, \sigma_r^{2(true)})$
- $\gamma_t^{(true)} = \gamma_{t-1}^{(true)} + \epsilon_{r(t)}$

Then, we got the true value for $\gamma_1^{(true)}$ to $\gamma_8^{(true)}$:

- $\gamma_1^{(true)} = 7.228863$
- $\gamma_2^{(true)} = 4.572165$
- $\gamma_3^{(true)} = 12.776484$
- $\gamma_4^{(true)} = 13.807305$
- $\gamma_5^{(true)} = 19.500778$
- $\gamma_6^{(true)} = 22.124091$
- $\gamma_7^{(true)} = 32.073988$
- $\gamma_8^{(true)} = 30.614541$

With the true value specified, we can generate the simulated data β_{it}^{sim}, π_{it}^{sim} and Y_{it}^{sim} by the below iterative procedure and we use the $Z_{i(t-1)}$ from the real data in this iterative procedure:

- In the iteration $i = 1, 2, 3, ...N$
  - In the iteration $t = 1, 2, 3, ...T$
    * $\beta_{it}^{sim} = \beta_0^{true} + \gamma_t^{(true)} \cdot Z_{i(t-1)}$

    * $\pi_{it}^{sim} = \frac{\exp(\pi_{it}^{sim})}{1+\exp(\pi_{it}^{sim})}$
    * We draw $Y_{it}^{sim} \sim Bernoulli(\pi_{it}^{sim})$

We generate the simulated data from 2009 to 2017. The simulated data from 2009 to 2016 will be used to estimate the model parameters and latent coefficients. The data in 2017 will be used for testing.

5.2 Parameters Estimation

The estimates are computed for the model parameters β_0, σ_r^2 and latent coefficients γ_1,...γ_8 using the simulated data from 2009 to 2016. In our estimation, we choose the following initial values of model parameters to be far from the corresponding true values:

- $\beta_0^{(0)} = 0.3$
- $\sigma_r^{2(0)} = 5$
- $\gamma_0^{(0)} = 1$

  With $\beta_0^{(0)}$, $\sigma_r^{2(0)}$, and $\gamma_0^{(0)}$, we can generate the $\gamma_1^{(0)}$ to $\gamma_8^{(0)}$ by the below iterative procedure:

  In the iteration $t = 1, 2, 3, ...T$

  - We draw $\epsilon_{r(t)} \sim Normal(0, \sigma_r^{2(0)})$
  - $\gamma_t^{(0)} = \gamma_{t-1}^{(0)} + \epsilon_{r(t)}$

  Then, we got the initial values for $\gamma_1^{(0)}$ to $\gamma_8^{(0)}$:

- $\gamma_1^{(0)} = 5.873067$
- $\gamma_2^{(0)} = 5.135799$
- $\gamma_3^{(0)} = 12.016019$
- $\gamma_4^{(0)} = 14.115418$
- $\gamma_5^{(0)} = 16.631101$
- $\gamma_6^{(0)} = 21.505064$
- $\gamma_7^{(0)} = 34.072877$

- $\gamma_8^{(0)} = 29.090717$

We run three Markov chains for each model parameters and latent coefficients to get the posterior distribution. Each chain has 10000 iterations and first 5000 iterations were burn-in.

Table 1 reports the posterior mean, the initial parameters, and the true values of the corresponding parameters and latent coefficients using simulated data. We can see in Table 1, that our MCMC method can accurately estimate the model parameters and latent coefficients. Even though we started our initial values at some values far away from the true values, the mean of parameters and latent coefficients estimates are very close to the true values we set in the simulation.

Figure 1 presents trace-plots of three chains for parameters β_0, σ_r^2 and latent coefficients $\gamma_1, ...\gamma_8$, respectively. Each trace-plot shows 5000 iterations after the burn-in period, and the red cross on the y-axis indicates the true value of the corresponding parameters and latent coefficients. We can see that the chains are converging stably and each of the model parameters and latent coefficients estimates are within a respective permissible range of the true values.

Figure 2 show the density plot of MCMC samples of parameter σ_r^2. The 95% credible interval is marked by the two blue vertical lines, which indicate the 2.5% and 97.5% quantiles, respectively. The true value of σ_r^2 is marked by the red vertical line and and true value of σ_r^2 is 20. As we can see, in the figure 2, the red line is in the 95% credible interval.

Figure 3 show the density plot of MCMC samples of parameter β_0. The 95% credible interval is marked by the two blue vertical lines, which indicate the 2.5% and 97.5% quantiles, respectively. The true value of β_0 is marked by the red vertical line and true value of β_0 is 0.5. As we can see, in the figure 3, the red line is in the 95% credible interval.

Figure 4 plot the 95% credible interval for the latent coefficients $\gamma_1, ...\gamma_8$. The year 2009 to 2016 on the x-axis are corresponding to γ_1 to γ_8. The two blue lines represent the 2.5% and 97.5% quantile line. The intervals between the 2.5% and 97.5% quantiles are the 95% credible intervals. The red line shows the true value for $\gamma_1, ...\gamma_8$, respectively and the true values are $\gamma_1 = 7.228863$, $\gamma_2 = 4.572165$, $\gamma_3 = 12.776484$, $\gamma_4 = 13.807305$, $\gamma_5 = 19.500778$, $\gamma_6 = 22.124091$, $\gamma_7 = 32.073988$ and $\gamma_8 = 30.614541$. The black line shows the posterior mean of the latent coefficients $\gamma_1, ...\gamma_8$. As we can see, the red line is not only in the 95% credible interval but also closely match the blue line.

Collectively, this simulation study demonstrates that our MCMC method can do an accurate job in estimating model parameters and latent coefficients.

6 Empirical Study

We use a real data set provided by a public university in the United States in empirical study. The data set record the donation behavior of 3699 alumni in a time period from 2009 to 2017. In this project, the data set from 2009 to 2016 is used for estimating the model parameters and latent coefficients. The data in 2017 is used for testing.

The posterior draws for the model parameters β_0, σ_r^2 and latent coefficients $\gamma_1, ...\gamma_8$ are drawn in the same fashion as before. For the initial values, we firstly chose the same initial values as we use in the simulation studies and generated the $\gamma_1^{(0)}$ to $\gamma_8^{(0)}$ in the same iterative procedure as we did in simulation study :

- $\beta_0^{(0)} = 0.3$
- $\sigma_r^{2(0)} = 5$
- $\gamma_0^{(0)} = 1$
- $\gamma_1^{(0)} = 1.864795$
- $\gamma_2^{(0)} = 4.157879$
- $\gamma_3^{(0)} = 5.238034$
- $\gamma_4^{(0)} = 7.018678$
- $\gamma_5^{(0)} = 6.933848$
- $\gamma_6^{(0)} = 8.592888$
- $\gamma_7^{(0)} = 9.280738$
- $\gamma_8^{(0)} = 8.421772$

Table 2 shows the estimates of the parameters and the latent coefficients and their standard deviations for the first group of initial values. We chose the burn-in period to be 5000 iterations. So, we reported the means and standard deviations of the posterior samples as parameters and latent coefficients estimates for the last 5000 iterations.

