Ministry of Higher Education &
Scientific Research
University of Al-Qadisiyah
College of Administration & Economics
Department of Statistics

Inference with Gamma and Inverse Gamma Prior Densities in Tobit Regression Problem with a Practical Application

A Thesis by

Esmaeel Ali Hamad Al-Silmawi

Submitted to the Council of the College of Administration & Economics / University of Al-Qadisiyah in Partial Fulfilment of the Requirements for Master Degree (MSc) in Statistical Sciences

Supervised by

Prof. Dr. Rahim J. Al-Hamzawi

2021 AD **1443 AH**

بِسْمِ اللّهِ الرَّحْمَنِ الرحِيمِ

وَيَسْأَلُونَكَ عَنِ الرُّوحِ قُلِ الرُّوحُ مِنْ أَمْرِ رَبِّي وَمَا

أُوتِيتُم مِّنَ الْعِلْمِ إِلاَّ قَلِيلاً

صدق الله العلي العظيم

سورة الاسراء الآية (85)

ACKNOWLEDGMENTS

First of all, I must thank Allah Almighty, Who enabled me to complete this thesis, which I hope will be of benefit to all who read it and seek help from it.

Also. I would like to express my deep thanks to my supervisor, Professor Dr. Rahim J. Al-Hamzawi, for his valuable observations and assistance throughout the period of study and writing the thesis.

Hence, A word of gratitude goes to the staff of Statistics Department for their help and assistance throughout my study.

I must mention my gratitude to the staff of statistics department, especially Prof. Dr. Mohammed Al-Sharoot, Prof. Dr. Ahmad Naeem, Prof. Dr. Ali Al-kenani, Dr. Muhannad Al-Saadony, Dr. Taha Mohamed, Dr. Fadhil Alhusseini, Prof. Dr. Tahir Dikheel and Dr. Hassan Sami for their help and assistance throughout my journey of knowledge.

I also extend my thanks and appreciation to my teacher, Prof. Dr. Abbas Lafta Al-eqabi.

I also thank Dr. Sabah Mohil, a nephrologist, who helped me get the data that was used in the chapter class.

Finally, I thank my dear parents, , brothers, family, sisters, all my friends and especially My MA Colleagues.

DEDICATION

To the endless tender, the still fruitful tree

My dear father

To the source of tenderness

My dear mother

To My support in this life

My dear brothers

To my eyes

My sisters

To my sons

Reda, Kumail, Ali

To My Beloved

Iraq

To everyone who wished me success, I dedicate the fruit of my humble effort.

Esmaeel

Approval of the linguistics expert

This is to hereby that the entitled (**Inference with Gamma and Inverse Gamma Prior Densities in Tobit Regression Problem with a Practical Application**) has been reviewed in terms of stylistics and linguistics (grammar and spelling). Therefore, after the modification of all recommended notes thesis has become free of all linguistics errors and ready to be defended and used as a scientific method to award the degree of master in statistical sciences.

Signature:

Name: Asst. Prof. Mazin Jasim Al-Hilu (PhD)

Date:

Supervisor's recommendation

I certify that the thesis entitled (**Inference with Gamma and Inverse Gamma Prior Densities in Tobit Regression Problem with a Practical Application)**has been under my supervision for the student (Esmaeel Ali Hamad)in the Department of Statistics / College of Administration & Economics / University of Al-Qadisiyah as it is part of the requirements for master's degree in statistics sciences.

Signature:

Supervisor: **Prof. Dr. Rahim Al-Hamzawi**

Date:

Recommendation of the head of the Graduate Studies Committee

Based on the available recommendation , I would like to forward this thesis for discussion.

Signature:

Name: Asst. Prof. Dr. Muhannad F.Al-Saadony

Chairman of the Higher Studies Committee in department of statistics

ABSTRACT

The procedure of the variable selection (VS) methods was used to evaluate the relationship between the set of explanatory variables and the dependent variable. Many methods for variable selection have been proposed over the years. In the recent years, there has been active research on variable selection using Bayesian regression methods.

In this thesis, the researcher has developed Bayesian methods for variable selection in left censored, right censored data, and binary data that lead to new Gibbs sampler methods with tractable full conditional posteriors. Through extensive simulation examples and real data analyses, we compare the performance of our proposed method for left censored data with some of the existing Bayesian and non-Bayesian methods. Results show that our proposed method for left censored data performs very well in comparison to the existing approaches.

List of Abbreviations

Abbreviations	Meaning
VS	Variable Selection
AIC	Akaike Information Criterion
MLE	Maximum Likelihood Estimation
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
MSE	Mean Square Errors
S.D	Standard Deviation
RSS	Residual Sum of Squares
OLS	Ordinary least Squares
Lasso	Least Absolute Shrinkage and Selection Operator
MCMC	Markov Chain Monte Carlo
SMN	Using a Scale Mixture Normal
SMU	Scale Mixture of Uniform
GG	Generalized Gaussian
LAD	Least Absolute Deviation
MMSE	Median of Mean Squared Errors
TR	Tobit regression
LTR	Lasso Tobit regression
BLTR	Bayesian Lasso Tobit regression
NewBLR	New Bayesian Left regression
eGFR	Estimated Glomerular Filtration Rate

Table of Contents

Table of Contents

List of Figures

Table of Contents

Table of Contents

Chapter One

Introduction

and

Literature Review

1.1 Introduction

Left censored regression is a statistical technique used to illustrate who is the response variable is censored from lower. These models are used in several fields, such as medicine, astronomy, finance, etc. Also, left censored regression is one of the most important methods used to evaluate the relationship between the set of explanatory variables and the dependent variable. In recent years, high dimensional data arise in many fields such as, biology, ecology, economics, medicine finance, social sciences. Thus, one of the most important problems in building left censored linear regression model is how to remove irrelevant predictors from the final model. Removing irrelevant predictors improve the prediction for the final model and obtain better interpretation.

Many methods for variable selection have been proposed over the years. In the recent years, there has been active research on variable selection using Bayesian regression methods.

In this thesis, Bayesian methods have been developed for variable selection in left censored data. We also extended the proposed methods in tobit and binary regression models.

1.2 Literature review

Left-censored linear regression models are quite popular models and have been deeply considered in the last three decades, This model assumes:

$$y_i = max\{c, y_i^*\} \qquad i = 1, \dots, n,$$

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \qquad (1.1)$$

where c is a left-censored point, $\boldsymbol{x_i}$ is a vector of predictors, $\boldsymbol{\beta}$ is a vector of the regression coefficients and ε_i is an error term, The zero censored model (tobit model) is a special case from (1) and is defined as:

$$y_i = max\{0, y_i^*\} \qquad i = 1, \dots, n,$$

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \qquad (1.2)$$

high-dimensional statistical modeling, assumes that the underlying true model is sparse, meaning that there are many duplicate variables, and un unknown only subset of predictors is active in the regression model (1.1). So, the subset selection problem is to find these predictors when analyzing problems data and these data are some explanatory variables and dependent variables. In any scientific research containing these variables, it is very important that the relevant variables be correctly identified. Some explanatory variables or predictions are not important. These variables should be removed from the form. Therefore, many researchers focus on Classical methods of (VS) to get the suitable model.

For example, Akaike Information Criterion (AIC), that was suggested by Akaike (1974) the formula of AIC procedure is:

$$AIC = -2\, ln\, l + 2k, \qquad (1.3)$$

where l be the maximum likelihood estimation function (MLE), k is the number of parameters, and which is one of the most popular methods that is used for (VS).

Bozdogan (1987) presented the AIC criterion as important, only when compared with other AIC values of the same data group.

Nishii (1984) showed that (AIC) produces the disconsonant model. Consequently, if the sample size is large, the model defined using AIC is inconsistent (Dziak et al., 2005; Javed and Mantalos, 2013).

Schwarz (1978) proposition the Bayesian Information Criterion (BIC) to avoid the problem in (AIC), and the formula of this criterion;

$$BIC = -2\,ln\,l + k\,ln\,n, \tag{1.4}$$

where n is the sample size, which was presented to treating the problem in the (AIC) (Javed and Mantalos, 2013; Mallick, 2015), and at the same time choosing model with perfect features, as well it works fine when it is $(k < n)$. Also, the variable selection problem cannot be treated by this criterion.

Mallows (1973) suggested the criteria to choosing a variable known as:

$$C_k = \frac{RSS(k)}{S^2} - n + 2k, \tag{1.5}$$

where S^2 is the Mean Square Error (MSE), RSS (k) is the residual sum of squares, k is a number of co-variances in the model and n is the sample size of data. This model will be comparatively exact if the value of C_k is teeny.

Nishii (1984) showed that C_k is inconsistent in choosing the correct model. Woodroofe (1982) showed that C_k choose the moderate model. and frequently selecting a big model when $n \to \infty$.

George and McCulloch (1993) the most important main ideas in this criterion are first, adding vector to the binary parameters.

Alhamzawi & Ali (2018) if the numbers of variables are large, according to this criterion the process of selecting the important variables takes a long time.

Recently, studies have indicated the regularization methods that utilize estimation and variables selection at the same time are effective. One of the advantages of these methods is to get better forecast precision from linear regression (Zhou,2006; Tibshirani, 1996). According to this property, it can estimate the parameters until if a number of covariates are greater than the sample size relatively Alhamzawi & Ali (2018).

Lately, due to the process of selecting and estimating sufficient parameters, the methods of organization have become more common. The form of regularization methods can be defined as follow:

$$\hat{\beta}_{RIDGE} = \underset{\beta}{argmin}\,(y - X\beta)'(y - X\beta) + \mathcal{G}_{\lambda}(\beta)\,, \qquad (1.6)$$

where $\mathcal{G}_{\lambda}(\beta)$ is a function of the model coefficients which controls the degree of penalty concerning inspect parameter $\lambda > 0$.

Donoho and Johnstone (1994) suggested the idea of VS by regularization, then Tibshirani (1996) developed it.

Hoerl and Kennard (1970) introduced methods of Ridge regression, which gives forecasting performance with minimized variance is better than (OLS) Ordinary least Squares estimates. Anyway, because the Ridge

regression keeps every predictor forever within the model, it cannot produce an optimal model.

Tibshirani (1996) suggested the Lasso regression which automatically choose the necessary variables by shrinking several uninfluential variables toward to zero.

Frank and Friedman (1993) suggested that Bridge regression has attractive properties like as Oracle and Impartiality , as well as the variable selection and parameter estimation of the model, however, the convergent covariance matrix and bootstrap studied criterion errors are unstable.

Zou (2006) suggested another technique, which is an update to the Lasso method, that is known as Adaptive Lasso regression, by adding dissimilar weights to different coefficients, which creates estimates consistent and unbiased .

Zou and Hastie (2005) proposed the elastic net regression model, which mediates among the Lasso (L1) and the Ridge (L2) penalties. In this way, the choice of the important variables and conjectting the parameters is being performed well If there is a large correlation between variables or even when they occurs when ($k > n$), but at the same time it complicated that needs to the aloft calculation cost.

Yuan and Lin (2006) suggested the Group Lasso, which was developed by Kim et al., (2006) to contain general loss functions. This method has the Oracle property. Also, the main idea of selecting important variables in this group it be chosen at the group level, but at the same time, insignificant variables cannot be deleted completely because they determine the variables at the selfsame group.

Wang and Leng (2008) proposed the adaptive group Lasso , to avoid the problems of the Group Lasso and mend its carrying out by forcing a weight on each variable. Moreover, the adaptive group Lasso is designed to specify the real model and is consistent with the Oracle feature. but can't choose the bi-level variable.

Meier et al. (2008) supposed the Lasso group with logistic regression.

Zhao et al. (2009) produced absolute public penalty method which is stretch of the Lasso method.

Huang et al. (2009) suggested a group Bridge regression, and it is have the ability to choosing a bi-level variable with oracle property and sparsity.

Griffin and Brown (2010) introduced Bayesian adaptive Lasso with non-convex penalization.

Park and Casella (2008) provided that by using a scale mixture normal (SMN) to create a hierarchical model, the Bayesian process can be utilized to speculate Lasso parameters. Bayesian process has been utilized by some researchers in their methods for example; Hans (2009) suggested Bayesian Lasso regression, Sun et al.(2010) proposed Bayesian adaptive Lasso and reiterated adaptive Lasso, Adaptive Lasso is more calculation actual than the usually used regression methods by forcing various adaptive weights and repetitive routine of these adaptive weights.

Li and Lin (2010) proposed a way to fix the problem of the elastic net problem by using a Gibbs sampler through Bayesian elastic net and avert the dual shrinkage problem in the elastic net.

Chen et al. (2011) upgraded a Gibbs sampler for Bayesian Lasso by changeable jump Markov Chain Monte Carlo (MCMC).

Simon et al. (2013) propositioned the sparse group Lasso, by utilizing the mixture of Lasso and Lasso group penalty on the parameter can be chosen bi-level variable. Nevertheless, because the shrinkage producing from the penalty forced on all parameters the estimators are passably biased.

Mallick and Yi (2014) by using the scale mixture of uniform (SMU) to show the Laplace density and inserting a modern hierarchical model for Bayesian Group Lasso, suggested a new Bayesian Lasso regression. Also, the performance of this method was compared well with the way the Bayesian Lasso method.

Mallick and Yi (2017) by forcing various shrinkage coefficients, to choose bi-level variables, suggested a Bayesian group Bridge. This proposal leads to gathering information among the variables within the group.

Mallick and Yi (2018) proposed Bayesian Bridge regression, utilized (SMU) as new characterization for Generalized Gaussian (GG) prior for it do the (MCMC) algorithm simple to carry out. This method has good estimates comparison with other procedures.

Powell (1986) introduced is to explain how the least absolute deviation (LAD) estimation method for the censored regression model can be expanded to wider quantiles.

Chib (1992) introduced Appropriate develops of Monte Carlo proceedings according to symmetric multivariate-t distributions.

Buchinsky and Hahn (1998) institute an alternate estimator for the linear censored quantile regression model.

Bilias, Y., S. Chen, and Z. Ying (2000) suggested arithmetical easy rechoosing methods Powell's approach in the resampling step.

Park and Casella (2008) motivated as to develop new Bayesian hierarchical model for the lasso binary regression model.

Bae and Mallick (2004) suggested a corresponding Gibbs sampler for probit binary regression.

Prentice (1988) introduced models for correlated binary data.

Manski (1985) introduced under weak distributional, the maximum score estimator of the coefficient vector of a binary response model is consistent.

Cavanagh (1987) and Kim and Pollard (1990) have shown converges between the random variable that maximizes a certain Gaussian process and the centered maximum score estimator.

Magel & Unruh (2013) introduced the generalized linear model with logit link is a natural fit for binary data because of the approaching curve and bounding range.

Alonzo & Pepe (2002) introduced that the perform parameter estimation can be utilities the generalized linear model methods for binary data.

Pepe (2000) based on the binary indicators had developed a method for fitting the regression model.

Albert & Chib (1993) for binary probit regression models showed an auxiliary variable approach under the Bayesian normal linear regression

model that renders the conditional distributions of the model parameters equivalent.

Manski (1975, 1985) determined the general semiparametric binary quantile regression estimator.

Kordas (2006) showed the simplicity of median for binary regression models other than the estimating quantiles.

Hilali & Alhamzawi (2019) suggested Bayesian adaptive Lasso Binary regression.

Benoit, & Alhamzawi (2013) suggested a Bayesian quantile regression method for binary data combined with a variable selection method.

Chapter Two

Bayesian Left Regression

2.1 Bayesian Left Regression

2.1.1 Introduction

Left-censored linear regression models are quite popular models and have been deeply considered in the last three decades, for example see, Powell (1986), Chib (1992), Buchinsky and Hahn (1998), Murphy et al. (1999), Bilias et al. (2000), Chay and Powell (2001), Chernozhukov and Hong (2003) and Müller and Van de Geer (2016). Suppose that the response y_i and the latent response $y_i{}^*$ are random variables connected by the following relationship

$$y_i = \max\{c, y_i{}^*\}, \;\; i = 1, \dots, n,$$

$$y_i{}^* = \boldsymbol{x_i}'\boldsymbol{\beta} + \varepsilon_i, \qquad (2.1)$$

where c is a left-censored point, $\boldsymbol{x_i}$ is a vector of predictors, $\boldsymbol{\beta}$ is a vector of the regression coefficients and ε_i is an error term, the zero censored model (tobit model) is a special case from (2.1) and is defined as:

$$y_i = \max\{0, y_i{}^*\}, \quad i = 1, \dots, n,$$

$$y_i{}^* = \boldsymbol{x_i}'\boldsymbol{\beta} + \varepsilon_i \qquad (2.2)$$

In high-dimensional data, we assume that only an unknown subset of predictors is active in the regression model (2.1), so that the subset selection problem is to find these predictors. In linear regression models, regularization methods are attractive methods that has received considerable attention over the last two decades for dealing with high dimensional data, see for example, lasso and it's Bayesian version (Alhamzawi and Taha Mohammad Ali, 2018; Park and Casella, 2008; Tibshirani, 1996), elastic net and it's Bayesian version (Alhamzawi et al., 2019; Alhamzawi and Ali, 2018; Li and Lin, 2010; Zou and Hastie,

2005), adaptive lasso and it's Bayesian version (Alhamzawi and Ali, 2018; Alhamzawi et al., 2012; Leng et al., 2014; Zou, 2006), and so on.

Andrews and Mallows (1974), suggested the following scale mixture of normal mixing with an exponential distribution,

$$\frac{l}{2}e^{-l|z|} = \int_0^\infty \frac{1}{\sqrt{2\pi w}} e^{-\frac{z^2}{2w}} \frac{l^2}{2} e^{-\frac{l^2 w}{2}} dl, \quad l > 0 \tag{2.3}$$

Park and Casella (2008) used the scale-mixture representation in (2.1) to proposed the Bayesian hierarchical model for the linear regression by assumption that,

$z = \beta, l = \frac{\lambda}{\sqrt{\sigma^2}}$ and $w = s^2\sigma^2$, then the scale mixture (2.3) became as follows,

$$\frac{\lambda}{2\sqrt{\sigma^2}} e^{-\frac{\lambda|\beta|}{2\sigma^2}} = \int_0^\infty \frac{1}{\sqrt{2\pi s^2\sigma^2}} e^{-\frac{\beta^2}{2s^2\sigma^2}} \frac{\lambda^2}{2} e^{-\frac{\lambda^2 s^2 \sigma^2}{2}} ds^2 \tag{2.4}$$

From representation (2.3), clearly that

$$\beta_j|\sigma^2, s_j{}^2 \sim N\left(0, \sigma^2 s_j{}^2\right), \qquad j = 1,2,\dots,k,$$

$$s_j{}^{-1} \sim \text{inverse} - \text{Gaussian}\left(\frac{1}{2}, \sqrt{\frac{\lambda^2\sigma^2}{\beta_j{}^2}}, \lambda^2\right), \qquad j = 1,2,\dots,k,$$

$$\sigma^2 \sim {}^{1}/_{\sigma^2}$$

Since Park and Casella (2008), different Bayesian regularization approaches have been proposed over the years, see for example, Carvalho et al. (2009), Carvalho et al. (2010), Griffin et al. (2010), Alhamzawi and Yu (2014), Bhattacharya et al. (2015), Alhamzawi (2015), Bhattacharya et al. (2016) and Alhamzawi (2017). Very recently, in the standard linear regression model, Bai and Ghosh (2018) considered normal scale mixture priors with beta prime densities for the regression coefficients. The

authors noted that this prior distributions can serve as both sparse and non-sparse priors. They showed that the beta prime density can be rewritten as a product of an independent gamma and inverse gamma densities. Specifically, Bai and Ghosh (2018) suggested the following prior distribution for the regression coefficients in standard linear regression model

$$\beta_j|\sigma^2, {\lambda_j}^2 {s_j}^2 \sim N\left(0, \sigma^2 {\lambda_j}^2 {s_j}^2\right), \qquad j = 1,2, \dots, k,$$

$$\lambda_j{}^2 \sim Gamma(a, 1), \qquad j = 1,2, \dots, k, \qquad (2.5)$$

$$s_j{}^2 \sim inverse - Gamma(b, 1), \qquad j = 1,2, \dots, k,$$

$$\sigma^2 \sim {}^{1}/_{\sigma^2}$$

In this chapter, we use this class of priors in left-censored regression. Under the above prior distribution, we develop a new Gibbs sampler for Bayesian left censored regression.

2.2 Methods

2.2.1 Model hierarchy and prior distributions

Based on the left censored regression model (2.1) and the prior distributions for the regression parameters in left censored regression model, we formulate our hierarchical representation as follows:

$$y_i = \begin{cases} y_i^* & , \quad y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \geq 0 \\ c & , \quad y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i < 0 \end{cases}, \qquad i = 1, \dots, n,$$

$$\begin{aligned} & y_i^* | \beta, \sigma^2 \sim N(\boldsymbol{x}_i'\boldsymbol{\beta}, \sigma^2), && i = 1,2, \dots, n, \\ & \beta_j | \sigma^2, \lambda_j^2, s_j^2, \sim \left(0, \sigma^2 \lambda_j^2 s_j^2\right), && j = 1,2, \dots, k, \\ & \lambda_j^2 \sim Gamma(a, 1), && j = 1,2, \dots, k, \\ & s_j^2 \sim inverse - Gamma(b, 1), && j = 1,2, \dots, k, \end{aligned} \tag{2.6}$$

$$\sigma^2 \sim {}^{1}/_{\sigma^2}$$

2.2.2 Full conditional posterior distributions

Following Bai and Ghosh (2018), the full conditional posterior distributions are given as

- Updating y_i^*, $\quad i = 1,2, \dots n.$

Let $\delta(y_i)$ denotes to a degenerate distribution, then y_i^* has a conditional distribution given by:

$$y_i^* | y_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda}^2, \boldsymbol{s}^2, \sigma^2 \sim \begin{cases} \delta(y_i), & if\ y_i > c, \\ N(\boldsymbol{x}_i'\boldsymbol{\beta}, \sigma^2) I(y_i^* \leq c), & otherwise, \end{cases}$$

where, $\boldsymbol{\lambda}^2 = (\lambda_1^2, \lambda_2^2, \dots, \lambda_k^2$ and $\boldsymbol{s}^2 = (s_1^2, s_2^2, \dots, s_k^2)$.