To see whether our estimates are sensitive to the initial value we chose, we set the second group of initial values which are far from the first group and generated the $\gamma_1^{(0)}$ to $\gamma_8^{(0)}$ in the same iterative procedure as we did in simulation study:

- $\beta_0^{(0)} = 1$
- $\sigma_r^{2(0)} = 10$
- $\gamma_0^{(0)} = 2$
- $\gamma_1^{(0)} = 3.827892$
- $\gamma_2^{(0)} = 2.342971$
- $\gamma_3^{(0)} = 8.610150$
- $\gamma_4^{(0)} = 9.149567$
- $\gamma_5^{(0)} = 10.888210$
- $\gamma_6^{(0)} = 18.649199$
- $\gamma_7^{(0)} = 22.649378$
- $\gamma_8^{(0)} = 22.935937$

Table 3 shows the estimates of the parameters and the latent coefficients and their standard deviations for the second group of initial values. We chose the burn-in period to be 5000 iterations. So, we reported the means and standard deviations of the posterior samples as parameters and latent coefficients estimates for the last 5000 iterations.

We can see from the table 2 and table 3, the estimates of model parameters and latent coefficients and their standard deviations are matched for different sets of initial values. It shows that the model parameters and latent coefficients will converge to the same values no matter what initial values we chose for the models.

The figure 5 plot the 95% credible interval for the latent coefficients $\gamma_1, ...\gamma_8$ using real data. The year 2009 to 2016 on the x-axis are corresponding to γ_1 to γ_8. The two blue lines represent the 2.5% and 97.5% quantile line. The intervals between the 2.5% and 97.5% quantiles are the 95% credible intervals. The black line shows the posterior mean for $\gamma_1, ...\gamma_8$.

We can see that the posterior mean of the latent coefficients $\gamma_1, ...\gamma_8$ are all positive. It indicates a positive relationship between the $Z_{i(t-1)}$ and β_{it}. That is, if alumni attend higher proportion of the activities held, the alumni will be more likely to donate in the next year. Additionally, the trend from γ_1 to γ_4 is monotonically decreasing, while the trend from γ_6 to γ_8 is monotonically increasing.

The figure 6 show the density plots for model parameters β_0, σ_r^2 and latent coefficients $\gamma_1, ...\gamma_8$, respectively. We chose the burn-in period to be 5000 iterations. So, the density plots shows 5000 iterations after the burn-in period.

7 Prediction

In this section, we predict the future alumni-giving behavior using simulated data and real date, respectively. We evaluate the prediction accuracy of our models by ROC Curve and discuss how the missing value will impact the prediction accuracy of our model.

ROC curve is a probability curve which shows the diagnostic ability of a binary classifier system when its discrimination threshold is varied. In our case, we plotted the ROC curve as a diagnostic tool for the estimated π_{it}. AUC measures the area under the ROC curve. AUC represents the degree of separability and shows how much the model is capable of distinguishing between classes. In our model, it is donate or not donate. In general, A high value of AUC means a model have high prediction accuracy.

We estimated our model parameters β_0, σ_r^2 and latent coefficients $\gamma_1, ...\gamma_T$ from the training data. The model parameters and latent coefficients estimated can fit our model which can predict $\hat{\gamma}_{T+1}$. With the $\hat{\gamma}_{T+1}$, we can predict $\hat{\beta}_{i(T+1)}$ and $\hat{\pi}_{i(T+1)}$.

In the iteration $i = 1, 2, 3, ...N$

- $\hat{\beta}_{i(T+1)} = \hat{\beta}_0 + \hat{\gamma}_{T+1} \cdot Z_{iT}$
- $\hat{\pi}_{i(T+1)} = \frac{\exp(\hat{\beta}_{i(T+1)})}{1+\exp(\hat{\beta}_{i(T+1)})}$

7.1 Prediction using simulated data

The Figure 7 plot the ROC Curve using simulated data. The AUC is the shaded area under curve and the value of AUC is 0.6223774.

In the real data, we found a large number of observation whose $Z_{i(t-1)}$ is 0. Since missing values in the real data were recorded as 0, most $Z_{i(t-1)}$ with value 0 are actually missing values. In order to exclude the effect of missing values, we are also interested in the prediction accuracy for the data without observations whose $Z_{i(t-1)}$ is 0.

The Figure 8 plot the ROC Curve using simulated data without observations whose Z_{i2016} is 0. As we can see, the ROC curve is more closed to the upper left corner in the figure, compared with the ROC curve in the figure 6. The AUC of the ROC curve is 0.7379325.

7.2 Prediction using real data

The Figure 9 plot the ROC Curve for real data. The AUC of the ROC curve is 0.5944497. At the first glance, the value of AUC is not good as we expected. However, as we explore further into the data set, we found that the data set for 2016 has serious missing value problem. There are 2705 observations whose Z_{i2016} are 0 and these observations share 73% of all observations in 2016. Although few alumni did not attend any activities on 2016, most of the 0 are due to the missing value.

In order to exclude the effect of missing values, we plot another ROC curve using the real data without observations whose Z_{i2016} are 0. The figure 10 show the ROC curve using the real data without missing value. As we can see, the ROC curve is more closed to the upper left corner, compared with the ROC curve in the figure 9. The AUC value of the ROC curve in the figure 10 is 0.7734138 which indicate a good prediction accuracy for our model.

7.3 Discussion for the prediction

As we can see, the AUC value is much larger when our model predict on the data without the observations whose Z_{i2016} are 0 than on the full data, no matter using simulated data or real data. This is simply because when the $Z_{i(t-1)}$ is 0, the parameter γ_t will lose its role in prediction and the parameter β_0 became the only model parameter which determine the π_{it}, thus lowering the prediction accuracy of our model.

8 Conclusion

Alumni-giving has already became an important funding source for universities and colleges in the United States. This paper helps to build a new model to prediction the future Alumni-giving behavior. In the simulation studies, We examined the accuracy of parameter estimates to ensure that our MCMC method can indeed identify model parameters and latent coefficients using the simulated data. In the empirical study, we show that our model are not sensitive to the initial value which we chose for the models. We predicted the future alumni-giving behavior using the simulated data and real data, respectively, and showed our model can have high accuracy of prediction using the data without missing values.

In order to raise more funding from donation, universities and colleges should share more fundraising efforts to the alumni who attend high proportion of the activities in that year. This is simply because there is a positive relationship between the $Z_{i(t-1)}$ and β_{it}. As we can see, the value of γ_t is changing year by year. In some year , even the alumni attend high proportion of activities held, the alumni may not donate in the next year due to a small value of γ_t for the years. In the end, the universities and colleges should exclude the missing value as possible as they can if they use our model for prediction. As we discussed in the prediction, the missing value will significantly lower the prediction accuracy of our model.