- Updating $\boldsymbol{\beta}$

The full conditional posterior distribution of β is defined as:

$$\boldsymbol{\beta}|y_i{}^*, X, \boldsymbol{\lambda^2}, \boldsymbol{s^2}, \sigma^2 \propto (y_i{}^*|X, \boldsymbol{\beta}, \sigma^2)\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\lambda^2}, \boldsymbol{s^2})$$

$$\boldsymbol{\beta}|y_i{}^*, X, \boldsymbol{\lambda^2}, \boldsymbol{s^2}, \sigma^2 \propto e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}^*-X\boldsymbol{\beta})'(\boldsymbol{y}^*-X\boldsymbol{\beta})} e^{-\frac{\boldsymbol{\beta}^2}{2\boldsymbol{\lambda}^2\sigma^2\boldsymbol{s}^2}}$$

$$\boldsymbol{\beta}|y_i{}^*, X, \boldsymbol{\lambda^2}, \boldsymbol{s^2}, \sigma^2 \propto e^{--\frac{1}{2\sigma^2}(\boldsymbol{y}^*-X\boldsymbol{\beta})'(\boldsymbol{y}^*-X\boldsymbol{\beta})} - \frac{1}{2\sigma^2}\boldsymbol{\beta}'D^{-1}\boldsymbol{\beta}$$

$$\boldsymbol{\beta}|y_i{}^*, X, \boldsymbol{\lambda^2}, \boldsymbol{s^2}, \sigma^2 \propto e^{-\frac{1}{2\sigma^2}\{(\boldsymbol{\beta}'(X'X)\boldsymbol{\beta}-2\boldsymbol{y}^*X\boldsymbol{\beta}+\boldsymbol{y}^{*\prime}\boldsymbol{y}^*)+\boldsymbol{\beta}'D^{-1}\boldsymbol{\beta}\}}$$

$$\boldsymbol{\beta}|y_i{}^*, X, \boldsymbol{\lambda^2}, \boldsymbol{s^2}, \sigma^2 \propto e^{-\frac{1}{2\sigma^2}\{(\boldsymbol{\beta}'(X'X+D^{-1})\boldsymbol{\beta}-2\boldsymbol{y}^*X\beta+\boldsymbol{y}^{*\prime}\boldsymbol{y}^*)\}}.$$

Now let,

$$C = X'X + D^{-1},$$

then we have,

$$\boldsymbol{\beta}|y_i{}^*, X, \boldsymbol{\lambda^2}, \boldsymbol{s^2}, \sigma^2 \propto e^{-\frac{1}{2\sigma^2}\{\boldsymbol{\beta}'C\boldsymbol{\beta}-2\boldsymbol{y}^*X\boldsymbol{\beta}+\boldsymbol{y}^{*\prime}\boldsymbol{y}^*\}}.$$

Now let,

$$(\boldsymbol{\beta} - C^{-1}X'\boldsymbol{y}^*)'C(\boldsymbol{\beta} - C^{-1}X'\boldsymbol{y}^*) = \boldsymbol{\beta}C'\boldsymbol{\beta} - 2\boldsymbol{y}^*X\boldsymbol{\beta} + \boldsymbol{y}^{*\prime}(XC^{-1}X')\boldsymbol{y}^*,$$

then we have,

$$\boldsymbol{\beta}|y_i{}^*, X, \boldsymbol{\lambda^2}, \boldsymbol{s^2}, \sigma^2 \propto e^{-\frac{1}{2\sigma^2}\{(\boldsymbol{\beta}-C^{-1}X'\boldsymbol{y}^*)'C(\boldsymbol{\beta}-C^{-1}X'\boldsymbol{y}^*)+\boldsymbol{y}^{*\prime}(I_n-XC^{-1}X')\boldsymbol{y}^*\}},$$

$$\boldsymbol{\beta}|y_i{}^*, X, \boldsymbol{\lambda^2}, \boldsymbol{s^2}, \sigma^2 \propto e^{-\frac{1}{2\sigma^2}\{(\beta-C^{-1}X'\boldsymbol{y}^*)'C(\beta-C^{-1}X'\boldsymbol{y}^*)\}}. \qquad (2.7)$$

By recalling the multivariate normal distribution, we found that (2.7) represent $\beta \sim N_k(\mu, \Sigma)$,

where

$$\mu = (X'X + \Omega^{-1})^{-1}X'\boldsymbol{y}^*,$$

and

$$\Sigma = \sigma^2(X'X + \Omega^{-1})^{-1},$$

where

$$\Omega = \left(\lambda_1{}^2 s_1{}^2, \dots, \lambda_k{}^2 s_k{}^2\right)'.$$

- Updating $\lambda_j{}^2$: The full conditional posterior distribution of $\lambda_j{}^2$is defined as:

$$\lambda_j{}^2|\beta, \sigma^2, s_j{}^2 \propto \left(\beta|\lambda_j{}^2, s_j{}^2, \sigma^2\right)\pi(\lambda^2)$$

$$\lambda_j{}^2|\boldsymbol{\beta}, \sigma^2, s_j{}^2 \propto (2\sigma^2\lambda^2)^{-\frac{1}{2}} exp\left\{-\frac{\beta_j{}^2}{2\sigma^2\lambda_j{}^2 s_j{}^2}\right\}\frac{1}{\Gamma(a)}\left(\lambda_j\right)^{a-1} e^{-\lambda_j{}^2}$$

$$\lambda_j{}^2|\boldsymbol{\beta}, \sigma^2, s_j{}^2 \propto \left(\lambda_j{}^2\right)^{-1/2}\left(\lambda_j\right)^{a-1} exp\left\{-\frac{\beta_j{}^2}{2\sigma^2\lambda_j{}^2 s_j{}^2} - \lambda_j{}^2\right\}$$

$$\propto \left(\lambda_j\right)^{(a-1/2)-1} exp\left\{-\frac{1}{2}\left(\frac{\beta_j{}^2/\sigma^2 s_j{}^2}{\lambda_j{}^2}\right) + \lambda_j{}^2\right\} \qquad (2.8)$$

From recall that the Generalized Inverse Gaussian (GIG)

$$\lambda_j{}^2|s_j{}^2, \beta_j, \sigma^2 \sim GIG\left(\frac{\beta_j{}^2}{\sigma^2 s_j{}^2}, 2, a - \frac{1}{2}\right) \qquad (2.9)$$

- Updating $s_j{}^2$: The conditional posterior distribution of $s_j{}^2$ is defined as the follows:

$$s_j{}^2|\beta_j, \lambda_j{}^2, \sigma^2 \propto \left(\beta_j|\sigma^2, \lambda_j{}^2, s_j{}^2\right)\pi\left(s_j{}^2\right),$$

$$s_j{}^2|\beta_j, \lambda_j{}^2 \propto \left(\frac{1}{s_j{}^2}\right)^{\frac{1}{2}} e^{-\frac{\beta_j{}^2}{2\sigma^2\lambda_j{}^2 s_j{}^2}} \times \frac{1}{\Gamma(b)}\left(s_j{}^2\right)^{-b-1} e^{-\frac{1}{s_j{}^2}},$$

$$s_j{}^2|\beta_j, \lambda_j{}^2 \propto e^{-\frac{\beta_j{}^2}{2\sigma^2\lambda_j{}^2 s_j{}^2} - \frac{1}{s_j{}^2}}\left(s_j{}^2\right)^{-b-1-\frac{1}{2}},$$

$$s_j{}^2|\beta_j,\lambda_j{}^2 \propto e^{-\frac{1}{s_j{}^2}\left[\frac{\beta_j{}^2}{2\sigma^2\lambda_j{}^2}+1\right]}\left(s_j{}^2\right)^{-\left(b+\frac{1}{2}\right)-1}. \qquad (2.10)$$

Recall that the Inverse Gamma distribution, then (2.10) represented Inverse Gamma posterior distribution for $s_j{}^2$ with shape parameter $\left(b+\frac{1}{2}\right)$, and scale parameter $\left[\frac{\beta_j{}^2}{2\sigma^2\lambda_j{}^2}+1\right]$,

$$s_j{}^2 \sim IG\left(b+\frac{1}{2},\frac{\beta_j{}^2}{2\sigma^2\lambda_j{}^2}+1\right) \qquad (2.11)$$

- Updating σ^2: The conditional posterior distribution of σ^2 is defined as the follows:

$$\sigma^2|y_i{}^*,X,\beta \propto \prod_{i=1}^{n}(y_i{}^*|X,\beta,\sigma^2)\prod_{j=1}^{k}\left(\beta|\sigma^2,\lambda_j{}^2,s_j{}^2\right)\pi(\sigma^2)$$

$$\sigma^2|y_i{}^*,X,\beta \propto (\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma_j{}^2}(\boldsymbol{y}^*-X\boldsymbol{\beta})'(\boldsymbol{y}^*-X\boldsymbol{\beta})}(\sigma^2)^{-\frac{1}{2}}e^{-\frac{\beta_j{}^2}{2\sigma_j{}^2\lambda_j{}^2 s_j{}^2}}(\sigma^2)^{-1}$$

$$\propto (\sigma^2)^{-\frac{n-3}{2}}exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}^*-X\boldsymbol{\beta})'(\boldsymbol{y}^*-X\boldsymbol{\beta})+\boldsymbol{\beta}'D^{-1}\boldsymbol{\beta}+d\right\}, \qquad (2.12)$$

where $D = diag\left(\lambda_1{}^2 s_1{}^2,\dots,\lambda_k{}^2 s_k{}^2\right)$,

From (2.12) recall that the Inverse Gamma distribution and hence ,the posterior distribution of σ^2 is defined by,

$$\sigma^2 \sim inverse\ gamma\ \left(\frac{n-3}{2},\frac{1}{2}(\boldsymbol{y}^*-X\boldsymbol{\beta})'(\boldsymbol{y}^*-X\boldsymbol{\beta})+\boldsymbol{\beta}'D^{-1}\boldsymbol{\beta}+d\right)$$

Chapter Three

Extensions

3.1 Bayesian right censored Regression

In this section, we propose a simple Bayesian variable selection method for right censored regression model. Our method is based on the hierarchical model in chapter two. Right censored regression is a technique used when the response (dependent) variable is censored from the right. It is a very common model which occurs when a subject leaves the study before an event occurs, or the study ends before the event has occurred. Thus, the response variable is partially known. For example, suppose an experiment is conducted to measure the effect of a particular type of medication on mortality. Suppose it is known to us that the age of the individual (subject) at death is not less than 80 years. Right censored data occurs if the individual withdraws from the study or currently alive at the age of 80.

The standard Binary regression structure is as follows:

$$y_i = min\{u, y_i^*\} \qquad i = 1, \dots, n,$$

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad , \tag{3.1}$$

where u is a right censored point and y_i is observed response variable for the observation, y_i^* is the latent (unobserved) response variable, ε_i is the error term with $\varepsilon_i \sim N(0, \sigma^2)$, $\boldsymbol{x}_i$ is a vector of explanatory variables, and $\boldsymbol{\beta}$ is a vector of unknown coefficients, for example see, Alhamzawi and Yu (2013), Long and Freese.(2006) and Agresti (2007) for more details about the censored regression.

In practice, many of right censored data have the problem of high-dimensional. In such case, the analysis of these data tends to select a subgroup of predictor variables which are highly effected on the response

variable y_i and remove the irrelevant explanatory variables that do not have an effect upon y_i, this procedure is called the variable selection procedure. As with left censored regression model, choosing the active predictors is important in right censored regression. Excluding active predictors may yield biased estimators whereas including inactive predictors may lead to loss in estimation efficiency. So, the variable selection problem enables the data analyst to choose the more interpretable regression model that provides the best fit for the data. Consequently, when there are $(k > n)$ we found many of works that deal with this property in the data analysis.

Based on the Bayesian hierarchical model in chapter two, we develop a new Bayesian hierarchical model for the right censored regression. Thus, our Bayesian hierarchical model can be written as:

$$y_i = min\{u, {y_i}^*\} \qquad i = 1, \dots, n,$$

$${y_i}^* = \boldsymbol{x}_i{}'\boldsymbol{\beta} + \varepsilon_i \quad ,$$

$$\left.\begin{array}{ll} {y_i}^*|\boldsymbol{\beta}, \sigma^2 \sim N(\boldsymbol{x}_i{}'\boldsymbol{\beta}, \sigma^2), & i = 1,2, \dots, n \\ \beta_j|\sigma^2, {\lambda_j}^2, {s_j}^2 \sim N\ (0, \sigma^2, D), & j = 1,2, \dots, k \\ \qquad D = diag\ (\lambda_1^2 s_1^2, \dots, \lambda_k^2 s_k^2) & \\ {\lambda_j}^2 \sim Gamma(a, 1), & j = 1,2, \dots, k \\ {s_j}^2 \sim inverse - Gamma(b, 1), & j = 1,2, \dots, k \\ \qquad \sigma^2 \sim inverse - Gamma\ (g1, g2). & \end{array}\right\} \quad (3.2)$$

Based on the above Bayesian hierarchical model (3.2), the full conditional posterior of ${y_i}^*$ is given by:

$${y_i}^*|y_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda^2}, \boldsymbol{s^2}, \sigma^2 \sim \begin{cases} \delta(y_i), & if\ y_i < u \\ N(\boldsymbol{x}_i{}'\boldsymbol{\beta}, \sigma^2) I({y_i}^* \geq u), & otherwise \end{cases} \quad (3.3)$$

The full conditional posterior distributions for $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{s}, \sigma^2)$ is the same as in chapter two except that y_i^* is generated from the posterior distribution in (3.3).

3.2 Bayesian Binary regression

Binary regression is a technique used when the response (dependent) variable is categorical. The Binary regression model is a very common model in the sciences fields that represents the problems with binary outcomes, such as the medical trails (negative or positive), consumers where the decision is made based on buying or not buying, and so on. So, since the outcome (response) variable is binary (dichotomous), we cannot directly model data, that is, we can treat the Binary regression model as latent variable regression model. Benoit et al.(2013) considered the standard Binary regression structure as follows:

$$y_i = \begin{cases} 1, & y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \geq 0 \\ 0, & y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i < 0 \end{cases}, \qquad (3.4)$$

where y_i is observed response variable for the ith observation, y_i^* is the latent (unobserved) response variable, ε_i is the error term will, $\varepsilon_i \sim NI_n(0, \sigma^2)$, $\boldsymbol{x}_i$ is a vector of explanatory variables, and $\boldsymbol{\beta}$ is a vector of unknown coefficients, for example see, Alhamzawi and Yu (2013), Long and Freese.(2006) and Agresti (2007) for more details about the binary regression.

In this part, based on the Bayesian hierarchical model in chapter two, we develop a new Bayesian hierarchical model for the Binary regression. Thus, our Bayesian hierarchical model can be written as:

$$y_i = \begin{cases} 1, & y_i{}^* = \boldsymbol{x}_i{}'\boldsymbol{\beta} + \varepsilon_i \geq 0 \\ 0, & y_i{}^* = \boldsymbol{x}_i{}'\boldsymbol{\beta} + \varepsilon_i < 0 \end{cases}$$

$$\left.\begin{array}{ll} y_i{}^* | \boldsymbol{\beta}, \sigma^2 \sim N(\boldsymbol{x}_i{}'\boldsymbol{\beta}, \sigma^2), & i = 1,2,\dots,n, \\ \beta_j | \sigma^2, \lambda_j{}^2, s_j{}^2 \sim N\,(0, \sigma^2, D), & j = 1,2,\dots,k, \\ \qquad D = diag\,(\lambda_1^2 s_1^2, \dots, \lambda_k^2 s_k^2) & \\ \lambda_j{}^2 \sim Gamma(a, 1), & j = 1,2,\dots,k, \\ s_j{}^2 \sim inverse - Gamma(b, 1), & j = 1,2,\dots,k, \\ \qquad \sigma^2 \sim inverse - Gamma\,(c, d). & \end{array}\right\} \quad (3.5)$$

Based on the above Bayesian hierarchical model (3.3), the full conditional posterior of $y_i{}^*$ is given by

$$y_i{}^* | y_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda^2}, \boldsymbol{s^2}, \sigma^2 \sim \begin{cases} N(\boldsymbol{x}_i'\boldsymbol{\beta}, \sigma^2) \prod_{i=1}^n I(y_i{}^* \geq 0)\ if\ \boldsymbol{y_i} = 1 \\ N(\boldsymbol{x}_i'\boldsymbol{\beta}, \sigma^2) \prod_{i=1}^n I(y_i{}^* < 0)\ if\ \boldsymbol{y_i} = 0 \end{cases}$$

The full conditional posterior distributions for $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{s}, \sigma^2)$ is the same as in chapter two except that $y_i{}^*$ is generated from the posterior distribution in (3.5).

Chapter Four

Simulation Studies

4 Simulation Studies

For the simulated studies, we consider the median of mean squared errors (MMSE) where the median is taken over the 150 replications. In each replication, we generate a training set with 100 observations and a testing set with 200 observations. Models are fitted on the training set and MSE's are calculated on the test set. Methods in the comparison include: Tobit regression (TR), Lasso tobit regression (LTR), Bayesian Lasso tobit regression (BLTR) and our proposed method (referred to as `NewBLR').

4.1 Simulation 1

In this simulation, the data are simulated from the following model

$$y_i = \max\{c = 0, y_i{}^*\}, \quad i = 1,2, \dots, n,$$

$$y_i{}^* = \boldsymbol{x_i}'\boldsymbol{\beta} + \varepsilon_i \tag{4.1}$$

We consider a very sparse model with a strong level of correlation ($\rho = 0.95$). We set the true regression coefficients, including the intercept term, $\beta = (1,2,0,0,0,0,0,0,0)$, $\sigma^2 = \{1,2,3,4,5\}$ and $e_i \sim N(0, \sigma^2)$. The predictors' matrix X is simulated from the multivariate Gaussian distribution with mean 0, variance 1 and pairwise correlations between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ equal to ρ.

4.2 Simulation 2

In this simulation, we consider the sparse case = $(1,0.25,0.25,0.25,0.25,0.25,0.25,0.25)'$, leaving other setup exactly same as model simulation 1.

4.3 Simulation 3

In this simulation, we consider the dense case = $(1,0.5,0.5,0.5,0.5,0.5,0.5,0.5)'$ and $(\rho = 0.5)$, leaving other setups exactly the same as model Simulation 1.

4.4 Simulation 4

In this simulation, we consider the dense case = $(1, 3, 1.5, 0, 0, 2, 0, 0, 0)'$ and $(\rho = 0.95)$, leaving other setups exactly the same as model Simulation 1.

Table 1: Median mean squared error (MMSE) based on 100 replications for simulation 1. In the parentheses are standard deviations of the MSEs.

n	σ^2	Lasso	aLasso	NewBLR
25	1	1.647 (0.392)	1.502 (0.453)	0.876 (0.333)
25	2	2.247 (0.610)	2.079 (0.683)	1.213 (0.459)
25	3	1.819 (0.646)	1.692 (0.523)	1.077 (0.545)
25	4	2.003 (0.679)	2.042 (0.761)	1.147 (0.729)
25	5	2.119 (1.123)	2.073 (1.236)	1.252 (0.619)
50	1	1.564 (0.278)	1.562 (0.283)	0.854 (0.271)
50	2	1.594 (0.337)	1.578 (0.367)	0.984 (0.284)
50	3	1.733 (0.505)	1.669 (0.584)	1.078 (0.378)
50	4	1.769 (0.415)	1.751 (0.412)	1.083 (0.465)
50	5	1.887 (0.341)	1.850 (0.424)	1.229 (0.422)
100	1	1.532 (0.194)	1.514 (0.187)	0.713 (0.203)
100	2	1.551 (0.211)	1.525 (0.227)	0.892 (0.226)
100	3	1.680 (0.252)	1.645 (0.263)	1.015 (0.155)

100	4	1.700 (0.288)	1.655 (0.305)	1.142 (0.326)
100	5	1.750 (0.309)	1.750 (0.394)	1.215 (0.276)
200	1	1.508 (0.146)	1.468 (0.151)	0.674 (0.164)
200	2	1.529 (0.172)	1.525 (0.171)	0.815 (0.133)
200	3	1.606 (0.291)	1.576 (0.286)	0.932 (0.237)
200	4	1.653 (0.197)	1.653 (0.202)	1.094 (0.183)
200	5	1.639 (0.258)	1.646 (0.293)	1.122 (0.256)

Figure 1. Trace plots of tobit regression parameters for Simulation 1 when $n = 25$ and $\sigma^2 = 5$.

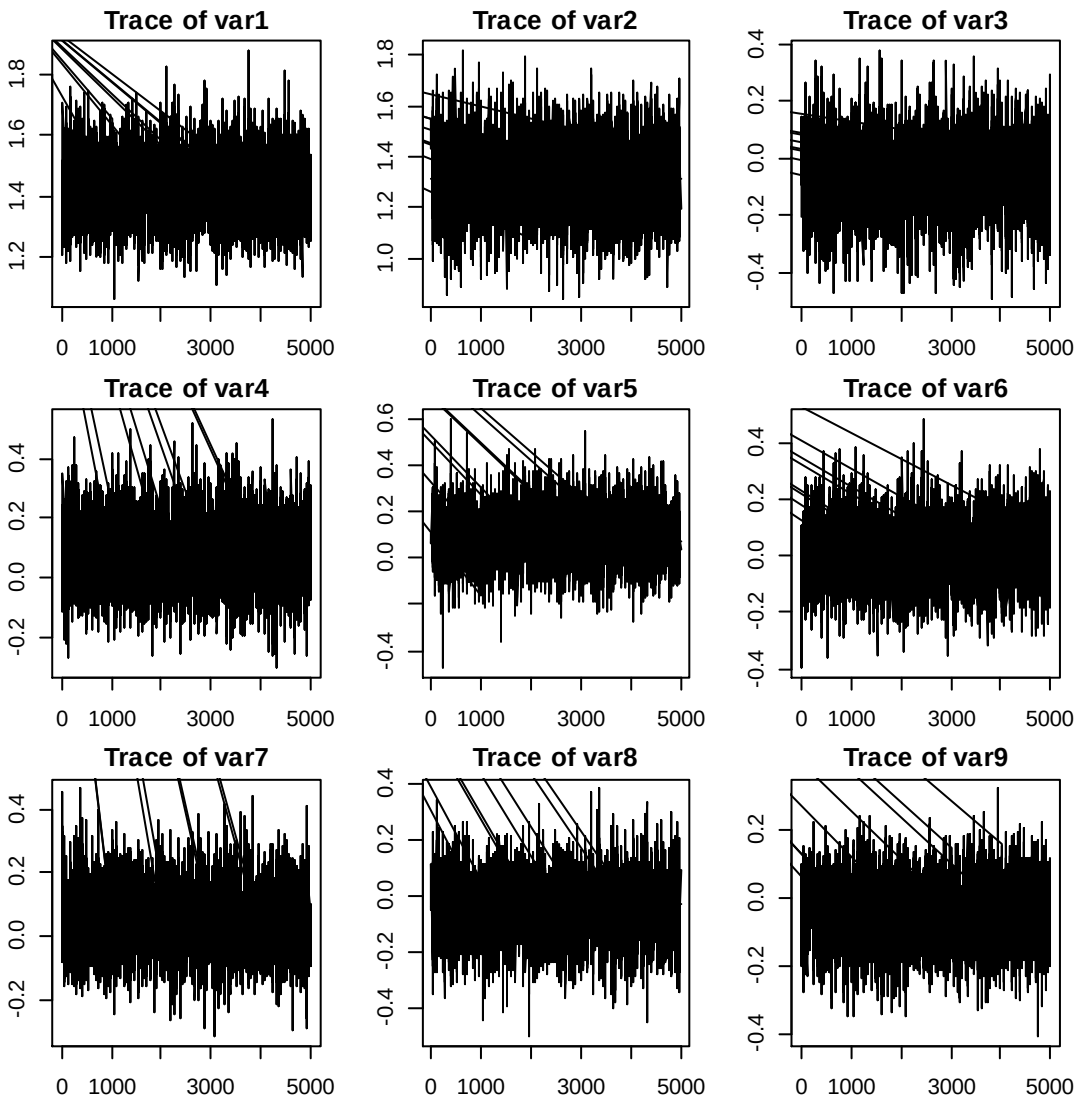


Figure 2. Histograms based on posterior samples of the parameters for Simulation 1 when $n = 25$ and $\sigma^2 = 5$.