A Appendix

	Mean	Initial Values	True Values
sigr2	35.99	5	20
b0	0.49	0.3	0.5
r1	8.16	5.873067	7.228863
r2	5.28	5.135799	4.572165
r3	12.75	12.016019	12.776484
r4	13.72	14.115418	13.807305
r5	20.29	16.631101	19.500778
r6	21.64	21.505064	22.124091
r7	34.14	34.072877	32.073988
r8	31.48	29.090717	30.614541

Table 1: This table reports simulation results on the accuracy of MCMC estimators of parameters σ_r^2, β_0 and latent coefficients $\gamma_1, ...\gamma_8$. In the table above, we report the mean of the estimated parameters and latent coefficients, the initial values of the parameters and latent coefficients, the true values of model parameters and latent coefficients used in the simulation. We discard the first 5000 iterations, and we use the next 5000 iterations in our MCMC estimation.

	Mean	SD	Initial Values
sigr2	14.41	10.62	5
b0	0.19	0.01	0.3
r1	24.64	2.14	1.864795
r2	21.69	1.61	4.157879
r3	19.70	1.66	5.238034
r4	13.59	1.61	7.018678
r5	15.96	1.40	6.933848
r6	13.55	1.22	8.592888
r7	19.00	1.57	9.280738
r8	21.72	1.70	8.421772

Table 2: This table provides MCMC estimates of the model parameters and latent coefficients using real data from 2009 to 2016 and the first set of initial values. Parameter and latent coefficients estimates and standard errors shown are the mean and standard deviation of posterior distributions of model parameters σ_r^2, β_0 and latent coefficients $\gamma_1, ...\gamma_8$. In the MCMC simulation, we discard the first 5000 iterations, and we use the next 5000 iterations in our MCMC estimation.

	Mean	SD	Initial Values
sigr2	14.53	10.67	10
b0	0.19	0.01	1
r1	24.50	2.14	3.827892
r2	21.74	1.65	2.342971
r3	19.70	1.70	8.610150
r4	13.56	1.58	9.149567
r5	16.03	1.53	10.888210
r6	13.64	1.27	18.649199
r7	19.27	1.63	22.649378
r8	21.70	1.79	22.935937

Table 3: This table provides MCMC estimates of the model parameters and latent coefficients using real data from 2009 to 2016 and the second set of initial values. Parameter and latent coefficients estimates and standard errors shown are the mean and standard deviation of posterior distributions of model parameters σ_r^2, β_0 and latent coefficients $\gamma_1, ...\gamma_8$. In the MCMC simulation, we discard the first 5000 iterations, and we use the next 5000 iterations in our MCMC estimation.

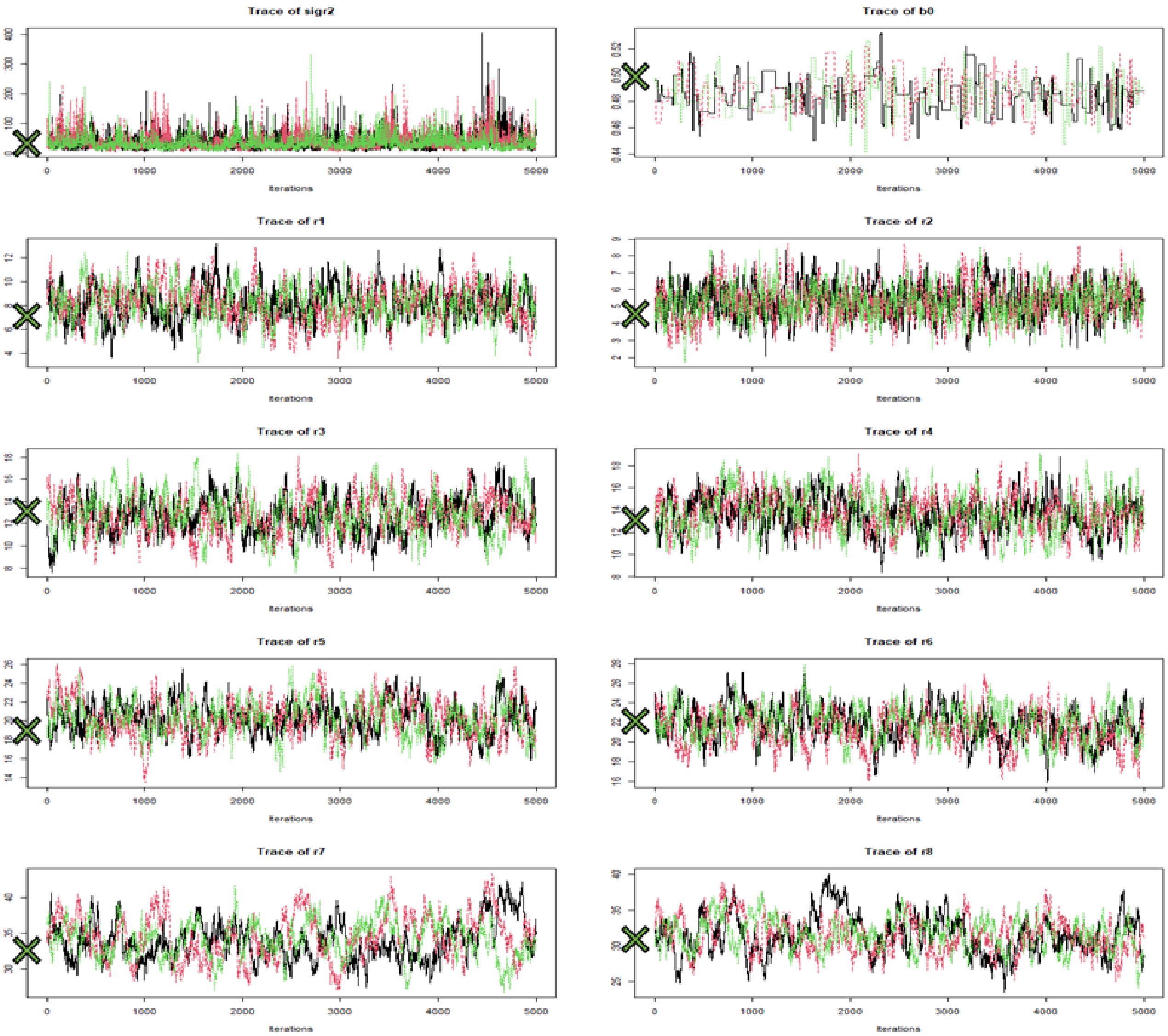

Figure 1: Trace-plots of σ_r^2, β_0, $\gamma_1, ...\gamma_8$ using simulated data. The red crosses indicate the true values of the parameters and latent coefficients. The true values are $\sigma_r^2 = 20$, $\beta_0 = 0.5$, $\gamma_1 = 7.228863$, $\gamma_2 = 4.572165$, $\gamma_3 = 12.776484$, $\gamma_4 = 13.807305$, $\gamma_5 = 19.500778$, $\gamma_6 = 22.124091$, $\gamma_7 = 32.073988$ and $\gamma_8 = 30.614541$

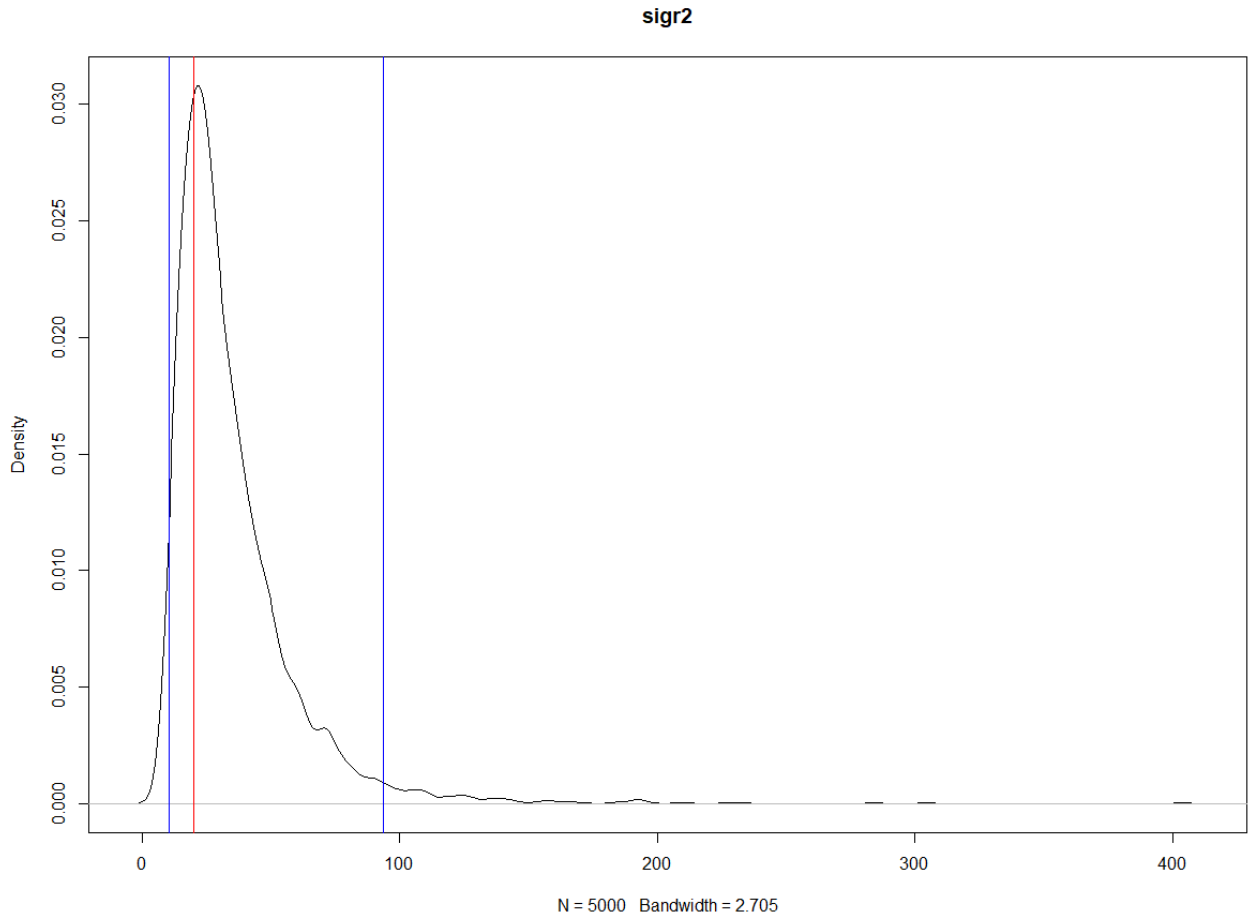

Figure 2: Density plot of MCMC samples of parameter σ_r^2 using simulated data. The 95% credible interval is marked by the two blue vertical lines, which indicate the 2.5% and 97.5% quantiles, respectively. The true value of σ_r^2 marked by the red vertical line and true value of σ_r^2 is 20.

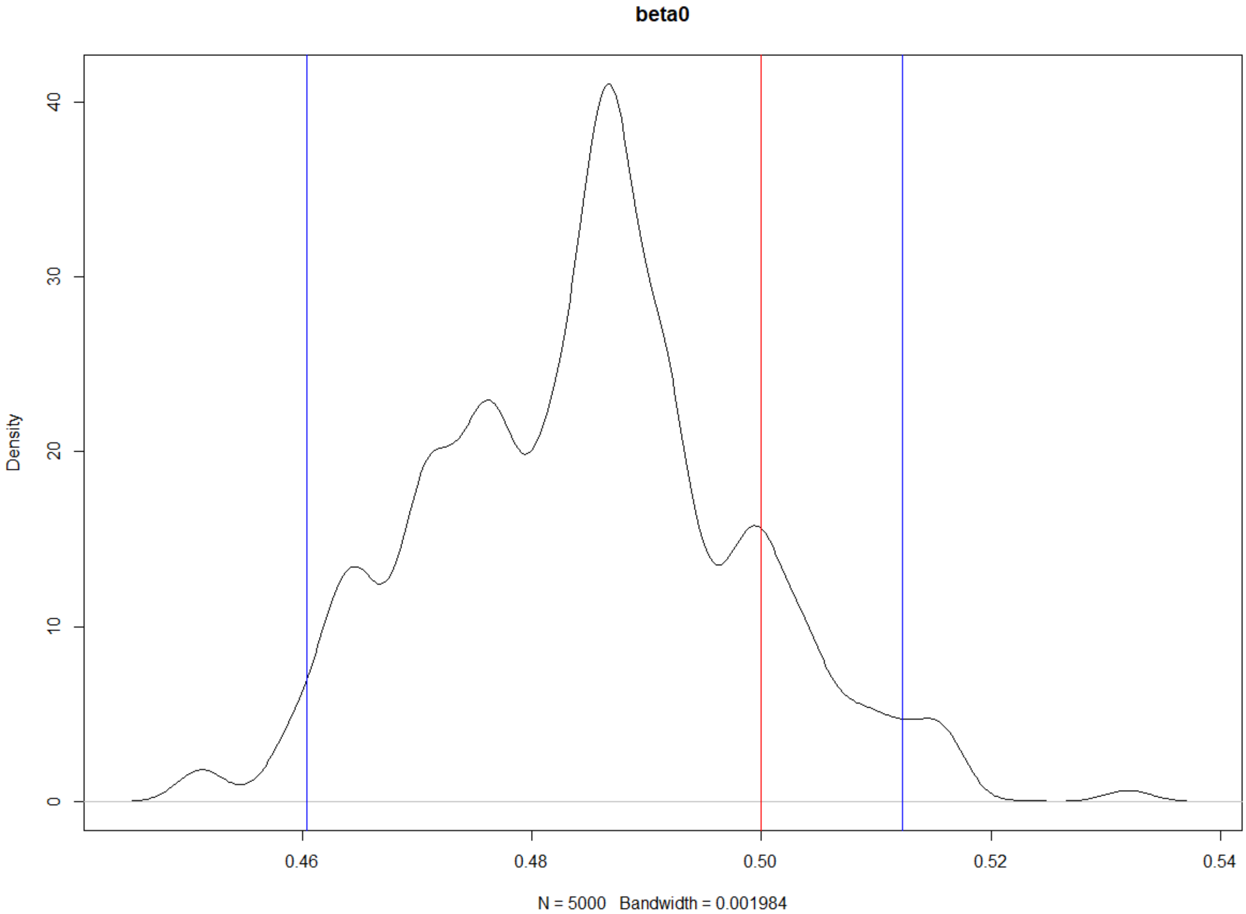

Figure 3: Density plot of MCMC samples of parameter β_0 using simulated data. The 95% credible interval is marked by the two blue vertical lines, which indicate the 2.5% and 97.5% quantiles, respectively. The true value of β_0 marked by the red vertical line and true value of β_0 is 0.5.