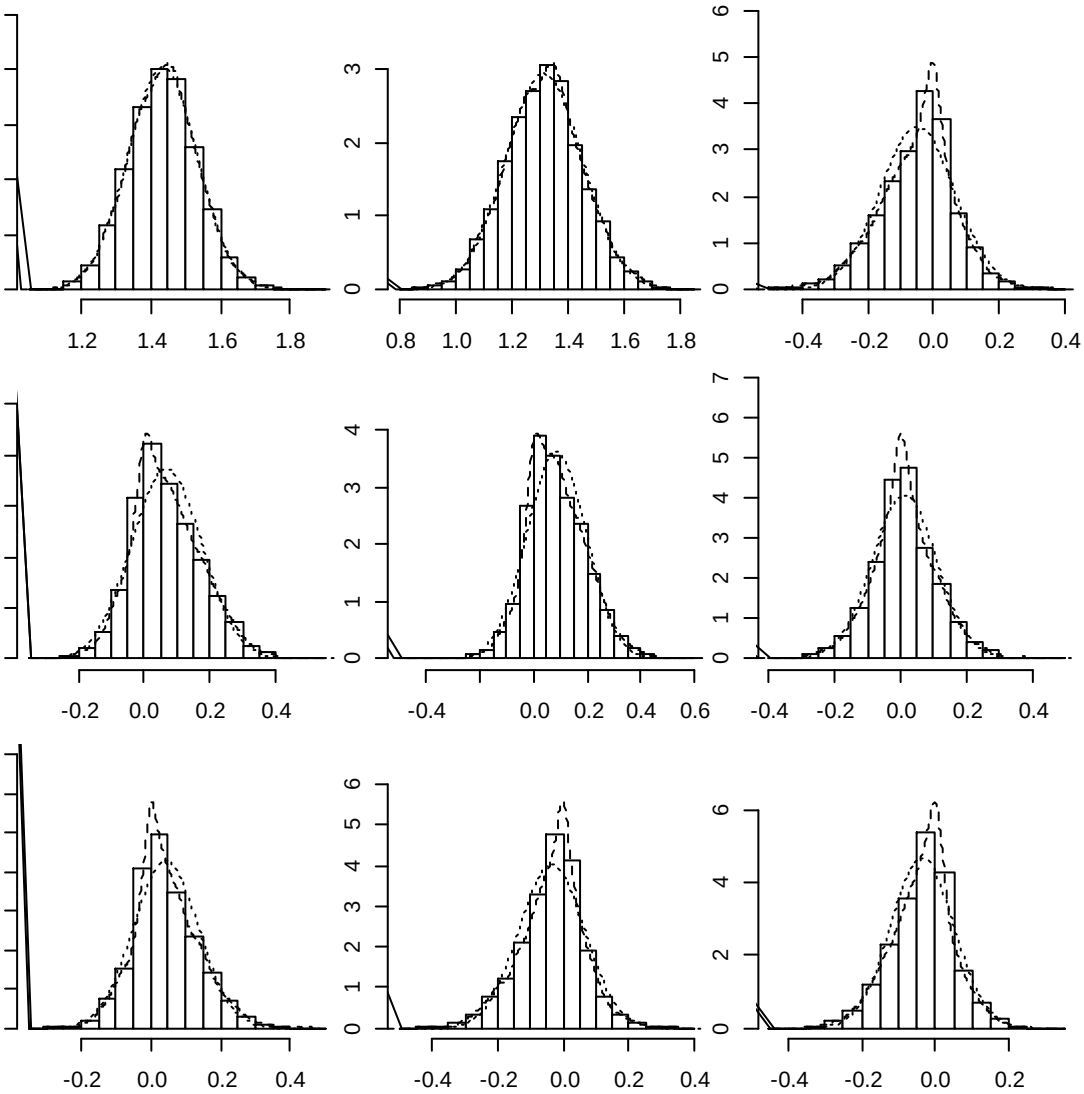


Table 2. Median mean squared error (MMSE) based on 100 replications for simulation 2. In the parentheses are standard deviations of the MSEs.

n	σ^2	**Lasso**	**aLasso**	**NewBLR**
25	1	1.390 (0.212)	1.420 (0.321)	0.392 (0.184)
25	2	1.357 (0.350)	1.413 (0.274)	0.513 (0.306)
25	3	1.364 (0.566)	1.457 (0.589)	0.567 (0.488)
25	4	1.440 (0.365)	1.336 (0.586)	0.754 (0.347)
25	5	1.670 (0.601)	1.704 (0.714)	0.820 (0.604)
50	1	1.205 (0.175)	1.236 (0.166)	0.274 (0.133)
50	2	1.369 (0.238)	1.374 (0.275)	0.422 (0.194)
50	3	1.397 (0.198)	1.403 (0.225)	0.567 (0.187)
50	4	1.496 (0.310)	1.436 (0.360)	0.792 (0.363)
50	5	1.448 (0.486)	1.510 (0.584)	0.820 (0.301)
100	1	1.239 (0.152)	1.234 (0.141)	0.302 (0.104)
100	2	1.266 (0.122)	1.305 (0.124)	0.351 (0.102)
100	3	1.323 (0.129)	1.347 (0.138)	0.536 (0.131)
100	4	1.458 (0.241)	1.429 (0.229)	0.672 (0.217)
100	5	1.374 (0.173)	1.389 (0.194)	0.694 (0.185)
200	1	1.223 (0.090)	1.207 (0.095)	0.311 (0.062)
200	2	1.238 (0.107)	1.240 (0.106)	0.443 (0.083)
200	3	1.263 (0.096)	1.283 (0.102)	0.490 (0.093)
200	4	1.310 (0.150)	1.363 (0.185)	0.582 (0.189)
200	5	1.292 (0.141)	1.284 (0.144)	0.692 .160)

Figure 3. Trace plots of tobit regression parameters for Simulation 2 when $n = 25$ and $\sigma^2 = 5$.

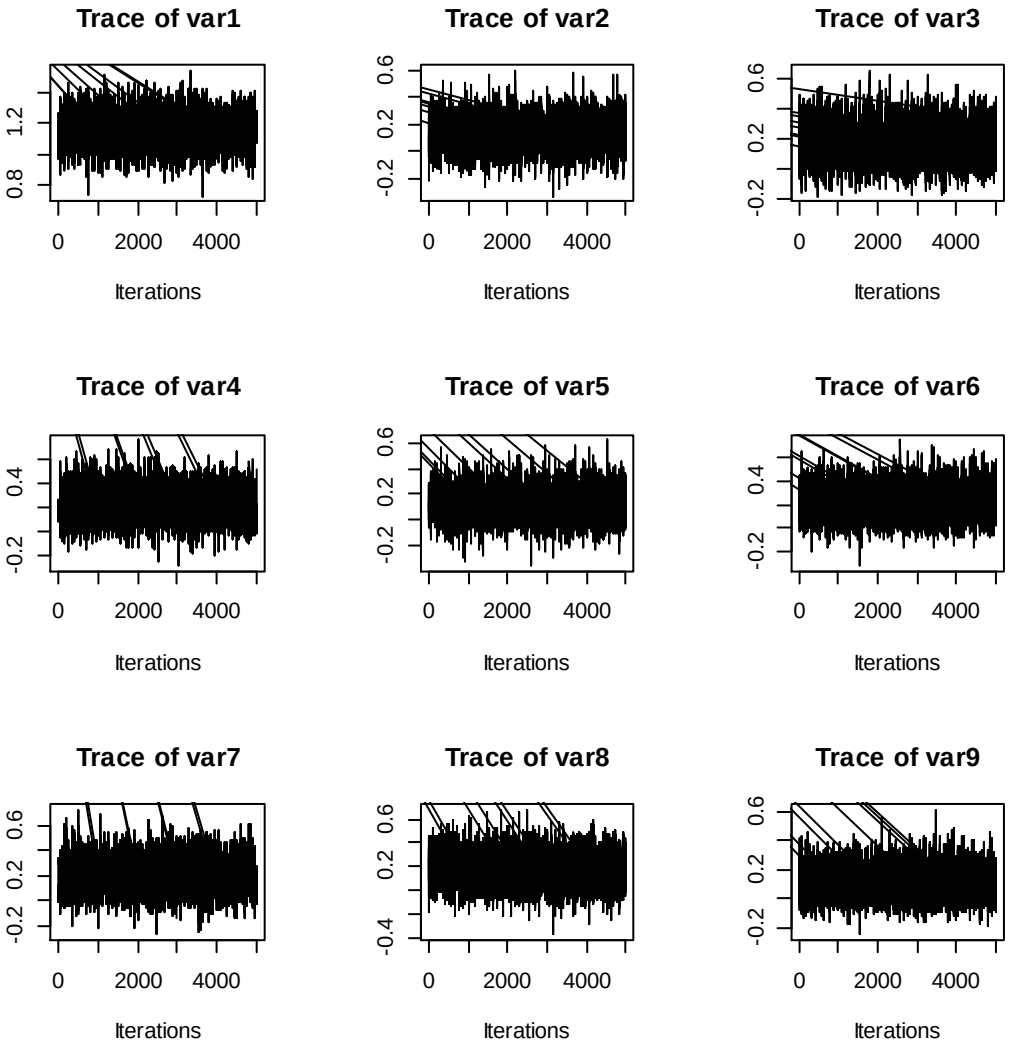


Figure 4. Histograms based on posterior samples of the parameters for Simulation 2 when $n = 25$ and $\sigma^2 = 5$.

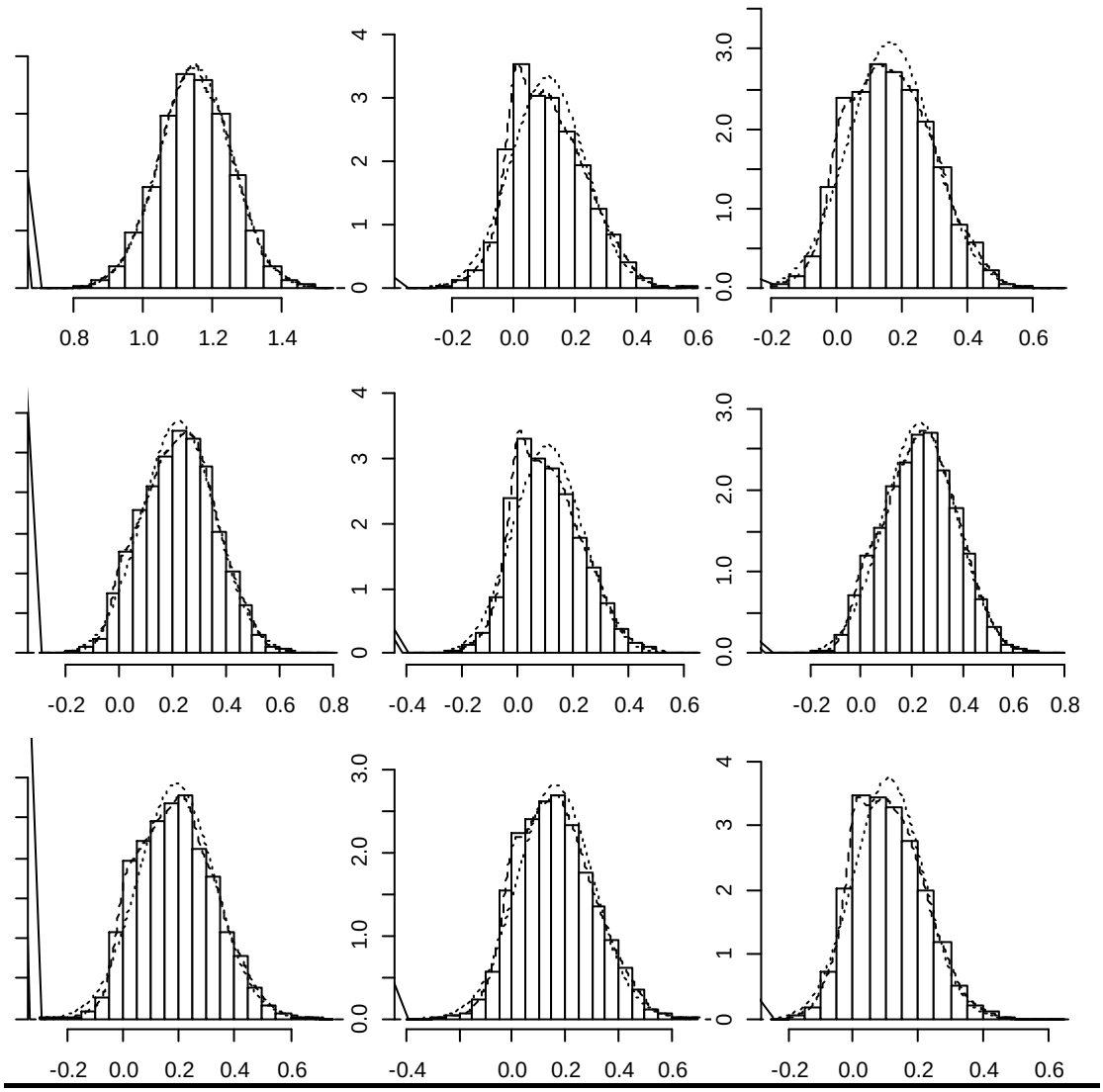


Table 3. Median mean squared error (MMSE) based on 100 replications for simulation 3. In the parentheses are standard deviations of the MSEs**.**

n	σ^2	Lasso	aLasso	NewBLR
25	1	1.647 (0.392)	1.502 (0.453)	0.876 (0.333)
25	2	2.247 (0.610)	2.079 (0.683)	1.213 (0.459)
25	3	1.819 (0.646)	1.692 (0.523)	1.077 (0.545)
25	4	2.003 (0.679)	2.042 (0.761)	1.147 (0.729)
25	5	2.119 (1.123)	2.073 (1.236)	1.252 (0.619)
50	1	1.564 (0.278)	1.562 (0.283)	0.854 (0.271)
50	2	1.594 (0.337)	1.578 (0.367)	0.984 (0.284)
50	3	1.733 (0.505)	1.669 (0.584)	1.078 (0.378)
50	4	1.769 (0.415)	1.751 (0.412)	1.083 (0.465)
50	5	1.887 (0.341)	1.850 (0.424)	1.229 (0.422)
100	1	1.532 (0.194)	1.514 (0.187)	0.713 (0.203)
100	2	1.551 (0.211)	1.525 (0.227)	0.892 (0.226)
100	3	1.680 (0.252)	1.645 (0.263)	1.015 (0.155)
100	4	1.700 (0.288)	1.655 (0.305)	1.142 (0.326)
100	5	1.750 (0.309)	1.750 (0.394)	1.215 (0.276)
200	1	1.508 (0.146)	1.468 (0.151)	0.674 (0.164)
200	2	1.529 (0.172)	1.525 (0.171)	0.815 (0.133)
200	3	1.606 (0.291)	1.576 (0.286)	0.932 (0.237)
200	4	1.653 (0.197)	1.653 (0.202)	1.094 (0.183)
200	5	1.639 (0.258)	1.646 (0.293)	1.122 (0.256)

Figure 5. Trace plots of tobit regression parameters for Simulation 3 when $n = 25$ and $\sigma^2 = 5$.

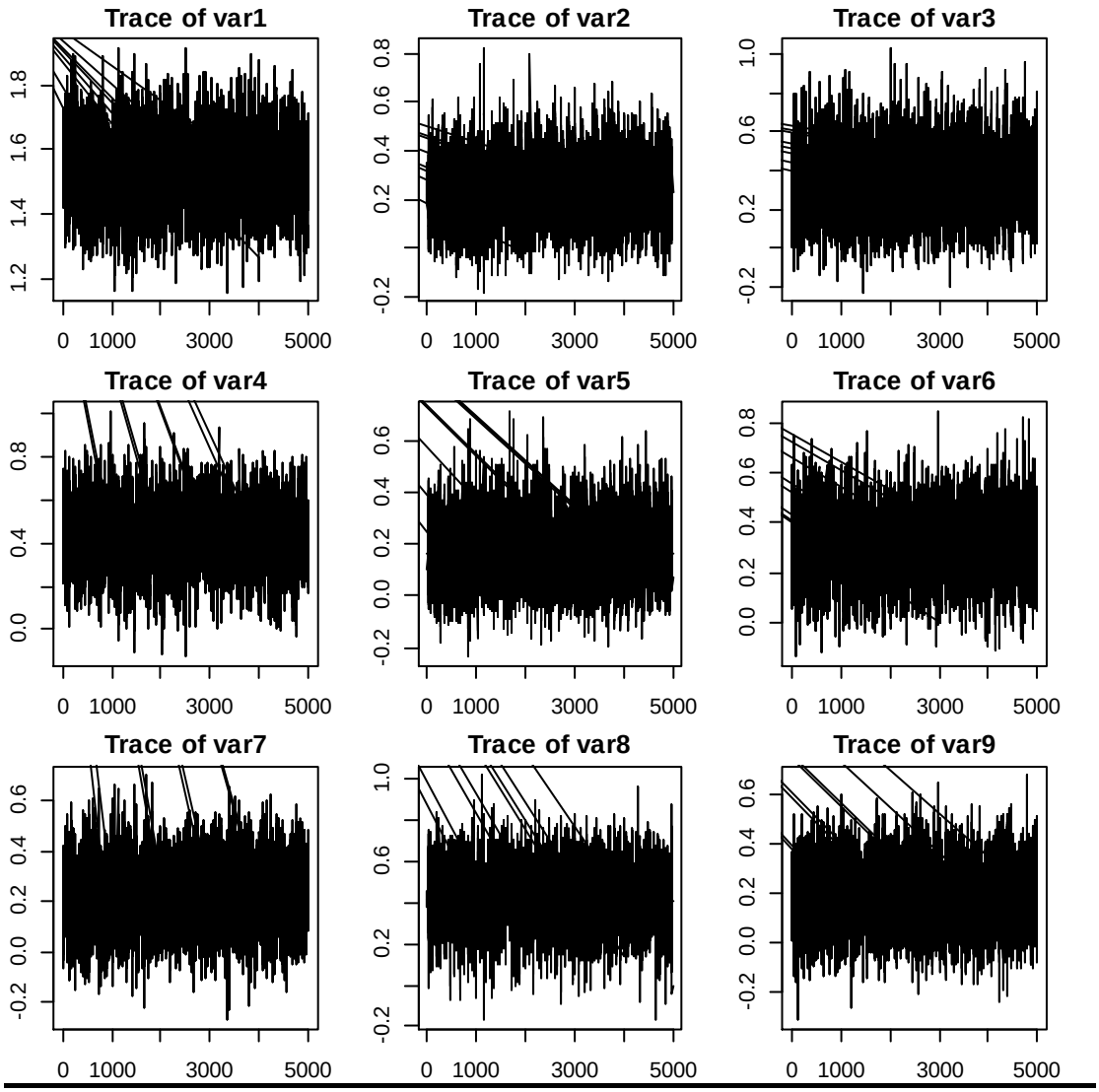


Figure 6. Histograms based on posterior samples of the parameters for Simulation 3 when $n = 25$ and $\sigma^2 = 5$.

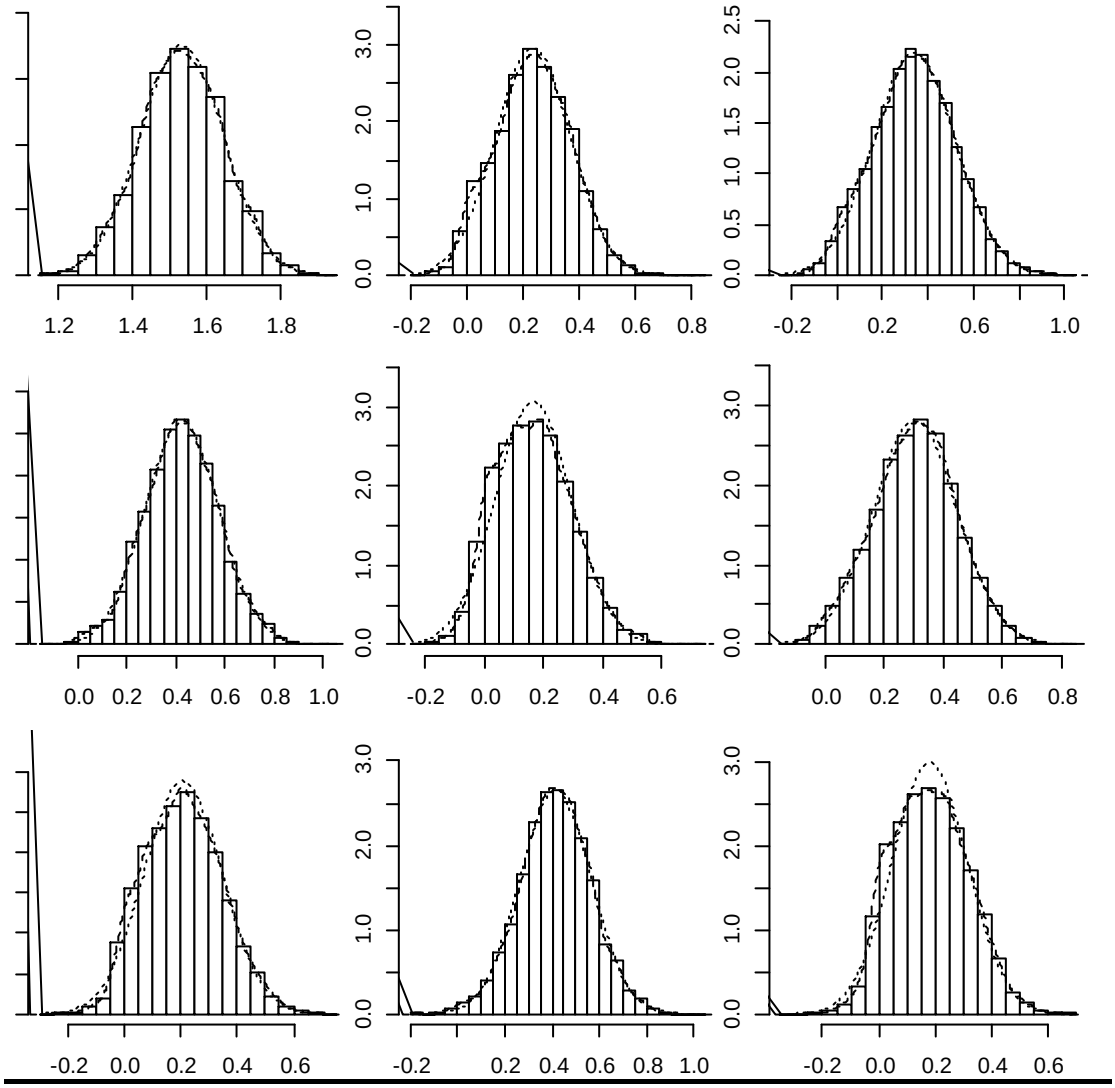


Table 4. Median mean squared error (MMSE) based on 100 replications for simulation 4. In the parentheses are standard deviations of the MSEs**.**

n	σ^2	Lasso	aLasso	NewBLR
25	1	10.642 (4.256)	9.680 (4.475)	9.133 (4.264)
25	2	8.190 (4.231)	8.884 (5.030)	8.821 (3.578)
25	3	9.715 (5.065)	10.143 (5.936)	9.420 (4.580)
25	4	10.561 (4.302)	10.048 (3.389)	9.095 (4.807)
25	5	10.610 (3.609)	11.106 (4.596)	10.052 (3.889)
50	1	9.844 (3.259)	8.792 (3.708)	8.046 (2.421)
50	2	9.425 (3.618)	9.752 (3.584)	8.881 (3.304)
50	3	8.728 (2.550)	8.724 (2.327)	7.237 (2.488)
50	4	10.041 (2.829)	10.991 (3.011)	10.706 (2.701)
50	5	9.712 (2.093)	9.353 (2.240)	9.694 (2.971)
100	1	8.573 (1.848)	8.437 (1.808)	8.871 (2.124)
100	2	8.912 (2.160)	8.951 (2.048)	8.613 (2.009)
100	3	8.772 (2.473)	9.067 (2.420)	8.106 (3.019)
100	4	7.783 (2.240)	7.537 (2.336)	7.580 (2.937)
100	5	10.268 (2.293)	10.333 (2.419)	10.944 (2.763)
200	1	8.038 (1.885)	7.965 (1.863)	7.228 (1.711)
200	2	9.011(1.022)	8.927 (1.437)	8.462 (1.755)
200	3	8.933 (0.837)	8.809 (0.878)	8.850 (1.751)
200	4	9.318 (1.989)	9.127 (1.954)	9.017 (2.210)
200	5	8.527 (1.410)	8.264 (1.464)	8.569 (1.892)

Figure 7. Trace plots based on posterior samples of the parameters for Simulation 4 when $n = 25$ and $\sigma^2 = 5$.