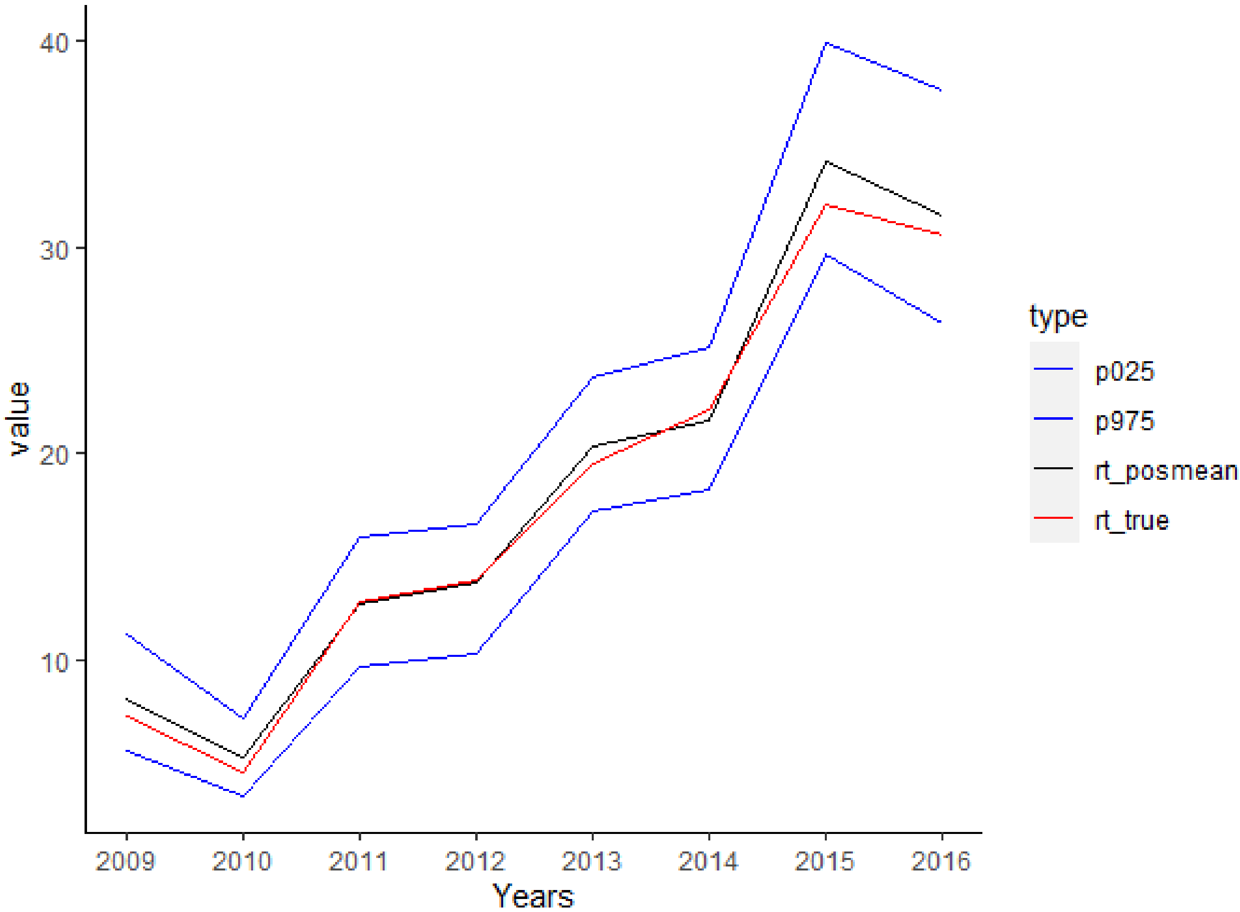

Figure 4: This figure plot the 95% credible interval for the $\gamma_1, ...\gamma_8$ using simulated data. The year 2009 to 2016 on the x-axis are corresponding to γ_1 to γ_8. The two blue lines represent the 2.5% and 97.5% quantile line. The intervals between the 2.5% and 97.5% quantiles are the 95% credible intervals. The red line shows the true value for $\gamma_1, ...\gamma_8$, respectively and the true values are $\gamma_1 = 7.228863$, $\gamma_2 = 4.572165$, $\gamma_3 = 12.776484$, $\gamma_4 = 13.807305$, $\gamma_5 = 19.500778$, $\gamma_6 = 22.124091$, $\gamma_7 = 32.073988$ and $\gamma_8 = 30.614541$. The black line shows the posterior mean for $\gamma_1, ...\gamma_8$.