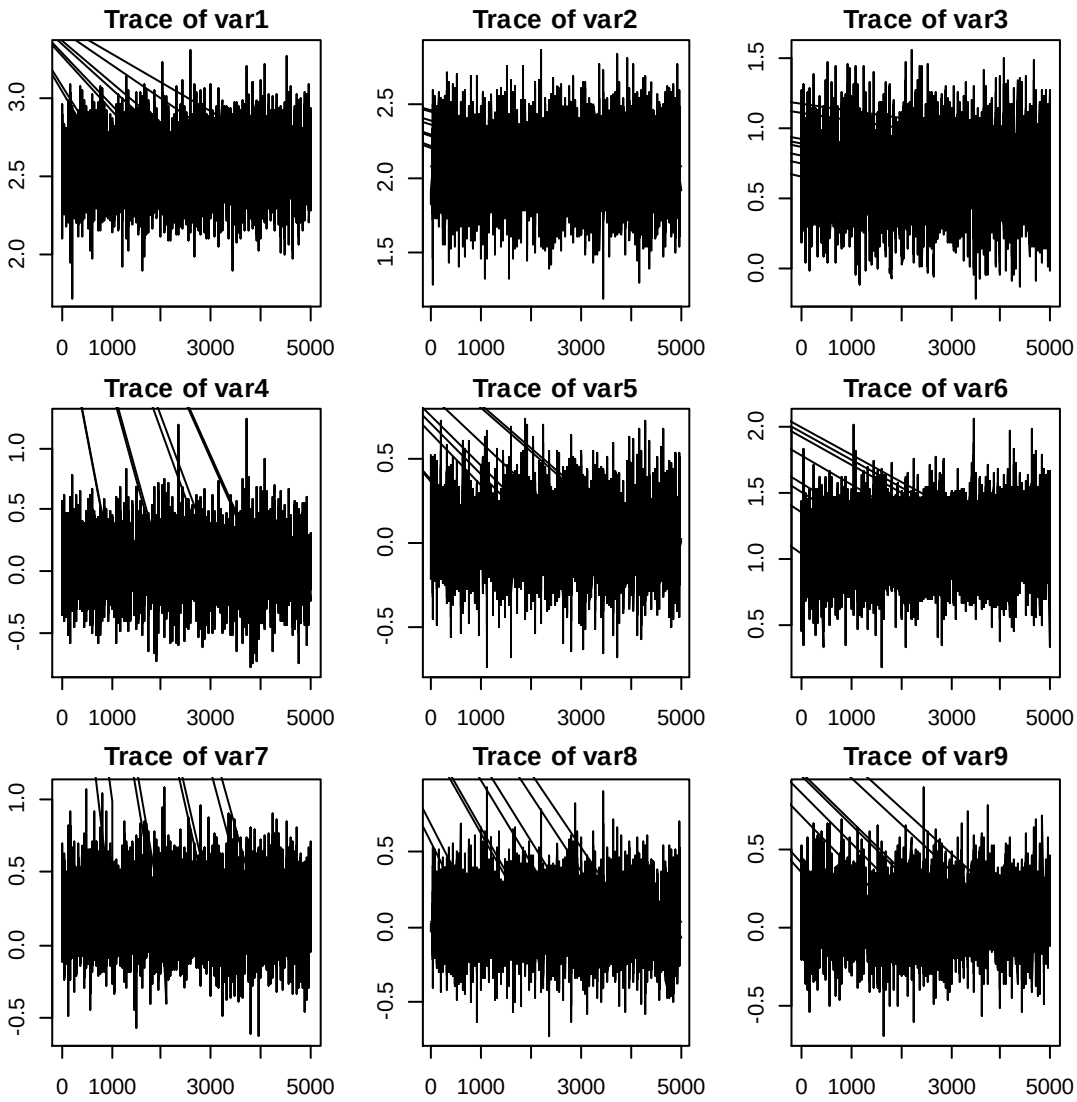
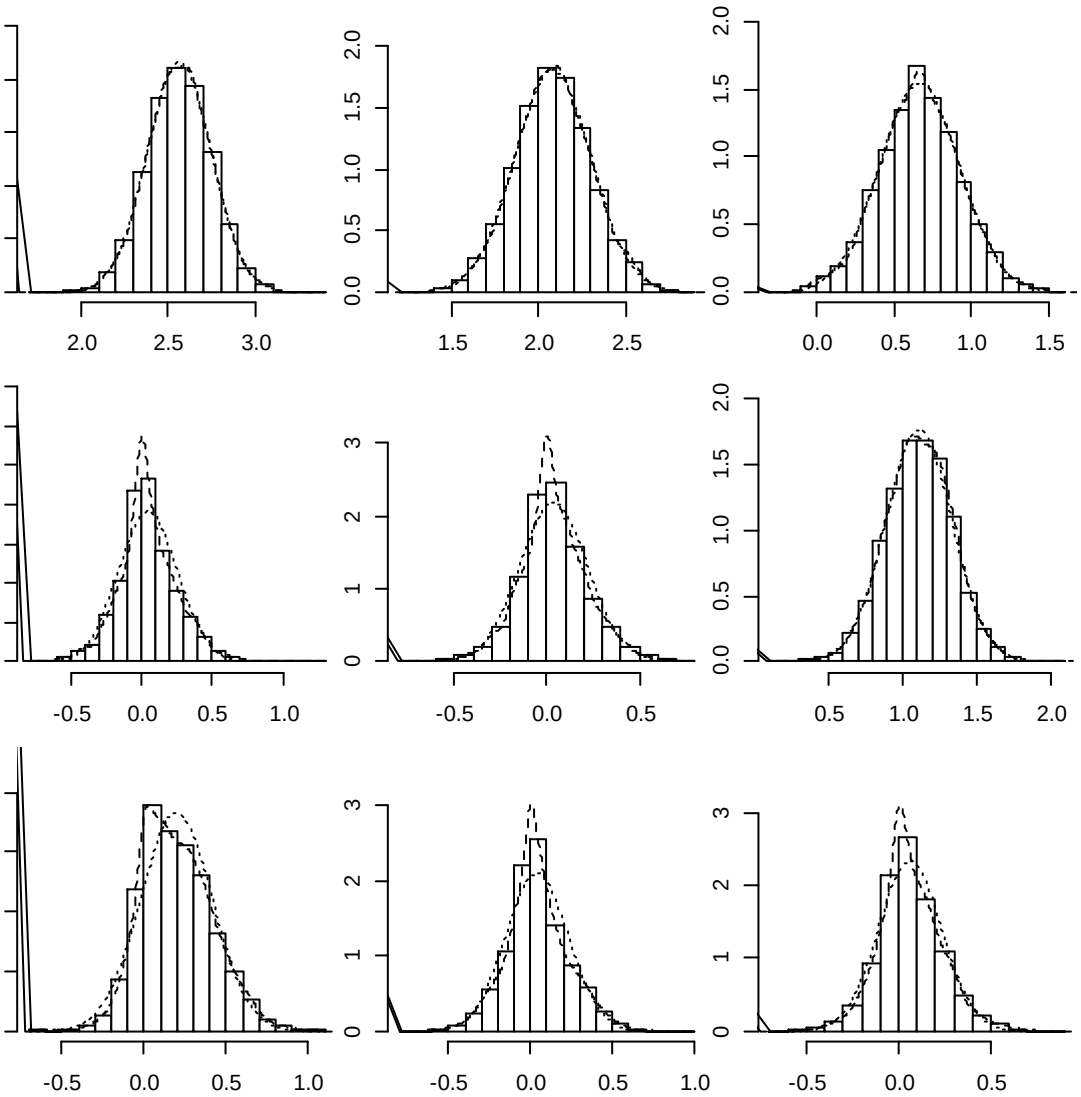


Figure 8. Histograms based on posterior samples of the parameters for Simulation 4 when $n = 25$ and $\sigma^2 = 5$.



Summary statistics of median mean squared error (MMSE) based on 100 replications for each simulation study are reported.

Simulation studies are summarized in Tables 1, 2, 3 and 4 which clearly suggest that the new Bayesian regression method for left censored data (NewBLR) outperforms the other methods across all simulation studies. We can observe that the NewBLR produces the smallest MMSE. These results show that the NewBLR exhibits promising performance in terms prediction accuracy**.** The mixing of an MCMC chain in Figures 1, 3, 5, and 7 show how rapidly our MCMC algorithm converges to the stationary distribution. Trace plot shows that the algorithm has a very good mixing property. The posterior histograms in Figures 2, 4, 6, and 8 reveal that the conditional posterior distributions are in fact the desired distributions.

Chapter Five

Real Data

5.1 Introduction

In this section of application chapter, We will apply our proposed methods to real data after we had shown the advantages of these methods in the simulation study in the previous chapter, and then we will analyze them. These data represent the factors affecting the change in glomerular filtration levels that determine the rate of renal failure. Before defining the explanatory variables, this disease must be mentioned briefly.

Kidney failure is defined as the inability of the kidneys to filter toxins and waste products from the blood. This disease is characterized by a decrease in the glomerular filtration rate (eGFR) to about 20-50% of the normal limit, which leads to a disturbance in the volume of fluid in the body leading to the formation of edema, excessive hyperkalemia, gastrointestinal anemia and vascular disturbance of the heart. Kidney failure is revealed by a high level of creatinine in the blood, a difference in the level of acidic body fluids, calcium, potassium, phosphorous, loss of protein in the urine, and sometimes a delay in the healing of broken bones and other long-term problems that have significant implications for other diseases.

5.2 Real Data

We use a sample of 100 randomly selected patients on 9 variables from the Medical Alaietimad Laboratory in the city of Kut (Iraq) to measure Estimated Glomerular Filtration Rate (eGFR). The response variable is the change in levels of eGFR. The other eight variables are explanatory variables as follows: gender (x_1), age (x_2), percentage of urea in the blood (x_3), the percentage of creatinine in the blood (x_4), calcium level in the blood (x_5), the percentage of potassium in the blood

(x_6), the percentage of sodium in the blood (x_7) and the percentage of phosphate in the blood (x_8).

In Table 4, we compare the Mean squared prediction errors by using our proposed method to those obtained using the Lasso and adaptive Lasso. It can be seen that the new method outperforms both Lasso and aLasso in terms of Mean squared prediction errors. The trace plot shown in Figure 1 indicates that the Gibbs draw jumps to the stationary distribution in relatively few steps. We also see that the histograms in Figure 2 based on 10,000 posterior samples reveal that the conditional posteriors are the required stationary distributions.

Hence, both the simulation studies and the real data results show strong support for the use of our proposed method.

Table 5. Mean squared prediction errors for three methods: Lasso, aLasso and NewBLR.

Lasso	aLasso	NewBLR
1.0372	1.0388	0.9736

Figure 9. Trace plots of Tobit regression parameters for real data.

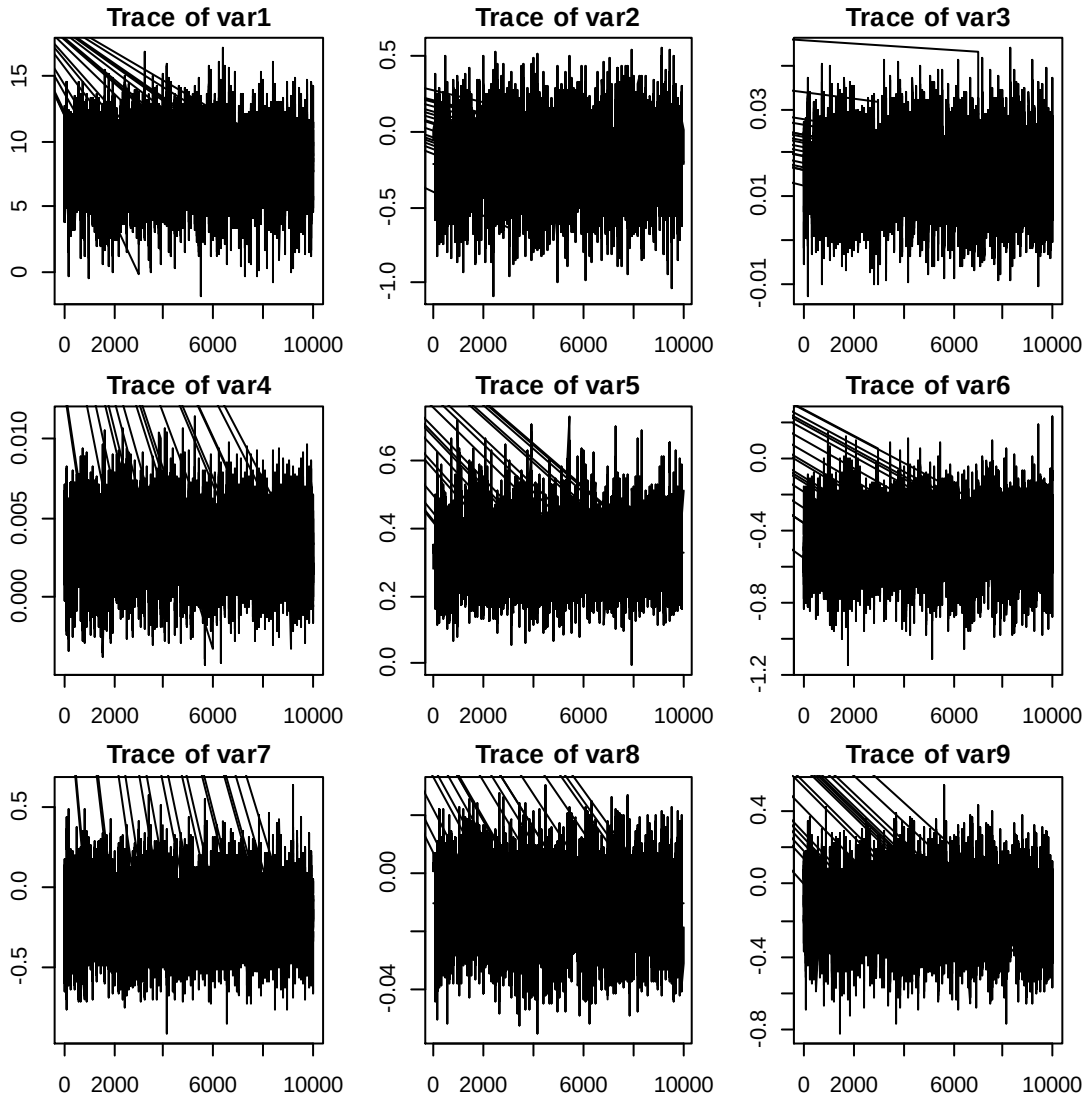
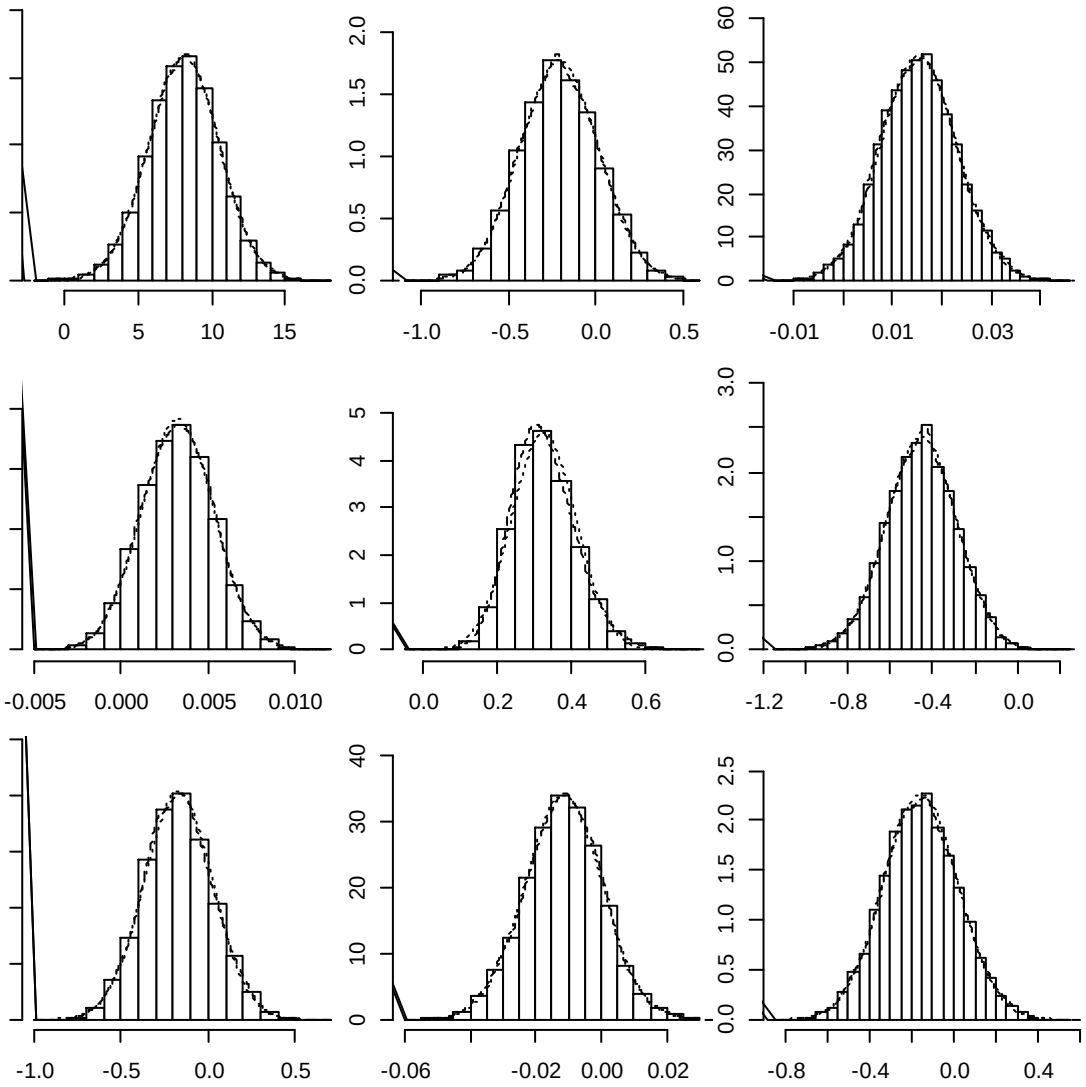


Figure 10. Histograms based on posterior samples of the parameters for real data.



Chapter Six

Conclusions and Recommendations

6.1 Conclusions

In practice, many of right censored data have the problem of high-dimensional data. Therefore, The present thesis presents a new extensions to variables selection technique, through, employing a simple Bayesian variable selection method for right censored regression model. New hierarchical model has developed for our proposed model, as well as we provided efficient Gibbs sampler algorithm for the proposed posterior distribution .

We employed the new method to illuminate the performance of the Bayesian right censored regression model based on the suggested good hierarchical model. In this thesis we focused on the comparison of the quality of the parameter estimation and variable selection procedure in a simulation study and in application of real data. Moreover, we used the Median mean squared error (**MMSE**) and standard deviation (**S.D**) criterions to measure the quality of fitting in the different censored regression models. In both of the simulation study and real data analysis, the proposed method give us promising results. In have concluded that the regularization methods **(NewBLR**) in the censored regression model under the proposed hierarchical model outperform the other methods.

6.2 Recommendation

The proposed method , Bayesian left censored Regression **(NewBLR**) will give feedback to the researchers to develop a new penalized Bayesian censored regression model, such as, Bayesian adaptive lasso right censored Regression , Bayesian fused lasso right censored Regression and Bayesian elastic net right censored Regression, and many other methods of Bayesian variables selection.

Bibliography

Bibliography

Agresti, A., & Franklin, C. (2007). The art and science of learning from data (Vol. 2). Upper Saddle River, NJ: Prentice Hall.

Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. Journal of the American statistical Association, 88(422), 669-679.

Alhamzawi, R. (2015). Model selection in quantile regression models. Journal of Applied Statistics, 42(2), 445-458.

Alhamzawi, R. (2017). Inference with three-level prior distributions in quantile regression problems. Journal of Applied Statistics, 44(11), 1947-1959.

Alhamzawi, R., Alhamzawi, A., & Ali, H. T. M. (2019). New Gibbs sampling methods for bayesian regularized quantile regression. Computers in biology and medicine, 110, 52-65.

Alhamzawi, R., & Ali, H. T. M. (2018). The Bayesian adaptive lasso regression. Mathematical biosciences, 303, 75-82.

Alhamzawi, R., & Taha Mohammad Ali, H. (2020). A new Gibbs sampler for Bayesian lasso. Communications in Statistics-Simulation and Computation, 49(7), 1855-1871.

Alhamzawi, R., & Yu, K. (2014). Bayesian Lasso-mixed quantile regression. Journal of Statistical Computation and Simulation, 84(4), 868-880.

Bibliography

Alhamzawi, R., & Ali, H. T. M. (2018). Bayesian tobit quantile regression with penalty. Communications in Statistics-Simulation and Computation, 47(6), 1739-1750.

Alhamzawi, R., Yu, K., & Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. Statistical Modelling, 12(3), 279-297.

Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. Journal of the Royal Statistical Society: Series B (Methodological), 36(1), 99-102.

Bae, K., & Mallick, B. K. (2004). Gene selection using a two-level hierarchical Bayesian model. Bioinformatics, 20(18), 3423-3430.

Bai, R., & Ghosh, M. (2018). On the Beta Prime Prior for Scale Parameters in High-Dimensional Bayesian Regression Models. arXiv preprint arXiv:1807.06539.

Benoit, D. F., Alhamzawi, R., & Yu, K. (2013). Bayesian lasso binary quantile regression. Computational Statistics, 28(6), 2861-2873.

Bhattacharya, A., Chakraborty, A., & Mallick, B. K. (2016). Fast sampling with Gaussian scale mixture priors in high-dimensional regression. Biometrika, asw042.

Bhattacharya, A., Pati, D., Pillai, N. S., & Dunson, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. Journal of the American Statistical Association, 110(512), 1479-1490.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, 52(3), 345-370.

Bibliography

Bilias, Y., Chen, S., & Ying, Z. (2000). Simple resampling methods for censored regression quantiles. Journal of Econometrics, 99(2), 373-386.

Buchinsky, M., & Hahn, J. (1998). An alternative estimator for the censored quantile regression model. Econometrica, 653-671.

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2009, April). Handling sparsity via the horseshoe. In Artificial Intelligence and Statistics (pp. 73-80).

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. Biometrika, 97(2), 465-480.

Chay, K. Y., & Powell, J. L. (2001). Semiparametric censored regression models. Journal of Economic Perspectives, 15(4), 29-42.

Chernozhukov, V., & Hong, H. (2003). An MCMC approach to classical estimation. Journal of Econometrics, 115(2), 293-346.

Chen, X., Wang, Z. J., & McKeown, M. J. (2011). A Bayesian Lasso via reversible-jump MCMC. Signal Processing, 91(8), 1920-1932.

Chib, S. (1992). Bayes inference in the Tobit censored regression model. Journal of Econometrics, 51(1-2), 79-99.

Donoho, D. L., & Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. biometrika, 81(3), 425-455.

Dziak, J., Li, R., & Collins, L. (2005). Critical review and comparison of variable selection procedures for linear regression. State College, PA: Pennsylvania State University.

Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. Technometrics, 35(2), 109-135.

Griffin, J. E., & Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. Bayesian analysis, 5(1), 171-188.

George, E. I., & McCulloch, R. E. 1993. Variable selection via Gibbs sampling. Journal of the American Statistical Association, 88(423): 881-889.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

Huang, J., Ma, S., Xie, H., & Zhang, C. H. (2009). A group bridge approach for variable selection. Biometrika, 96(2), 339-355.

Javed, F., & Mantalos, P. (2013). GARCH-type models and performance of information criteria. Communications in Statistics-Simulation and Computation, 42(8), 1917-1933.

Kim, Y., Kim, J., & Kim, Y. (2006). Blockwise sparse regression. Statistica Sinica, 375-390.

Kim, J., & Pollard, D. (1990). Cube root asymptotics. The Annals of Statistics, 191-219.

Kordas, G. (2006). Smoothed binary regression quantiles. Journal of Applied Econometrics, 21(3), 387-407.

Leng, C., Tran, M. N., & Nott, D. (2014). Bayesian adaptive lasso. Annals of the Institute of Statistical Mathematics, 66(2), 221-244.

Li, Q., & Lin, N. (2010). The Bayesian elastic net. Bayesian analysis, 5(1), 151-170.

Long, J. S., & Freese, J. (2006). Regression models for categorical dependent variables using Stata (Vol. 7). Stata press.

Magel, R., Unruh S. (2013). Determining Factors Influencing the Outcome of College Basketball Fames. Open Journal of Statistics, Vol. 3 No. 4, 2013, p.225-230.

Mallows, C. L. 1973. Some comments on C p. Technometrics, 15(4): 661-675.

Mallick, H., & Yi, N. (2014). A new Bayesian lasso. Statistics and its interface, 7(4), 571-582.

Mallick, H., & Yi, N. (2017). Bayesian group bridge for bi-level variable selection. Computational statistics & data analysis, 110, 115-133.

Mallick, H., & Yi, N. (2018). Bayesian bridge regression. Journal of applied statistics, 45(6), 988-1008.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. Journal of econometrics, 3(3), 205-228.

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. Journal of econometrics, 27(3), 313-333.

Meier, L., Van De Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(1), 53-71.

Müller, P., & Van de Geer, S. (2016). Censored linear model in high dimensions. Test, 25(1), 75-92.

Murphy, S. A., Van der Vaart, A. W., & Wellner, J. A. (1999). Current status regression.

Nishii, R. 1984. Asymptotic properties of criteria for selection of variables in multiple regression. The Annals of Statistics, 758-765.

Park, T., & Casella, G. (2008). The bayesian lasso. Journal of the American Statistical Association, 103(482), 681-686.

Alonzo, T. A., & Pepe, M. S. (2002). Distribution-free ROC analysis using binary regression techniques. Biostatistics, 3(3), 421-432.

Powell, J. L. (1986). Censored regression quantiles. Journal of econometrics, 32(1), 143-155.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. Biometrics, 1033-1048.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 461-464.

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2013). A sparse-group lasso. Journal of computational and graphical statistics, 22(2), 231-245.

Sun, W., Ibrahim, J. G., & Zou, F. (2010). Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. Genetics, 185(1), 349-359.

Wang, H., & Leng, C. (2008). A note on adaptive group lasso. Computational statistics & data analysis, 52(12), 5277-5286.

Williams, K. R., & Cavanagh, P. R. (1987). Relationship between distance running mechanics, running economy, and performance. Journal of Applied Physiology, 63(3), 1236-1245.

Woodroofe, M. (1982). Nonlinear renewal theory in sequential analysis. Society for Industrial and Applied Mathematics.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.

Zhao, P., Rocha, G., & Yu, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. The Annals of Statistics, 37(6A), 3468-3497.

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418-1429.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), 301-320.

Appendix

Figure 11. Trace plots based on posterior samples of the parameters for Simulation 1 when $n = 50$ and $\sigma^2 = 5$.

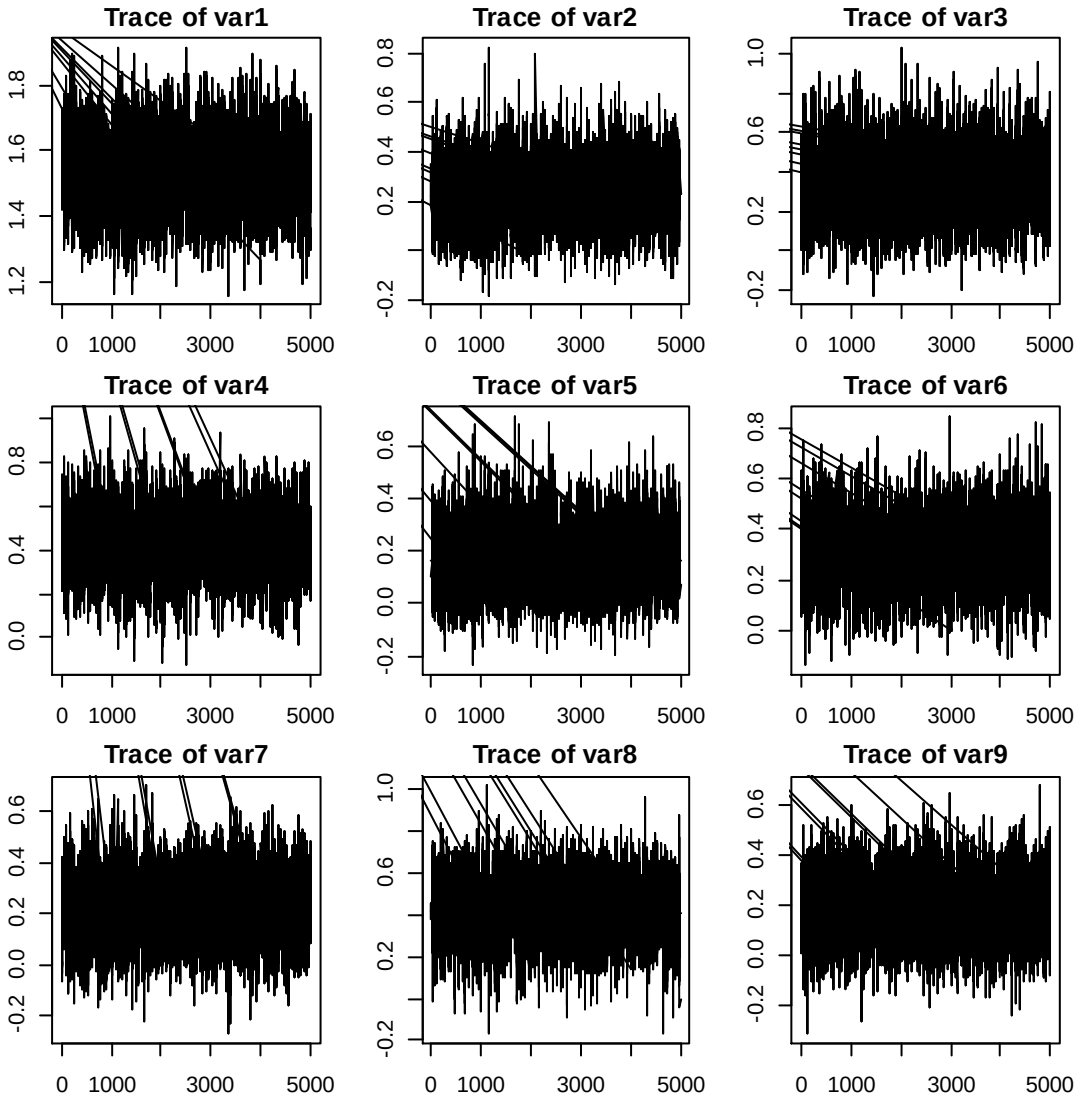


Figure 12. Histograms based on posterior samples of the parameters for Simulation 1 when $n = 50$ and $\sigma^2 = 5$.

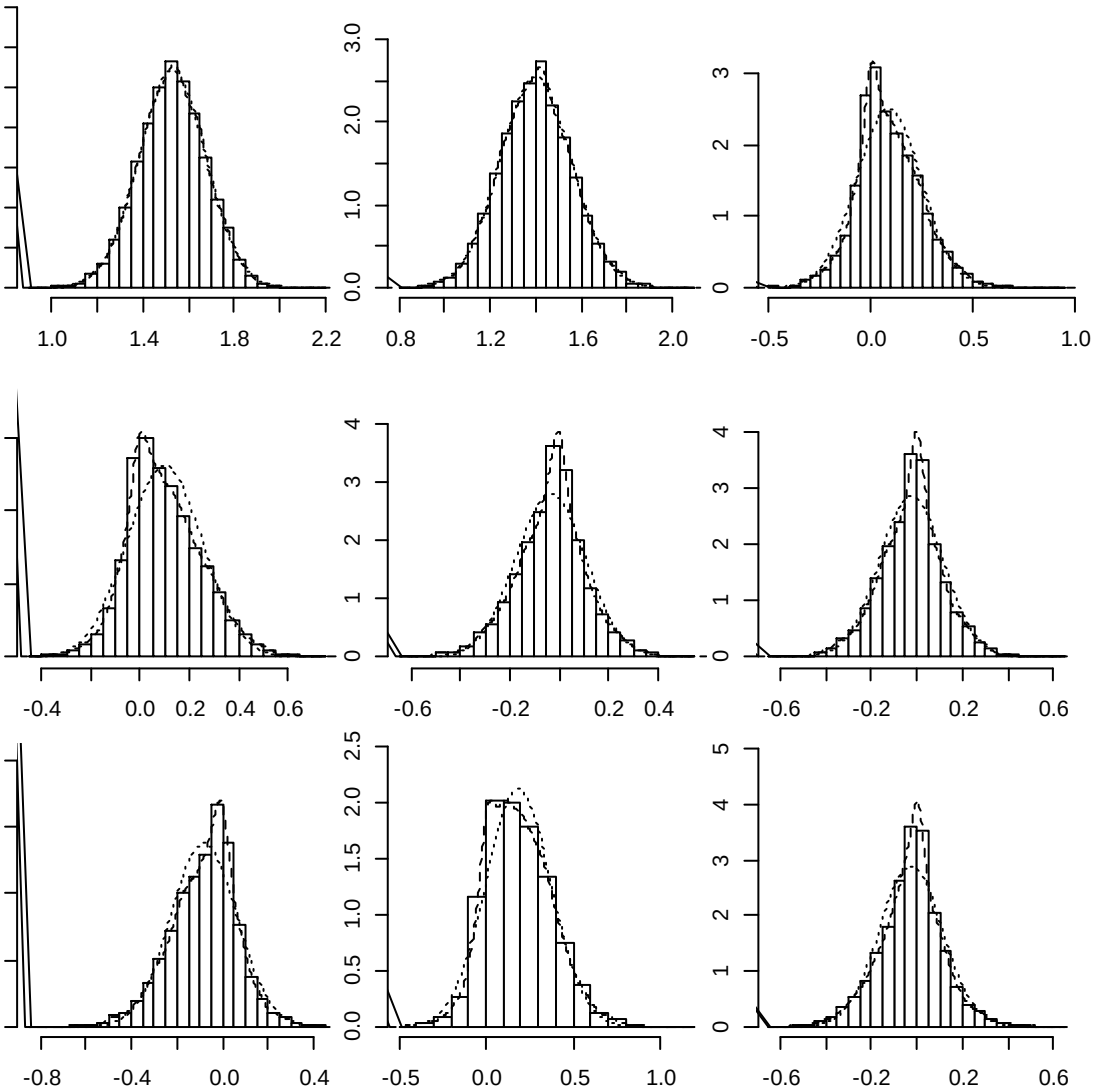


Figure 13. Trace plots based on posterior samples of the parameters for Simulation 1 when $n = 100$ and $\sigma^2 = 5$.

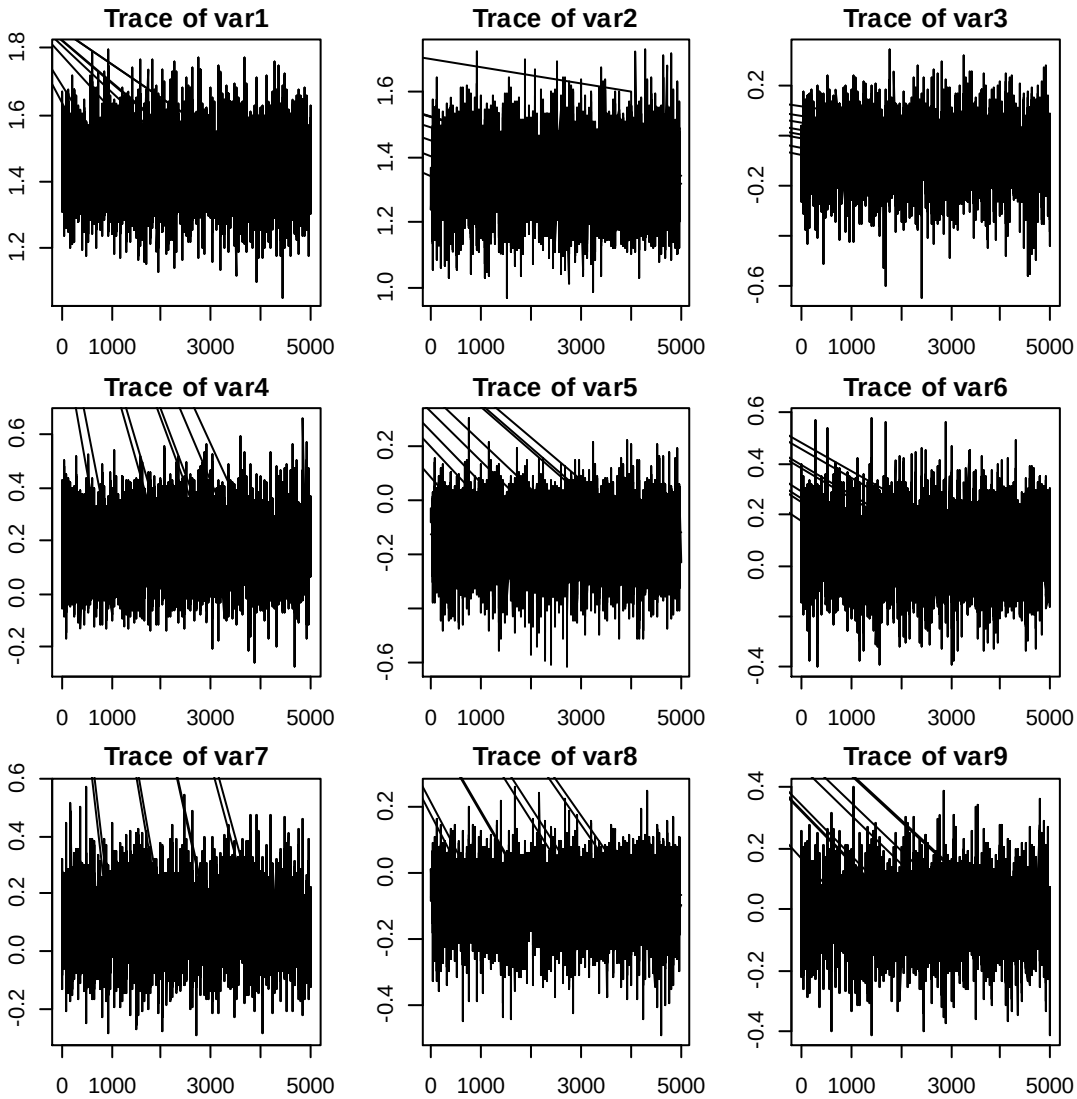


Figure 14. Histograms based on posterior samples of the parameters for Simulation 1 when $n = 100$ and $\sigma^2 = 5$.

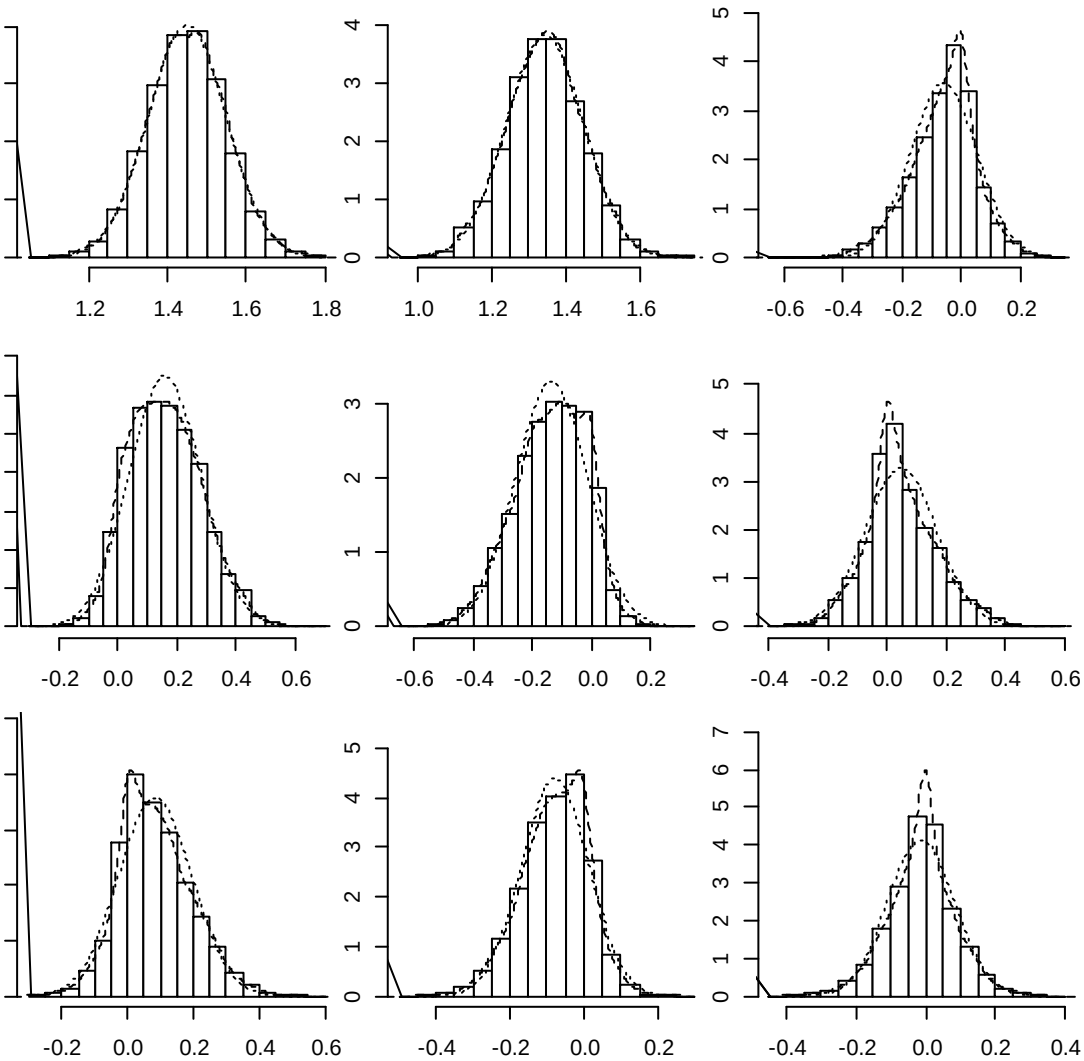


Figure 15. Trace plots based on posterior samples of the parameters for Simulation 1 when $n = 200$ and $\sigma^2 = 5$.

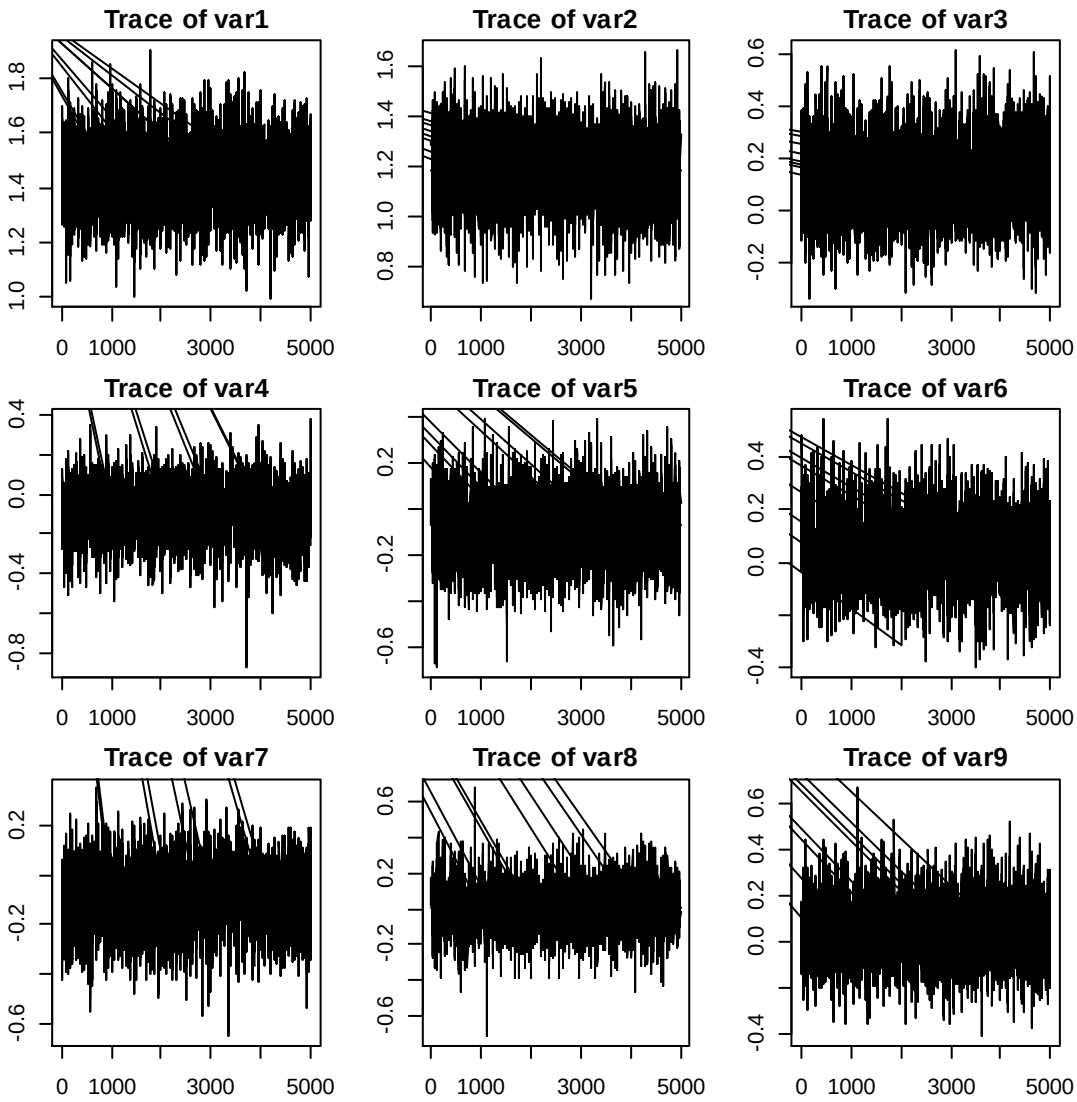


Figure 16. Histograms based on posterior samples of the parameters for Simulation 1 when $n = 200$ and $\sigma^2 = 5$.

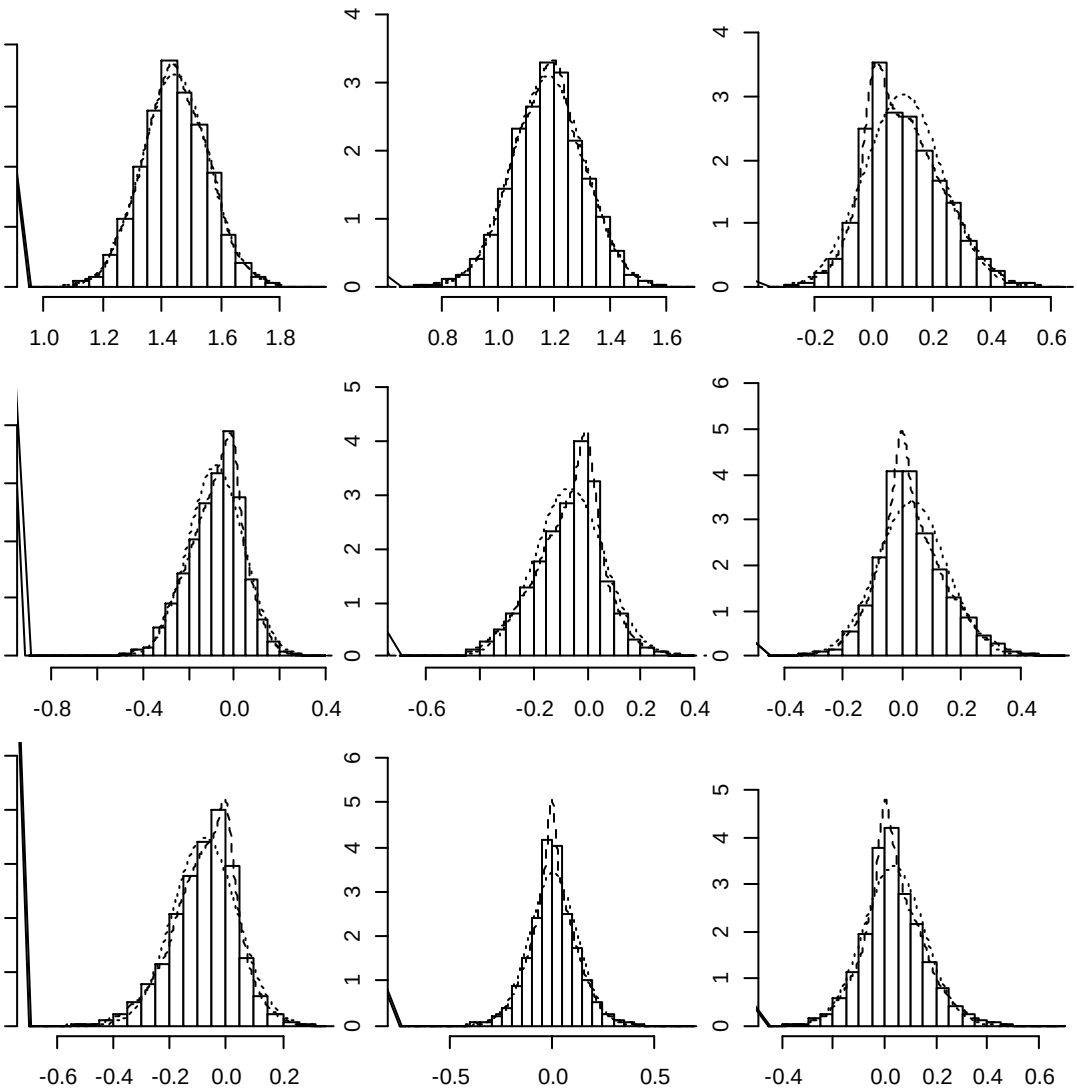


Figure 17. Trace plots based on posterior samples of the parameters for Simulation 2 when $n = 50$ and $\sigma^2 = 5$.