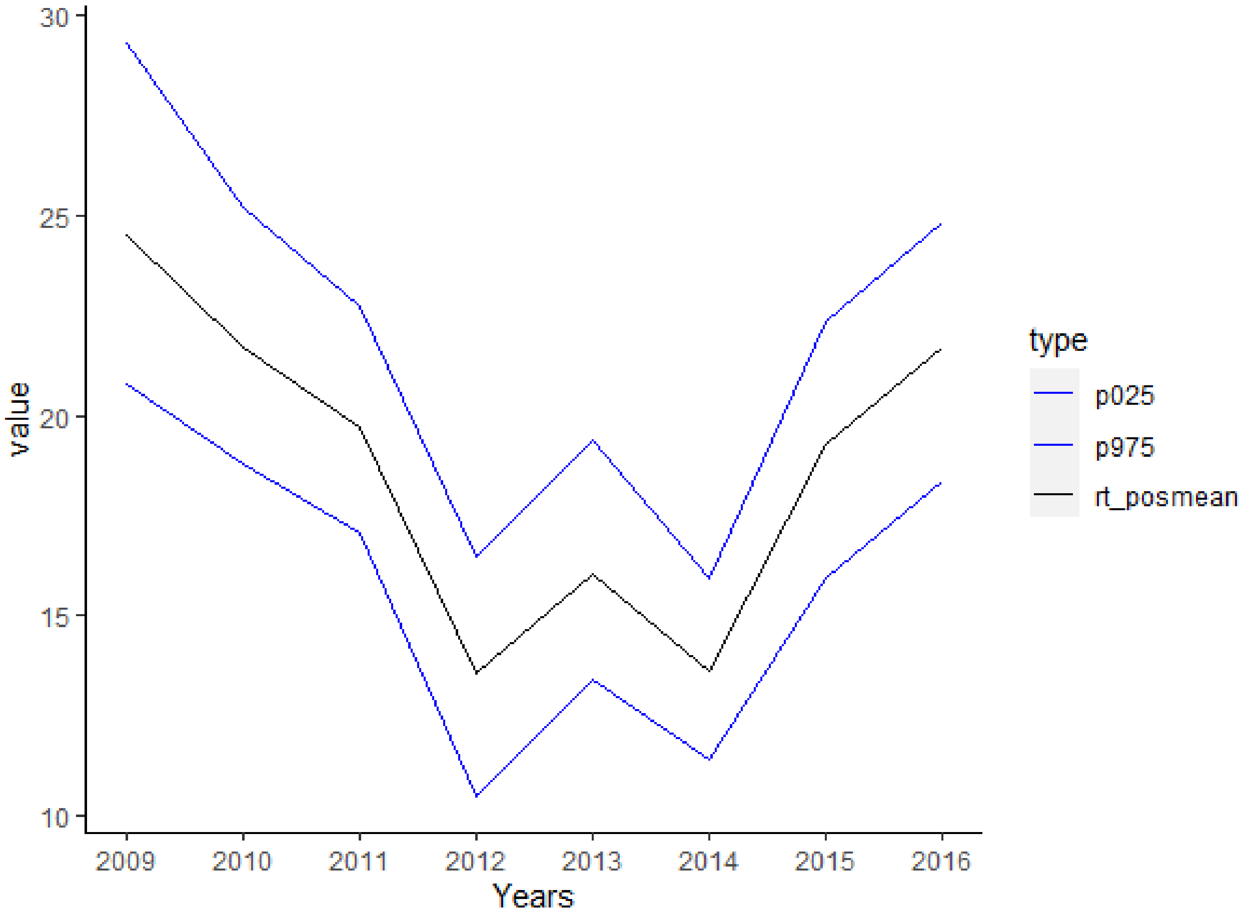

Figure 5: This figure plots the 95% credible interval for the $\gamma_1, ...\gamma_8$ using real data and the first set of initial values. The year 2009 to 2016 on the x-axis are corresponding to γ_1 to γ_8. The two blue lines represent the 2.5% and 97.5% quantile line. The intervals between the 2.5% and 97.5% quantiles are the 95% credible intervals. The black line shows the posterior mean for $\gamma_1, ...\gamma_8$.

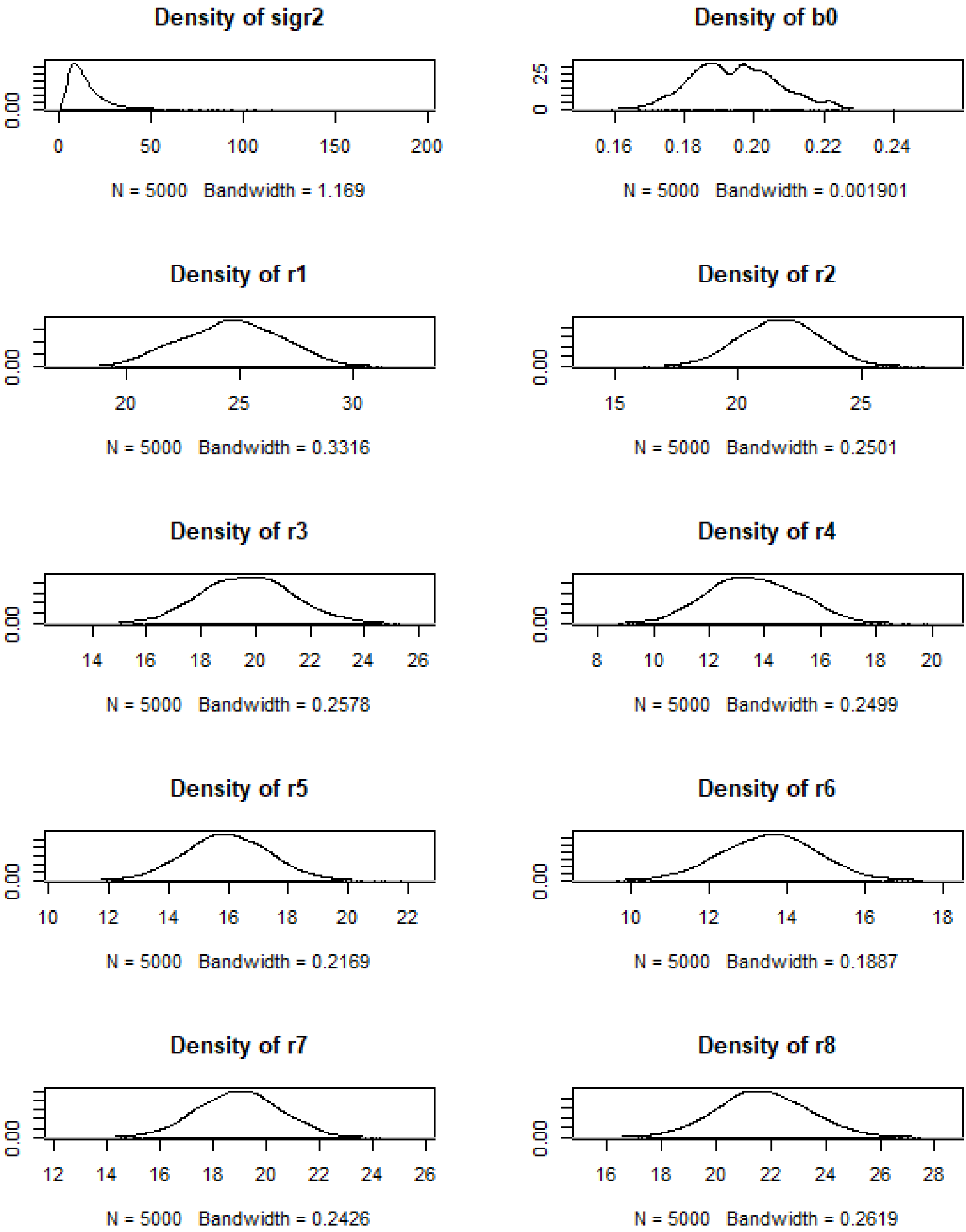

Figure 6: The figure shows the density plots for model parameters β_0, σ_r^2 and latent coefficients $\gamma_1, ...\gamma_8$ using real data and the first set of initial values, respectively. We chose the burn-in period to be 5000 iterations. So, the density plots shows 5000 iterations after the burn-in period.