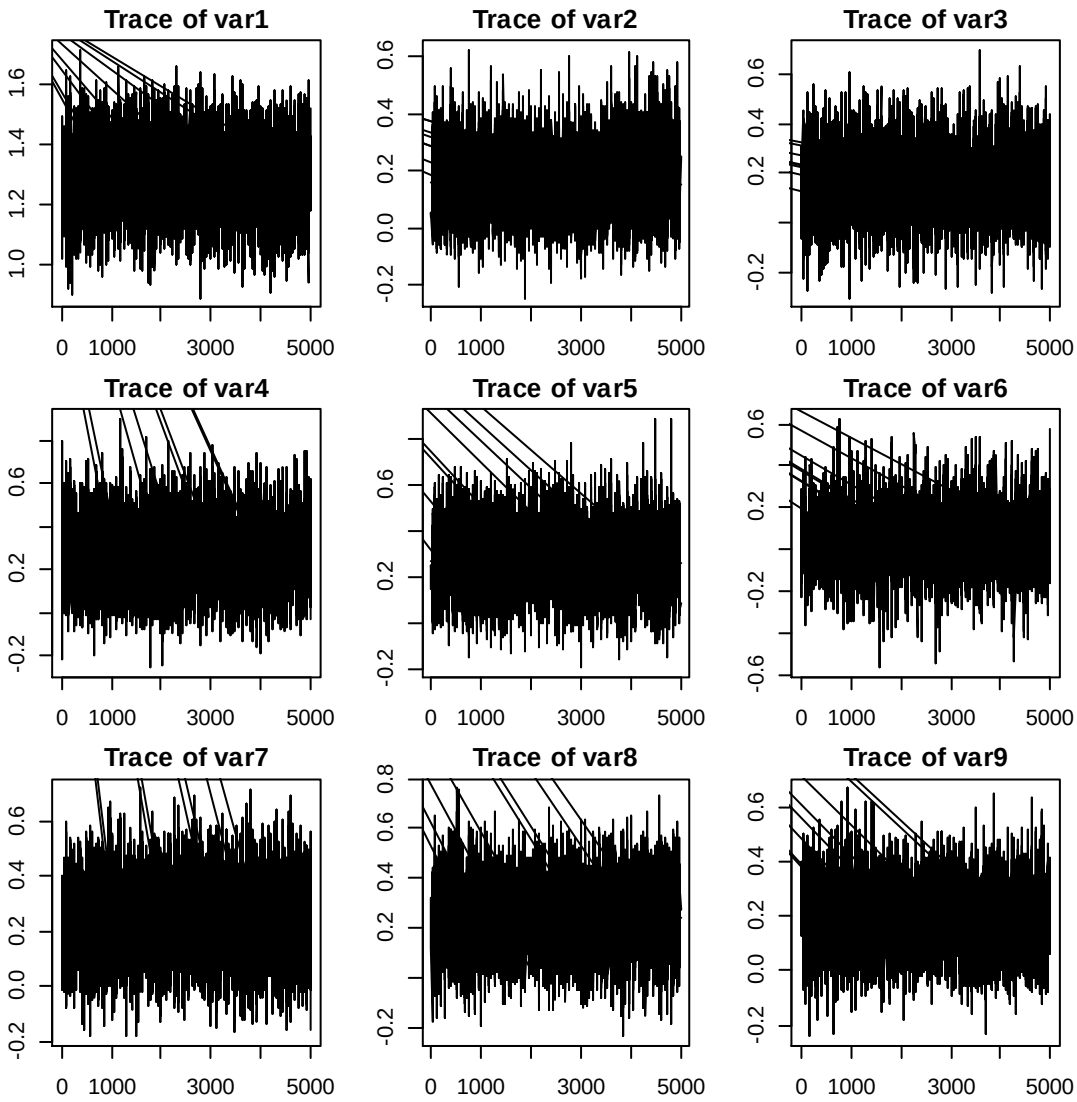


Figure 18. Histograms based on posterior samples of the parameters for Simulation 2 when $n = 50$ and $\sigma^2 = 5$.

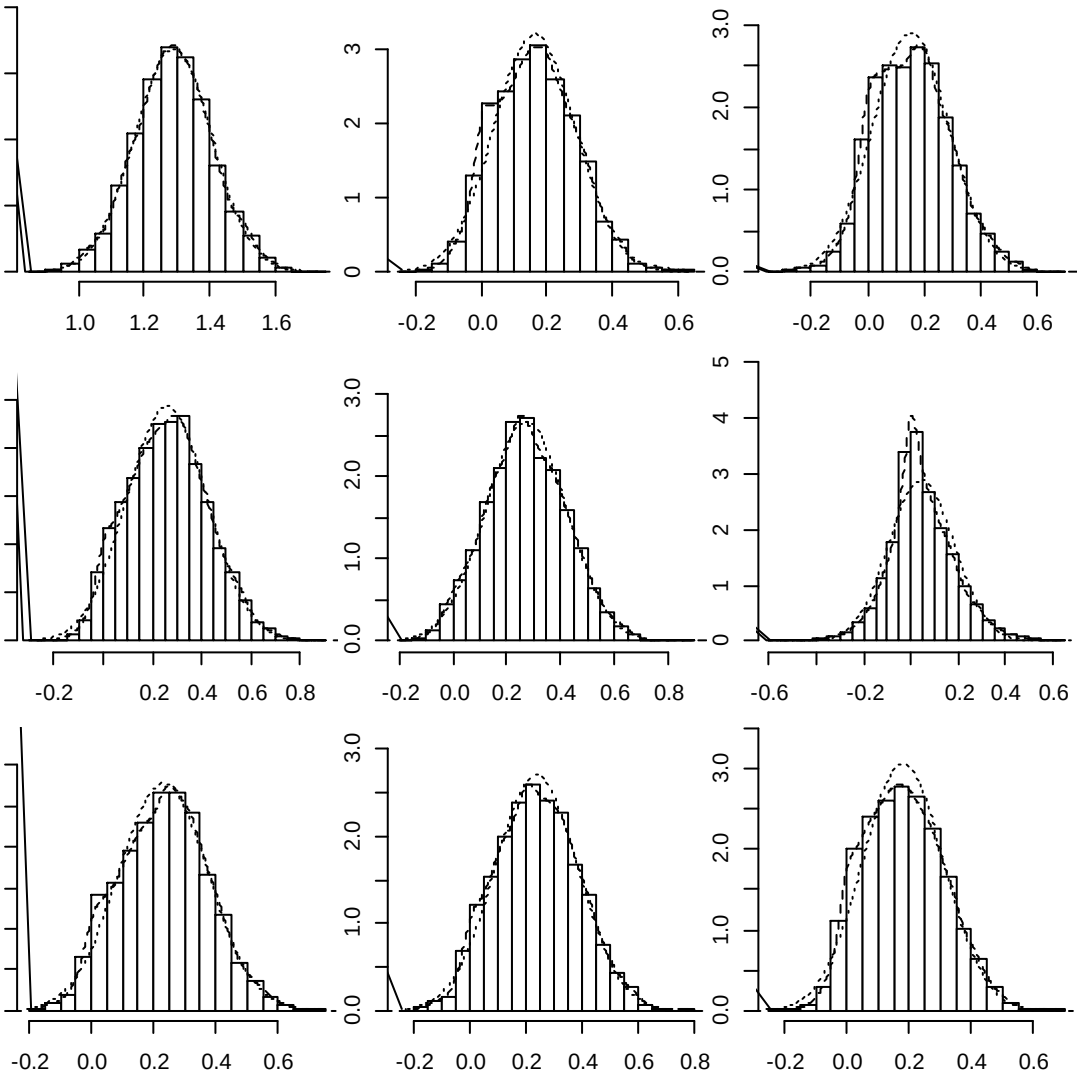


Figure 19. Trace plots based on posterior samples of the parameters for Simulation 2 when $n = 100$ and $\sigma^2 = 5$.

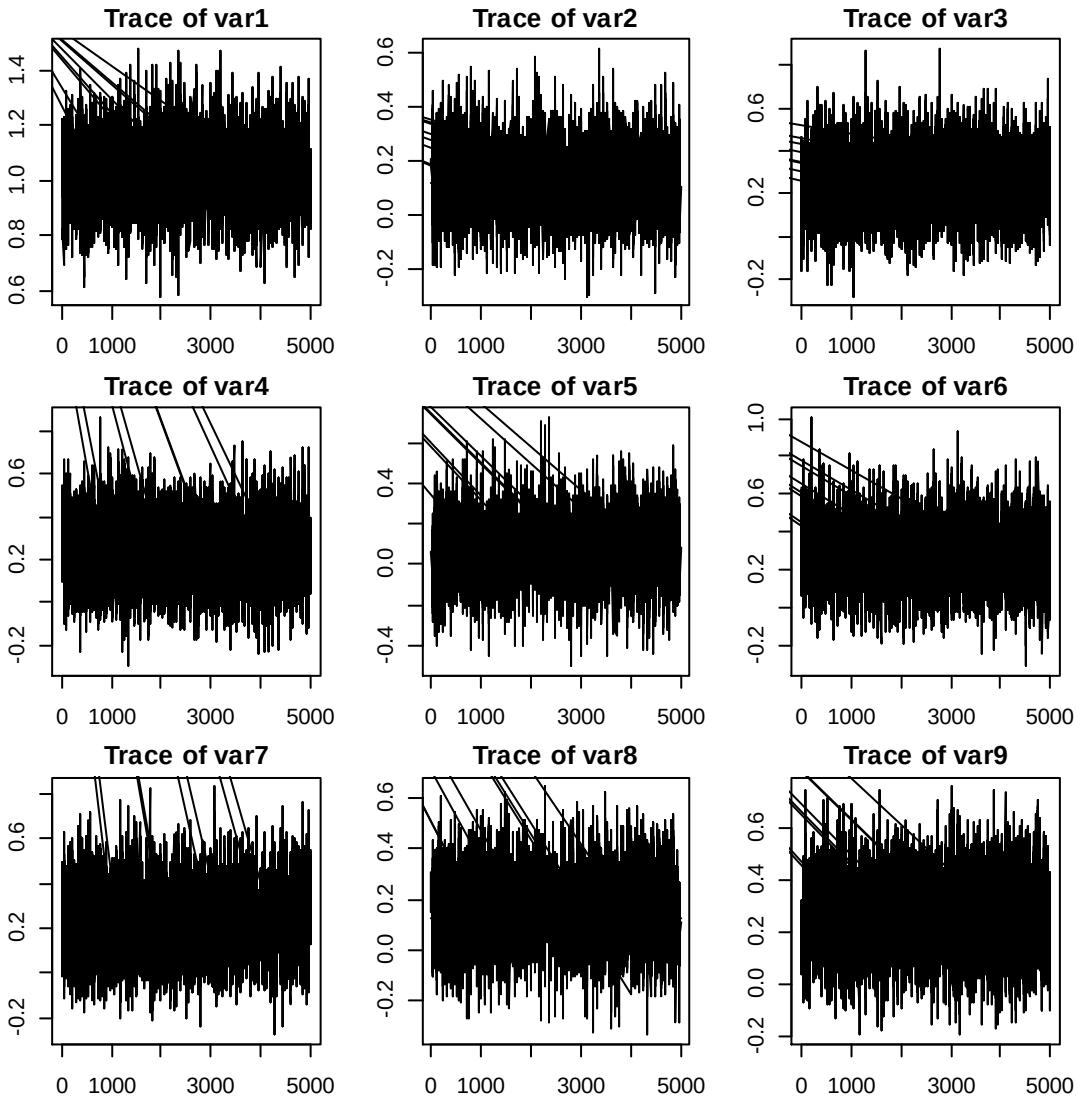


Figure 20. Histograms based on posterior samples of the parameters for Simulation 2 when $n = 100$ and $\sigma^2 = 5$.

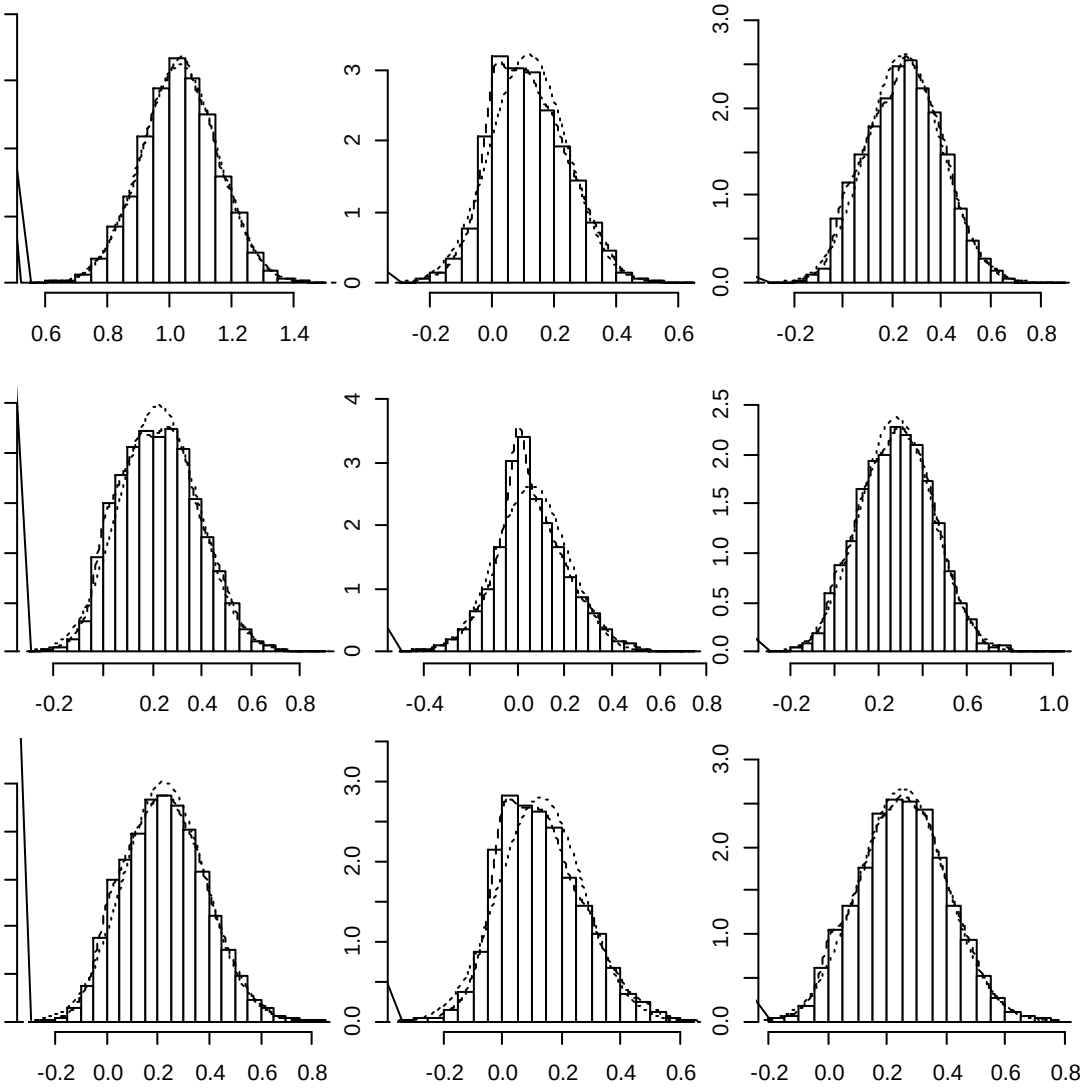


Figure 21. Trace plots based on posterior samples of the parameters for Simulation 2 when $n = 200$ and $\sigma^2 = 5$.

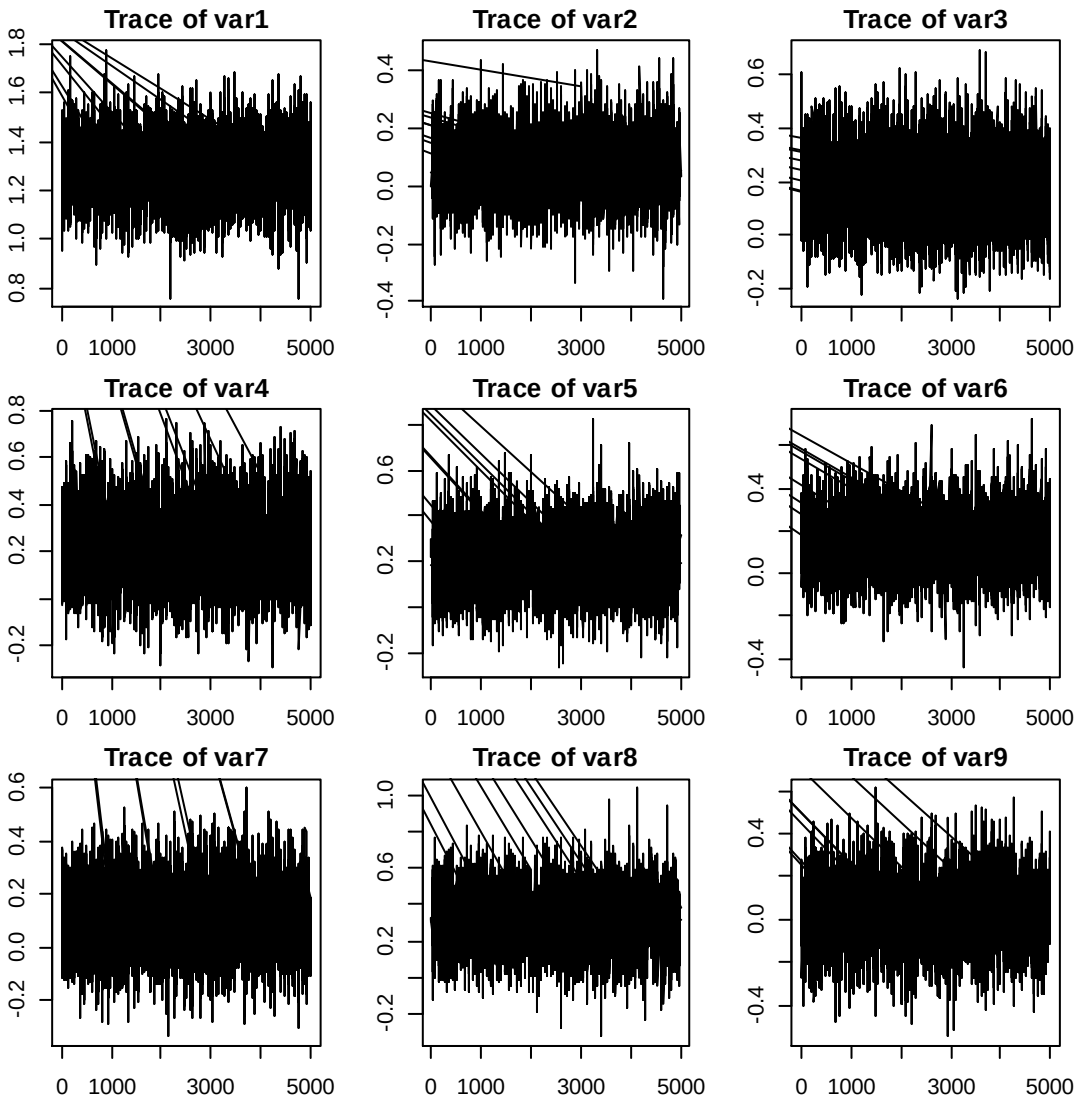


Figure 22. Histograms based on posterior samples of the parameters for Simulation 2 when $n = 200$ and $\sigma^2 = 5$.

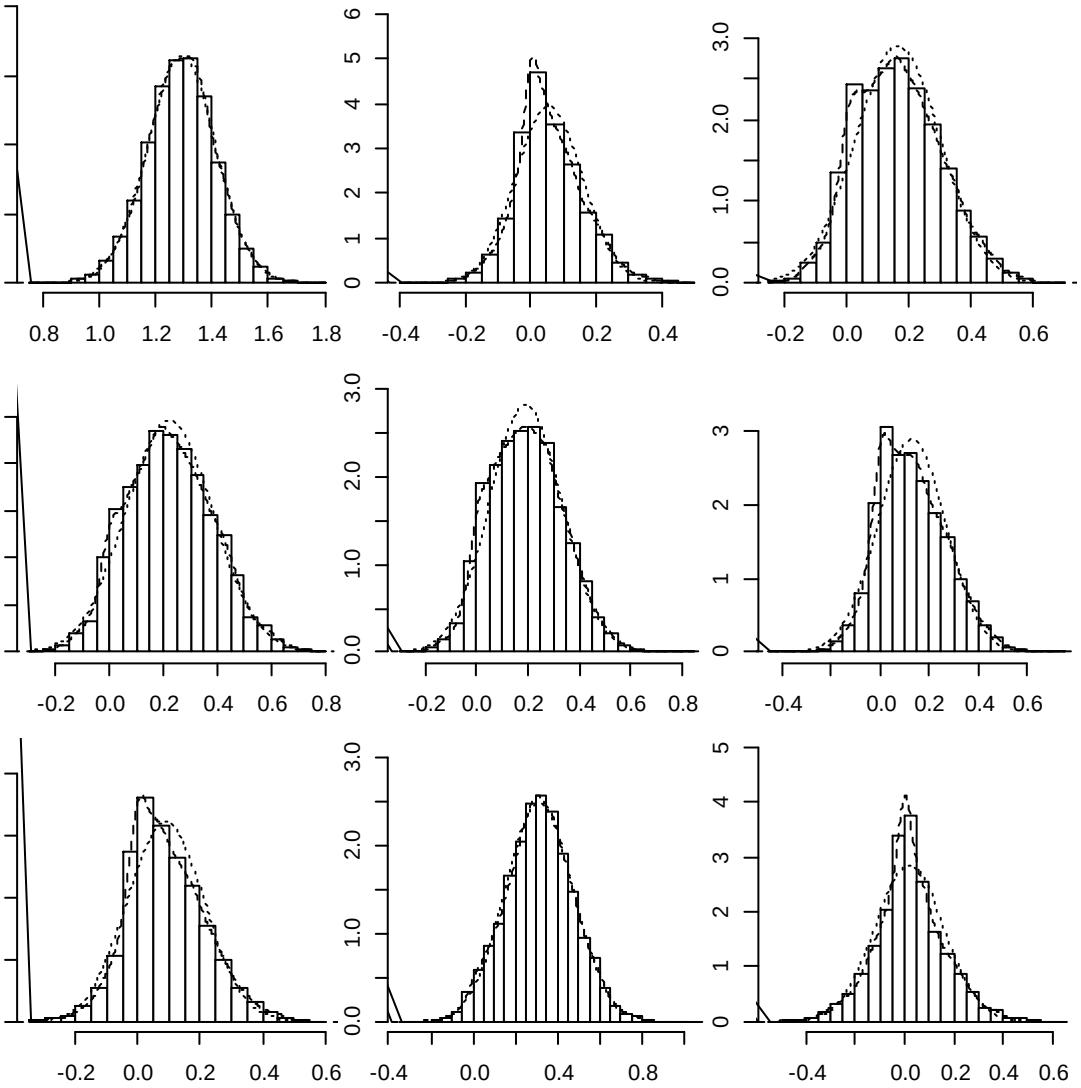


Figure 23. Trace plots based on posterior samples of the parameters for Simulation 3 when $n = 50$ and $\sigma^2 = 5$.

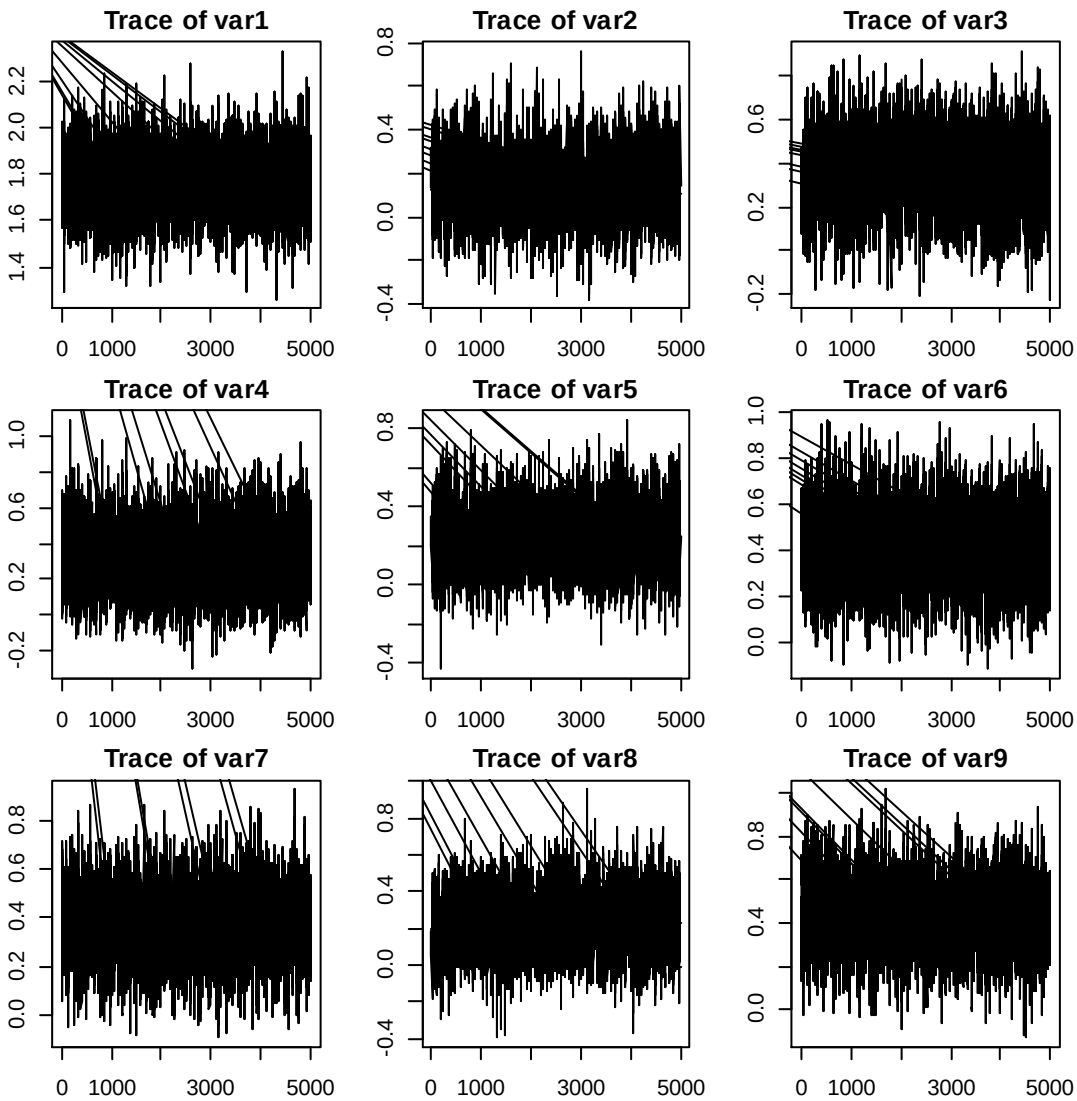


Figure 24. Histograms based on posterior samples of the parameters for Simulation 3 when $n = 50$ and $\sigma^2 = 5$.

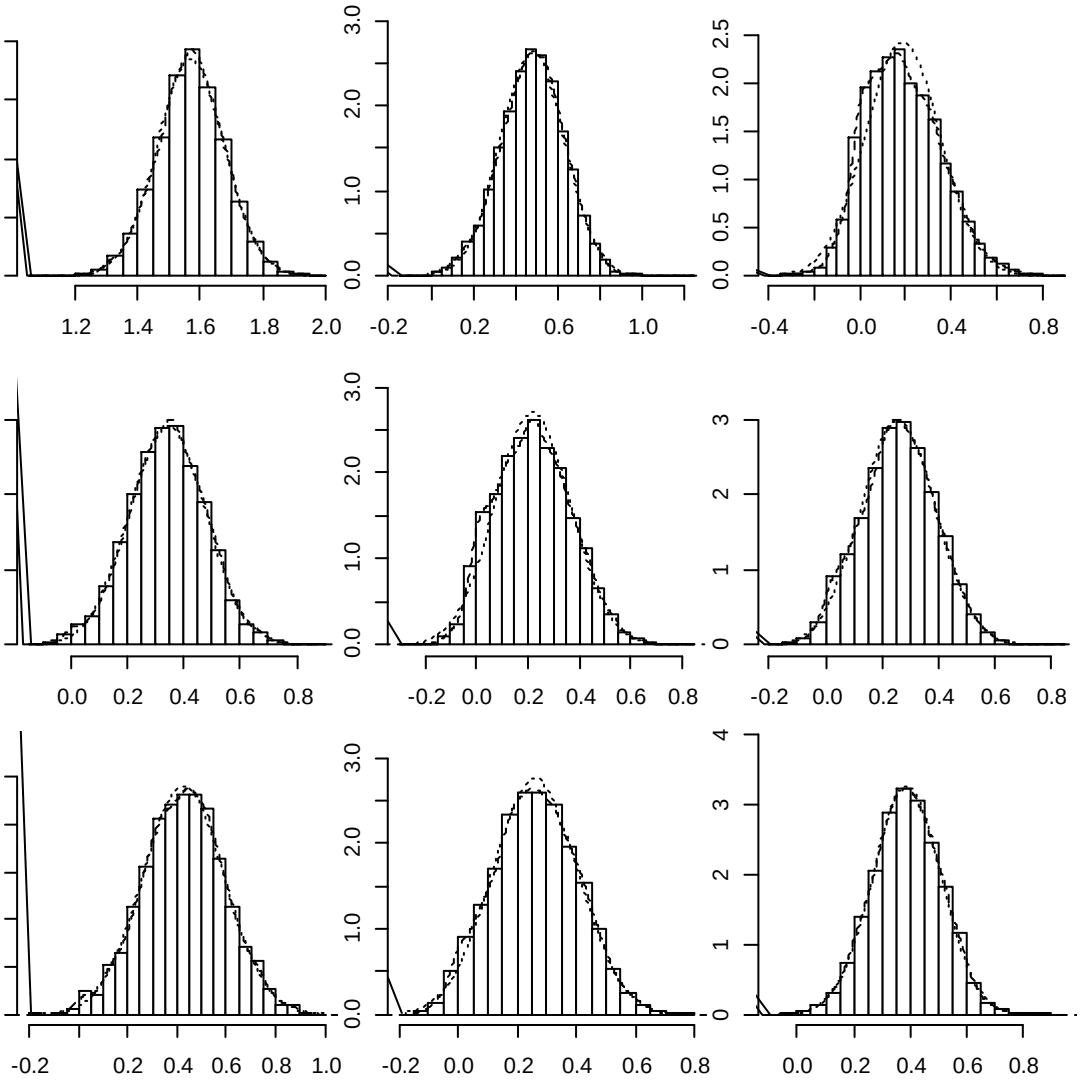


Figure 25. Trace plots based on posterior samples of the parameters for Simulation 3 when $n = 100$ and $\sigma^2 = 5$.

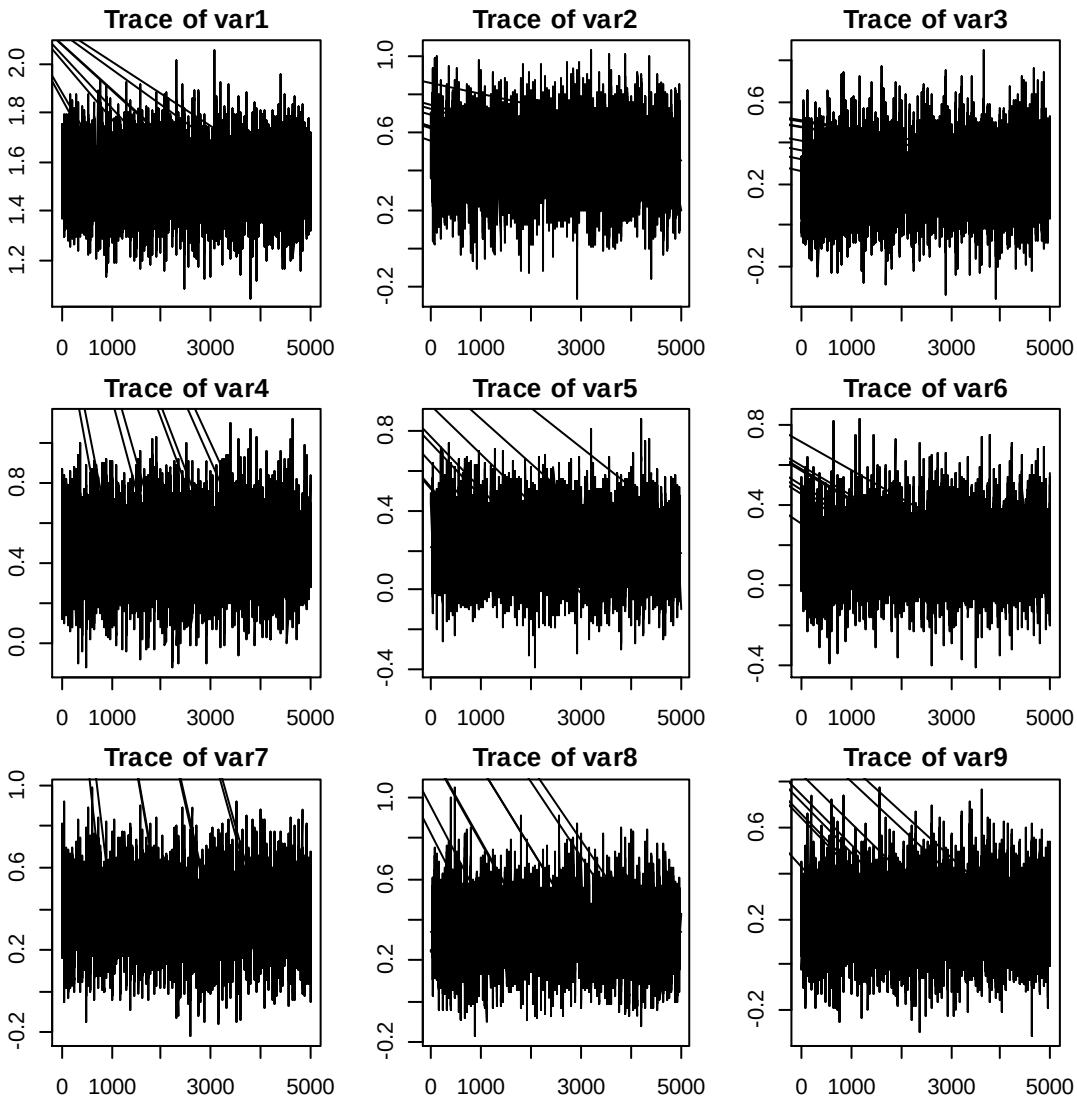


Figure 26. Histograms based on posterior samples of the parameters for Simulation 3 when $n = 100$ and $\sigma^2 = 5$.

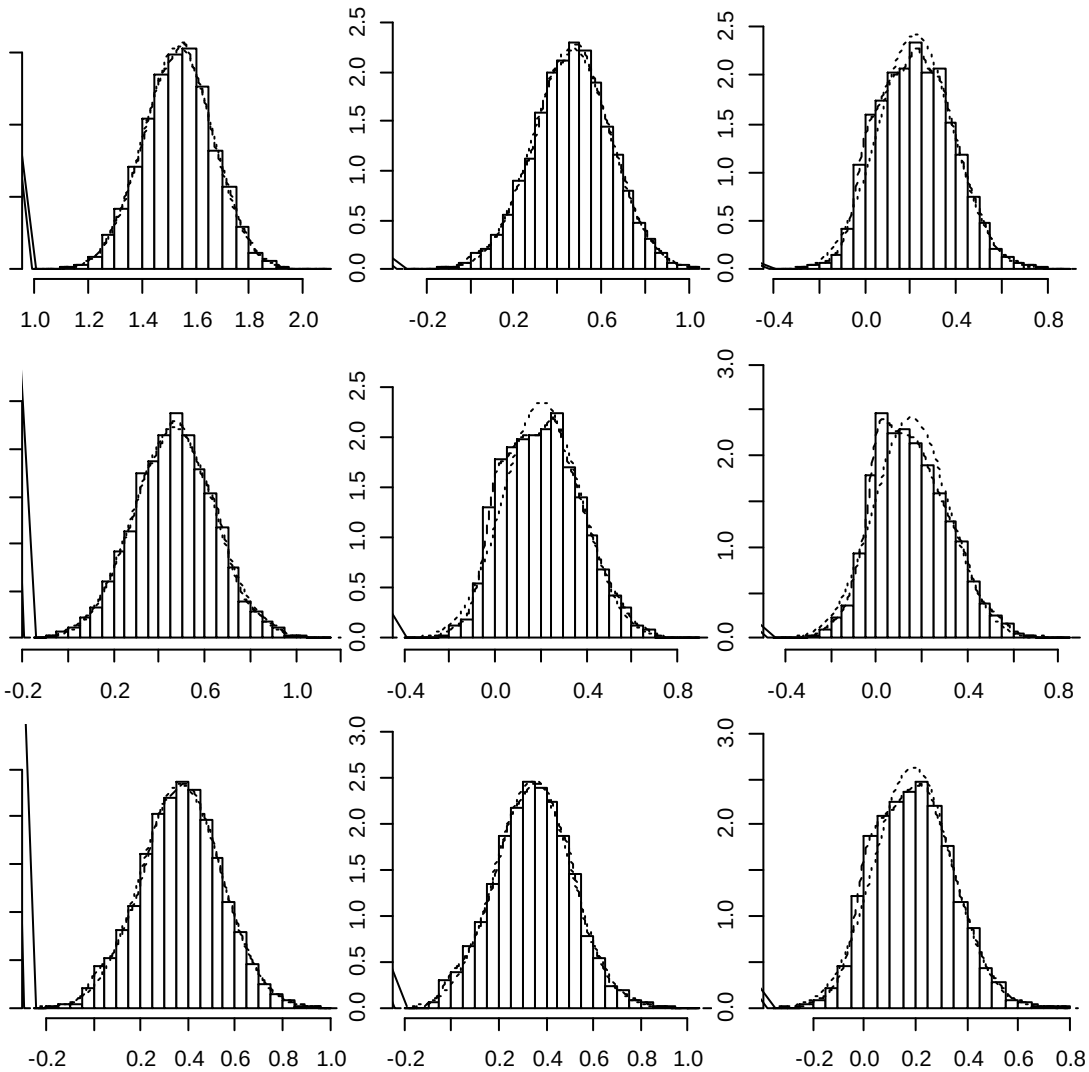


Figure 27. Trace plots based on posterior samples of the parameters for Simulation 3 when $n = 200$ and $\sigma^2 = 5$.

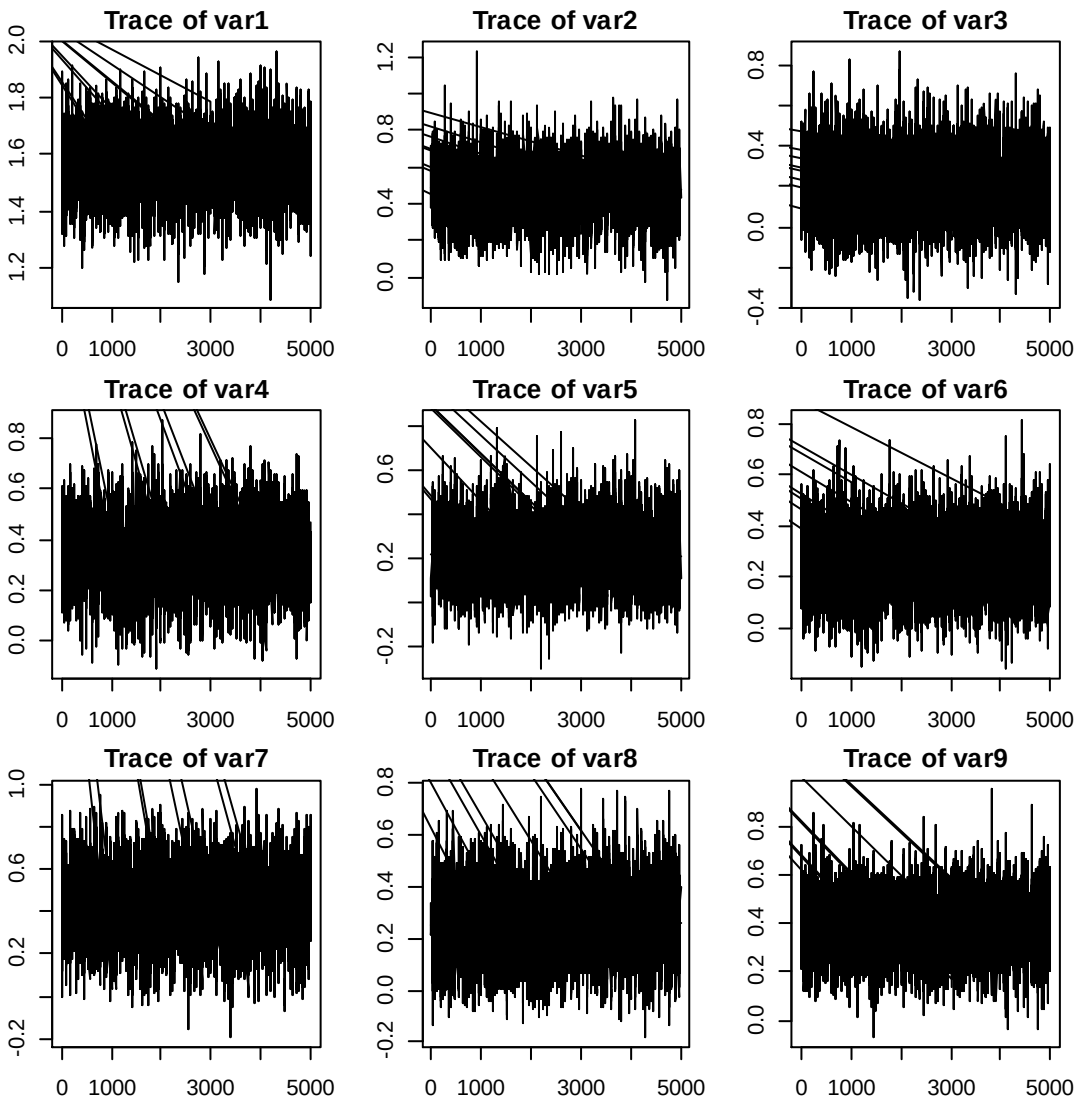


Figure 28. Histograms based on posterior samples of the parameters for Simulation 3 when $n = 200$ and $\sigma^2 = 5$.

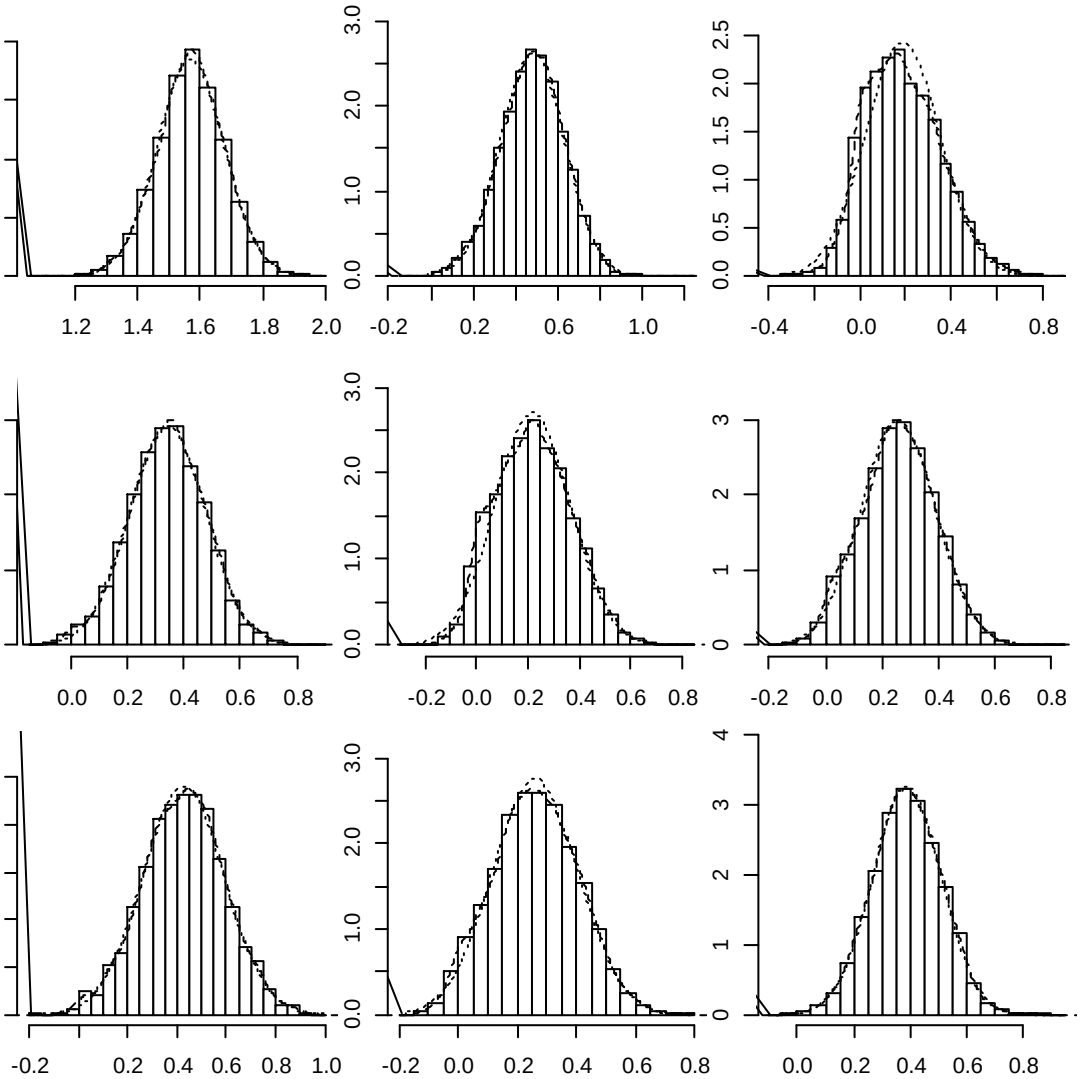


Figure 29. Trace plots based on posterior samples of the parameters for Simulation 4 when $n = 50$ and $\sigma^2 = 5$.