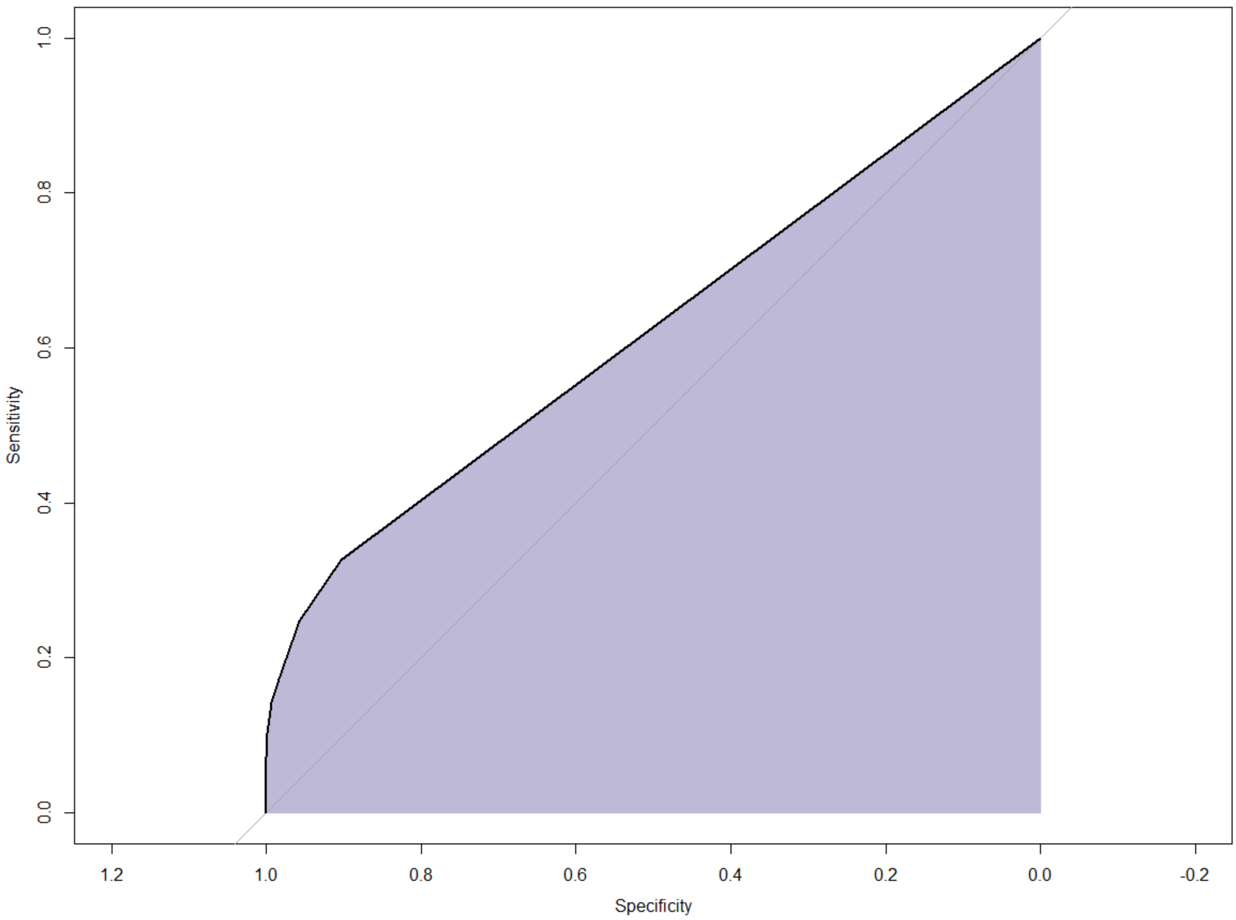

Figure 7: This figure plots the ROC Curve using simulated data. The AUC is the shaded area under curve and the value of AUC is 0.6223774.

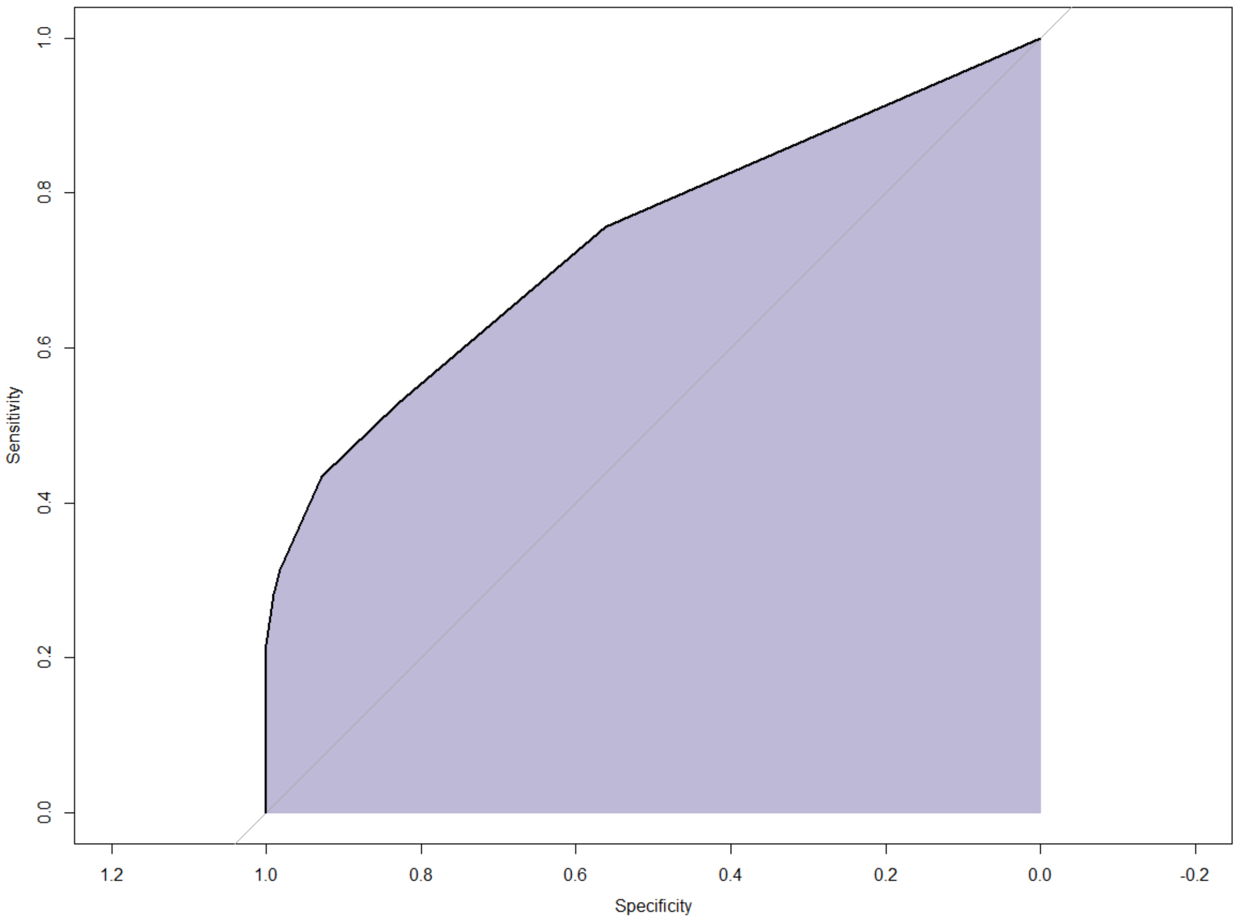

Figure 8: This figure plots the ROC Curve using simulated data without observations whose Z_{i2016} are 0. The AUC is the shaded area under curve and the value of AUC is 0.7379325.