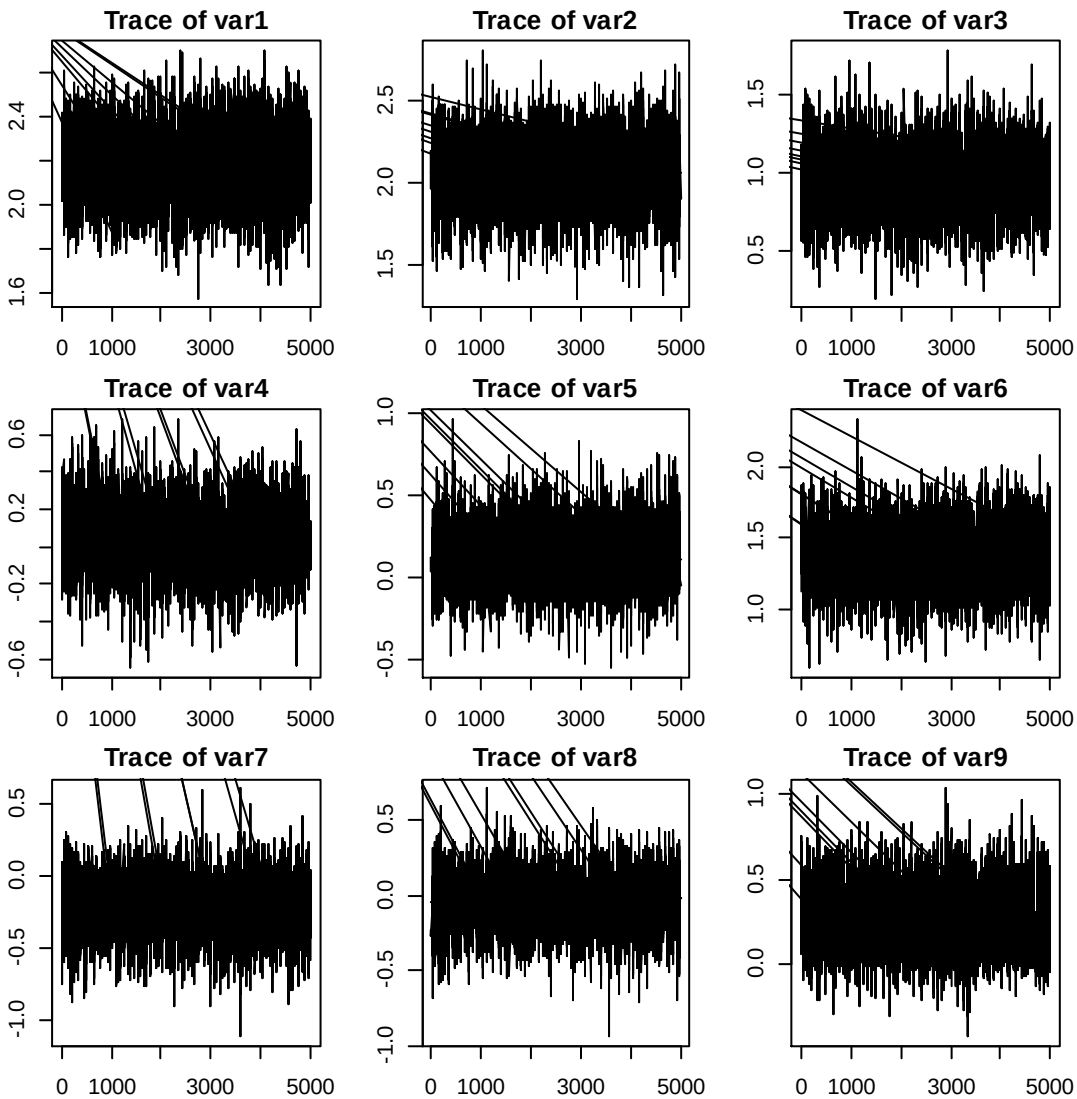


Figure 30. Histograms based on posterior samples of the parameters for Simulation 4 when $n = 50$ and $\sigma^2 = 5$.

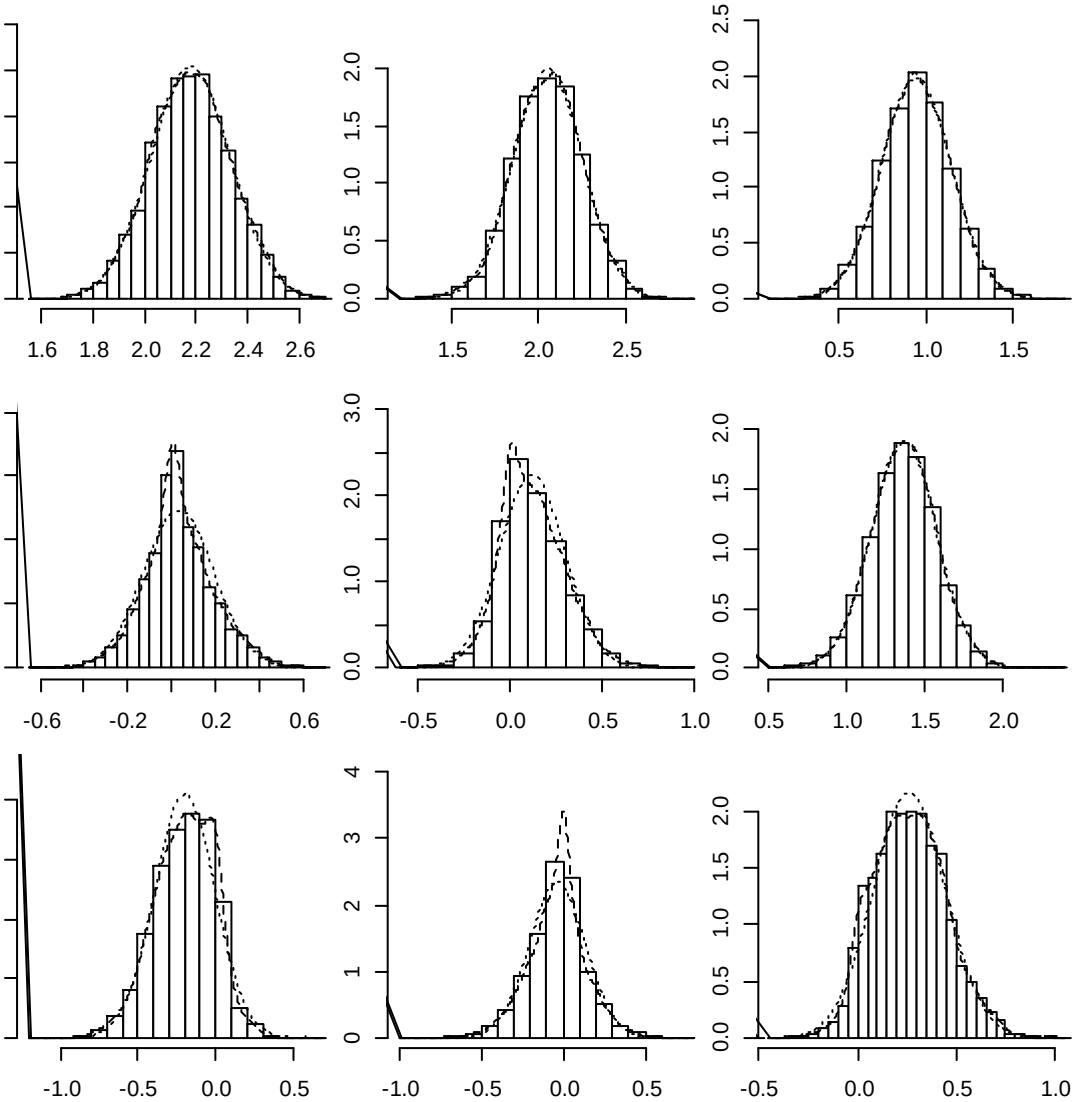


Figure 31. Trace plots based on posterior samples of the parameters for Simulation 4 when $n = 100$ and $\sigma^2 = 5$.

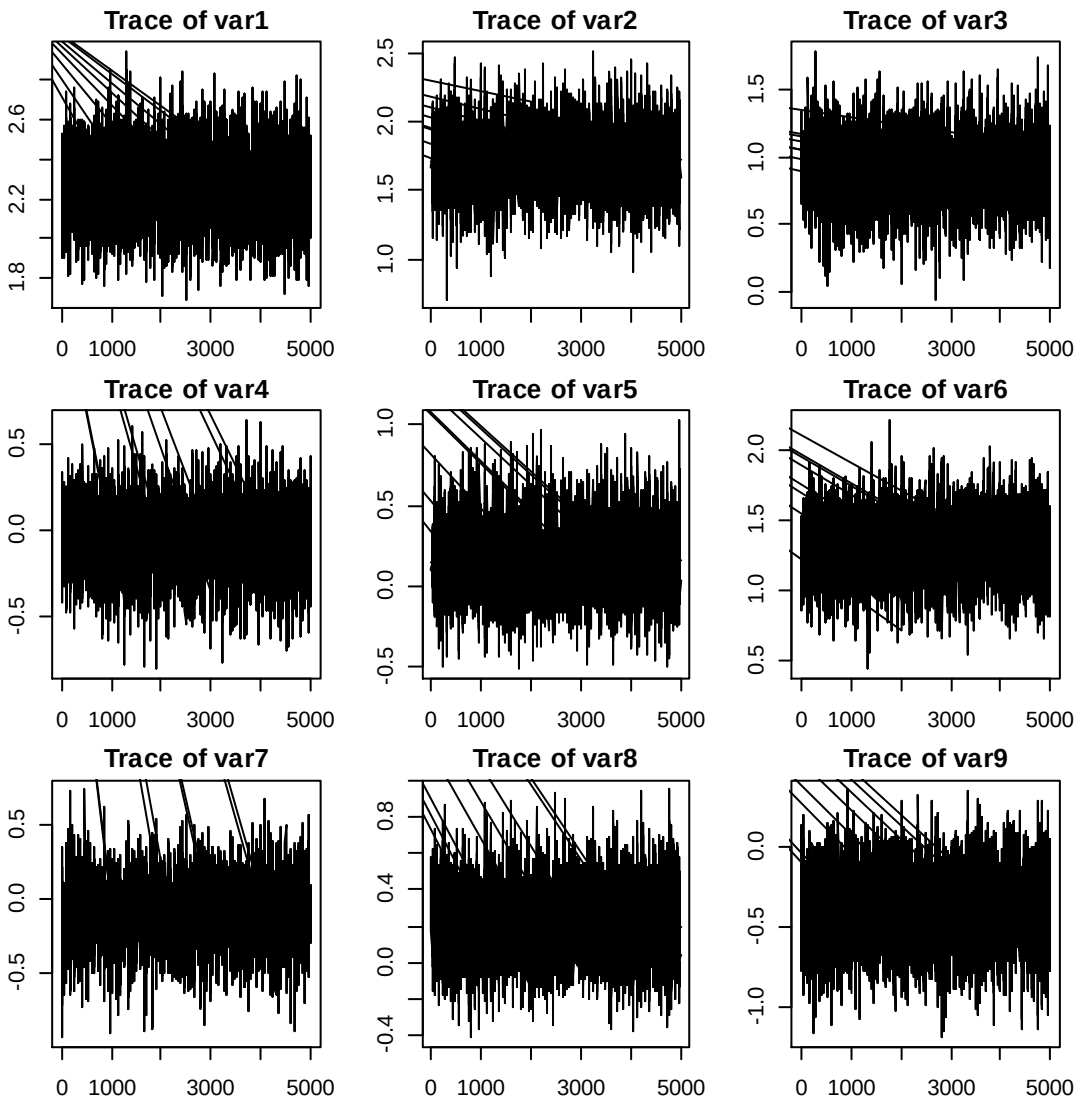


Figure 32. Histograms based on posterior samples of the parameters for Simulation 4 when $n = 100$ and $\sigma^2 = 5$.

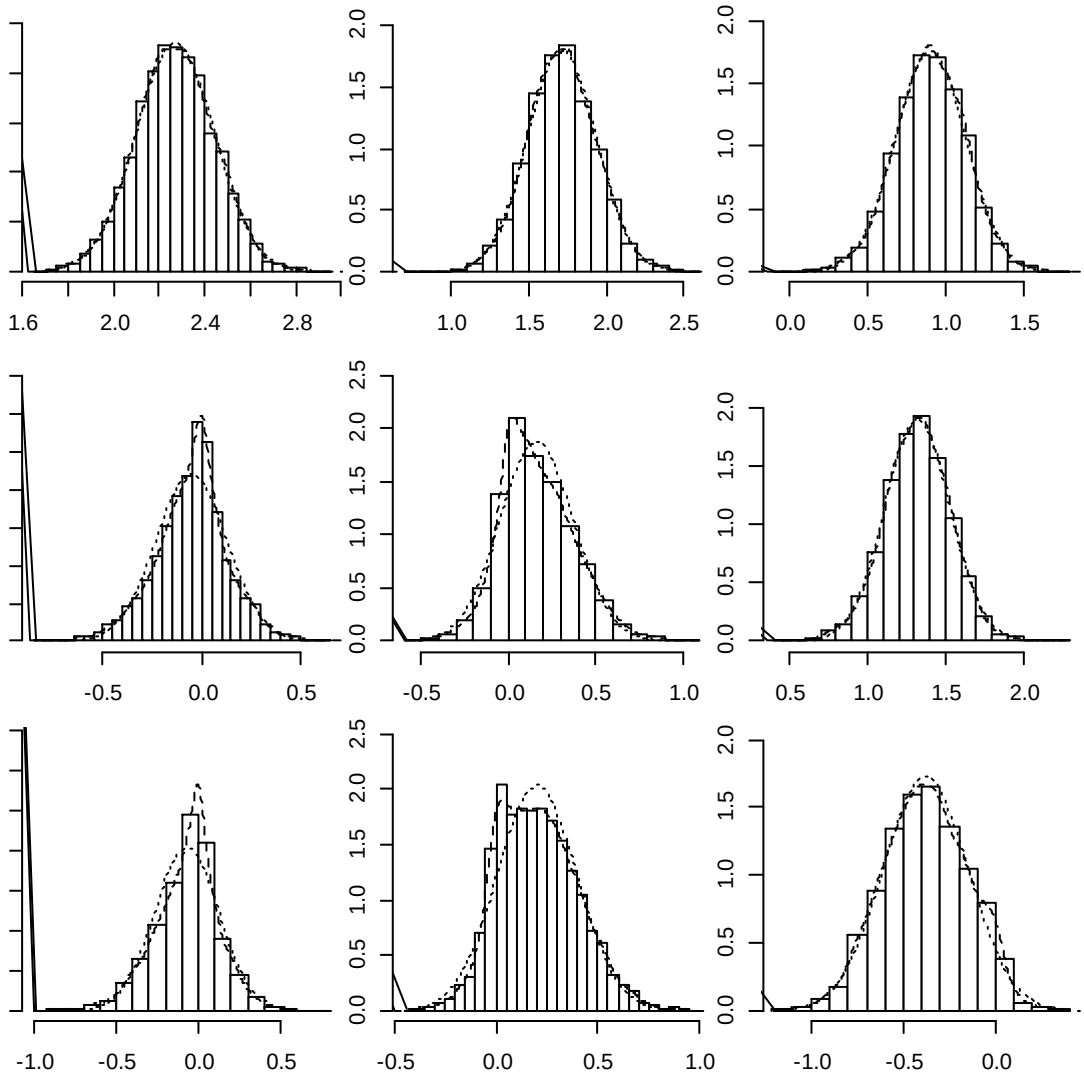


Figure 33. Trace plots based on posterior samples of the parameters for Simulation 4 when $n = 200$ and $\sigma^2 = 5$.

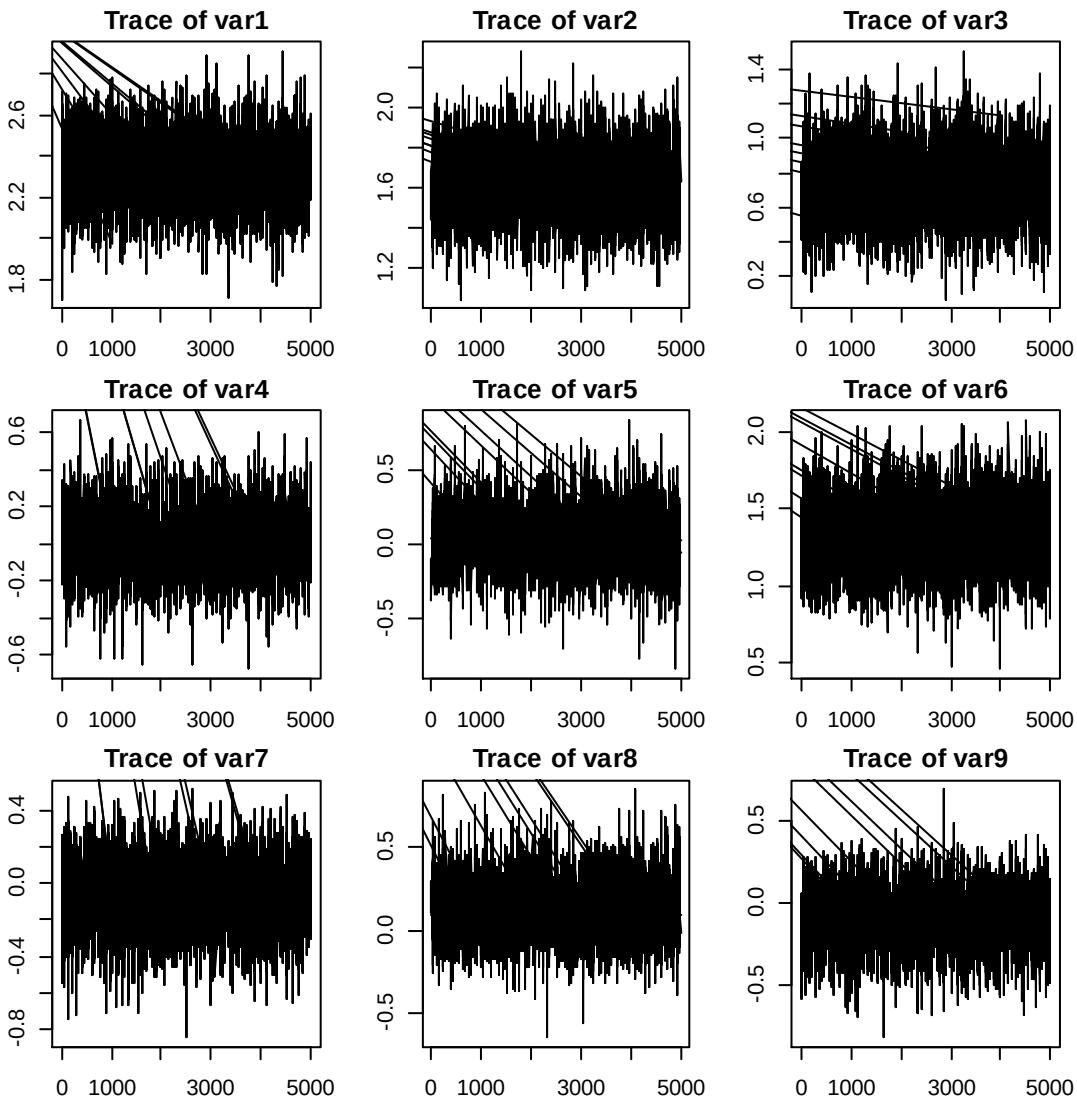
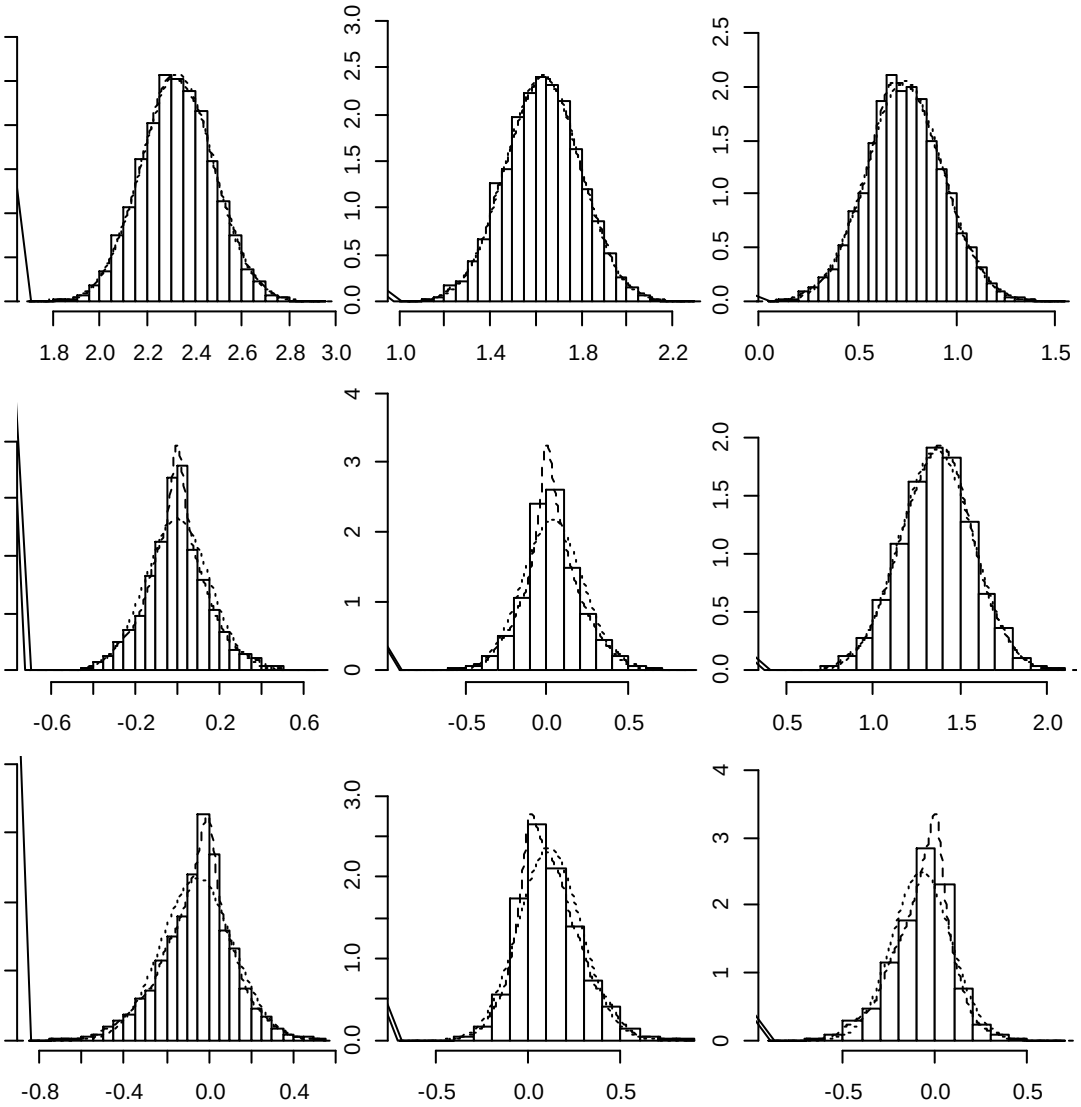


Figure 34. Histograms based on posterior samples of the parameters for Simulation 4 when $n = 200$ and $\sigma^2 = 5$.



The figures in the appendix show that the our Gibbs sampler perform very well in terms of convergences and stability.

المستخلص

إن إجراء طرق اختيار المتغير المستخدمة هي لتقييم العلاقة بين مجموعة المتغيرات التفسيرية والمتغير التابع. تم اقتراح العديد من طرق اختيار المتغير على مر السنين. في السنوات الأخيرة ، كان هناك بحث فعال حول اختيار المتغير باستخدام طرق الانحدار البيزي.

في هذه الرسالة ، قمنا بتطوير طرق بيز لاختيار المتغير في البيانات الخاضعة للرقابة من جهة اليسار ومن جهة اليمين، والبيانات الثنائية التي تؤدي إلى طرق جديدة لأخذ عينات Gibbs، مع امكانية الحصول على توزيعات شرطية كاملة مسبقة. من خلال أمثلة المحاكاة المكثفة وتحليلات البيانات الحقيقية ، ومقارنة أداء طريقتنا المقترحة للبيانات الخاضعة للرقابة من جهة اليسار مع بعض الطرق البيزية وغير البيزية.

وزارة التعليم العالي والبحث العلمي
جامعة القادسية
كلية الإدارة والاقتصاد
قسم الإحصاء
الدراسات العليا

الاستدلال باستخدام توزيعي كاما ومعكوس كاما في مشكلة انحدار توبت مع تطبيق عملي

رسالة مقدمة

الى مجلس كلية الإدارة والاقتصاد – جامعة القادسية

جزءاً من متطلبات نيل درجة الماجستير في علوم الإحصاء

من الطالب

إسماعيل علي حمد السلماوي

إشراف

أ. د. رحيم جبار الحمزاوي

2021 م 1443 هـ