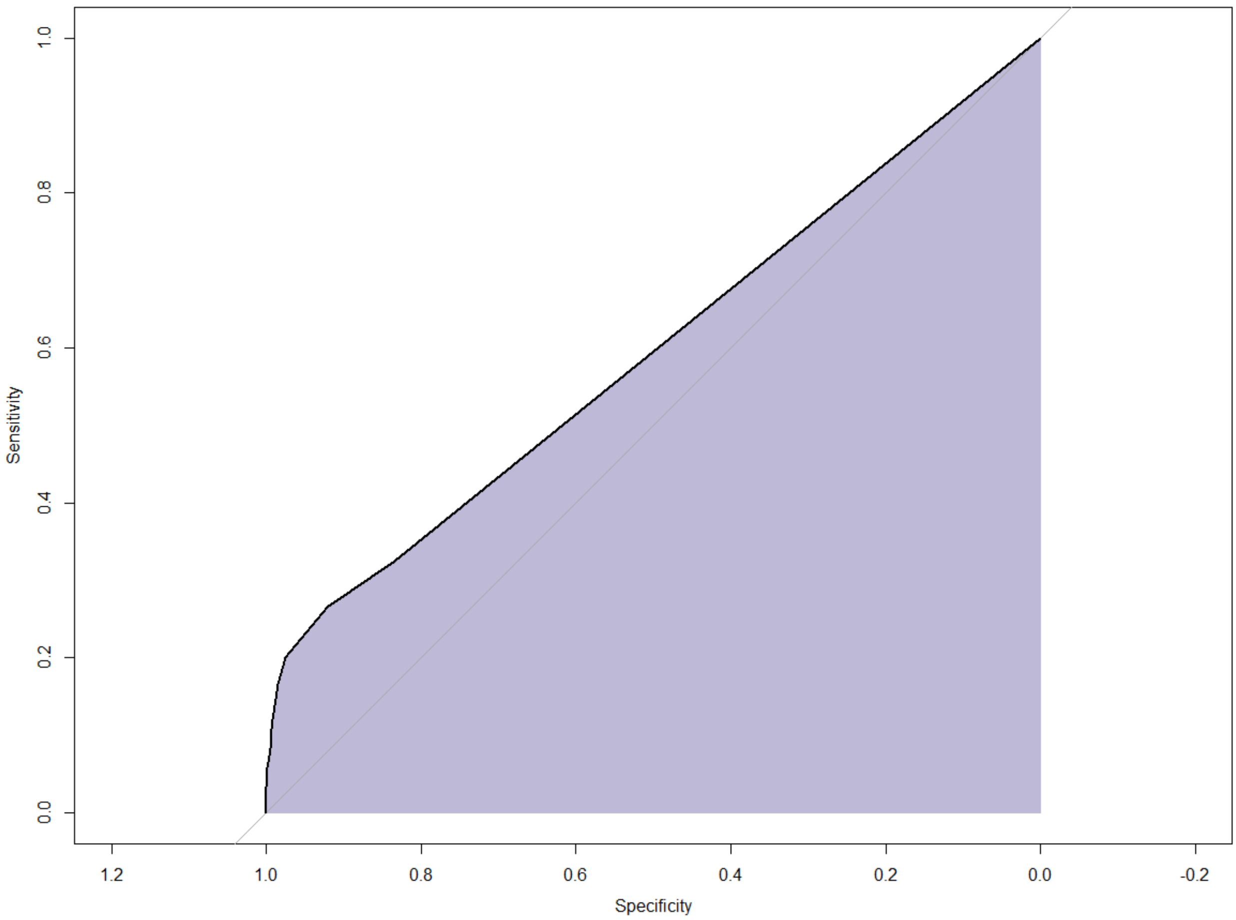

Figure 9: This figure plos the ROC Curve using real data. The AUC is the shaded area under curve and the value of AUC is 0.5944497.

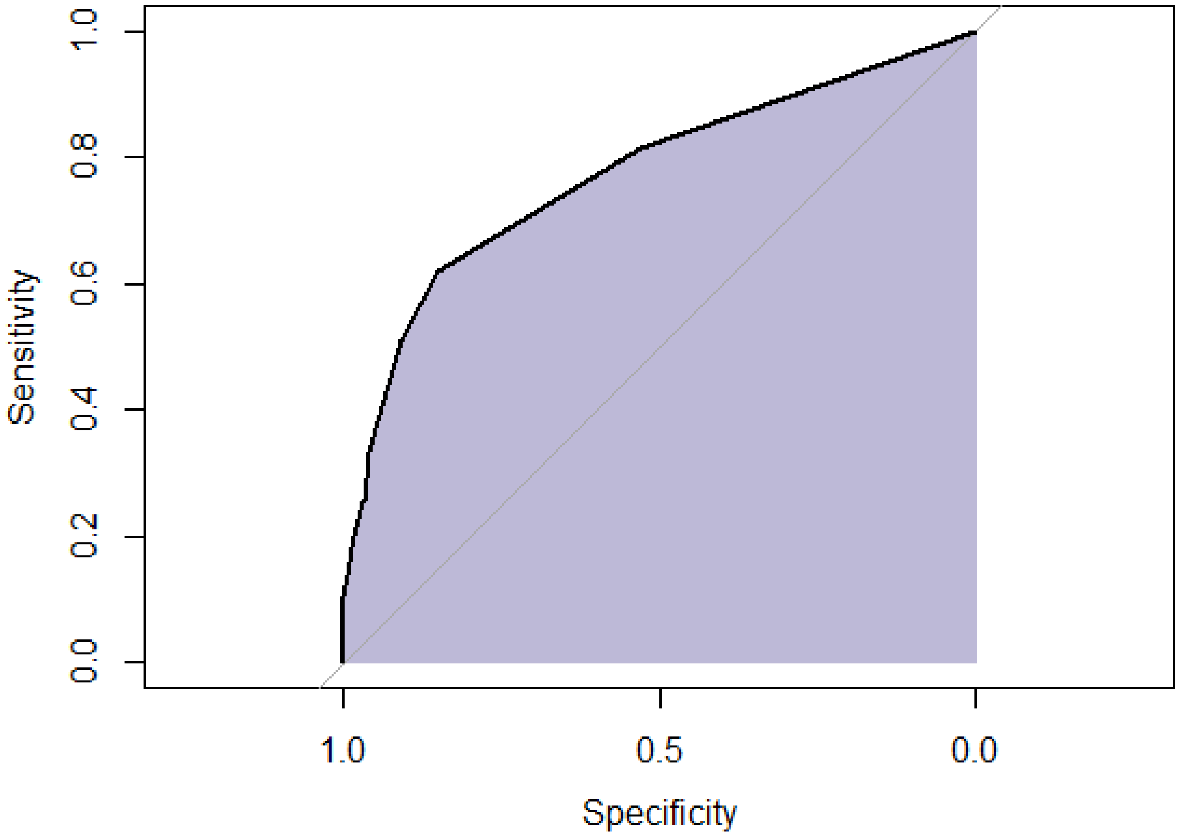

Figure 10: This figure plots the ROC Curve using real data without observations whose Z_{i2016} are 0. The AUC is the shaded area under curve and the value of AUC is 0.7734138.

References

[1] Wesley E. Lindahl and Christopher Winship. Predictive models for annual fundraising and major gift fundraising. *Nonprofit Management and Leadership*, 3(1):43–64, 1992.

[2] Oded Netzer, James M Lattin, and Vikram Srinivasan. A hidden markov model of customer relationship dynamics. *Marketing science*, 27(2):185–204, 2008.

[3] Xiaogeng Sun, Sharon C Hoffman, and Marilyn L Grady. A multivariate causal model of alumni giving: Implications for alumni fundraisers. *International Journal of Educational Advancement*, 7(4):307–332, 2007.