A Simulation Study on Bayesian Estimation of Half-Normal Regression Model with Deterministic Frontier

Submitted by

Krishnendu Paul	211322
Rahul Ghosh Dastidar	211353
Saprativa Bhowal	211371
Souraj Mazumdar	211393

Under the Supervision of

Dr. Arnab Hazra

Department Of Mathematics And Statistics,

IIT Kanpur

Abstract

In this project, we are going to review the paper **"Bayesian Estimation of Half-Normal Regression Model with Deterministic Frontier"** by **Francisco J. Ortega** and **Jose M. Gavilan** (published in the journal **Computational Statistics** in the year 2016). In the paper a regression model with deterministic frontier is considered. Analysis of this type of model is not very common because this is a non regular model which makes it difficult to apply maximum likelihood estimation. As an alternative to ML estimation, a Bayesian Methodology is proposed. Through the Gibbs algorithm, sampling from the posterior distribution and making inference about the model parameters is straightforward. We will see further how the method works through simulation studies and data application.

Acknowledgement

It is our pleasure to present a project on **"A Simulation Study on Bayesian Estimation of Half-Normal Regression Model with Deterministic Frontier"**. Every accomplishment has constant encouragement and advice from valuable and noble minds to guide us in putting our efforts in the right direction to bring out the project. We want to express our sincere gratitude to our instructor **Dr. Arnab Hazra** for his constant help and support throughout the completion of the project. Without his valuable guidance and motivation, it was nearly impossible to work on this project as a team and understand the practical aspect of the course **"MTH535A: AN INTRODUCTION TO BAYESIAN ANALYSIS"**. Also we are thankful to all faculty members and seniors without whose support at various stages, this project would not have materialized. Finally our earnest thanks go to our friends who were always beside us when we needed them without any excuses.

Krishnendu Paul

Rahul Ghosh Dastidar

Saprativa Bhowal

Souraj Mazumdar

Contents

1 Introduction

The concept and the analysis of the frontier model was started in 1957 (by Farrell). In 1968, Aigner and Chu proposed a model where the dependent variable (output) is expressed as a function of independent variables (inputs), some unknown parameters with a negative random perturbation. This is called as the deterministic frontier model (DFM). The deterministic part of the model explains the production frontier or maximum attainable value of output. The difference between the estimated production frontier and the actual production represents the inefficiency in the production process.

But, after the apperance of the stochastic frontier model (SFM) in 1977, the use and the analysis of the DFM declined drastically. There are a quite number of reasons behind the declination of the use of the deterministic frontier model.

Firstly, the model does not take into consideration the source error. The second problem of this model is that it possesses a non regular problem, i.e., the properties of the maximum likelihood estimator (MLE) are very uncertain in this case. These are some drawbacks of the DFM. The later drawback of the model is solved to some extent by assuming that the density of ϵ and its derivative at the point $\epsilon = 0$ are equal to 0. Distributions like log-normal, gamma with shape parameter ≥ 2 fullfill the condition whereas the exponential and the half-normal distributions fail to satisfy this condition.

There are also some drawbacks of SFM. In many cases, it is difficult to seperate out the random effect and the inefficiency of the composite error and there is a significance portion of samples that involve one or the other extreme.

The Bayesian approach to the DFM have some interesting advantages :

- ease of adding constraints involving both the parameters and the observations, which make the model non-regular
- the possibility of obtaining exact inferences for finite samples, for the parameters of the model ,as well as for the individual efficiencies.

In our project we will try to reproduce the findings in the paper, where some Bayesian analysis of the DFM has been performed using Gibbs sampling.

2 The Model : Prior and the Posterior Distributions

Let us consider the model -

$$y_i = \boldsymbol{x_i'\beta} - \epsilon_i,\ i = 1, 2, ..., n$$

where $y_i{'s}$ are the responses, $\boldsymbol{\beta} \in \mathbb{R}^k$ is a column vector of the unknown parameters, $\boldsymbol{x_i{'s}}$ are the vectors of the independent variables with $x_{i1} = 1\,\forall$i , i.e., the model contains an intercept term, $\epsilon_i{'s}$ (> 0) are the random error component that measures the inefficiency of the i^{th} observation, n is the sample size and k is the number of parameters to be estimated.

It is assumed that the ϵ_i follows a half normal distribution $(0, \sigma^2)$, i.e., the pdf of the ϵ_i is given by-

$$f(\epsilon_i) = \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\epsilon_i^2}{\sigma^2}},\ \epsilon_i > 0\,, \forall\, i = 1, 2, ..., n \tag{1}$$

Therefore the probability density function of y_i is given by-

$$f(y_i|\boldsymbol{x_i}, \boldsymbol{\beta}, \sigma) = \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\left(y_i - \boldsymbol{x_i'\beta}\right)^2},\ y_i \leq \boldsymbol{x_i'\beta}\,, \forall\, i = 1, 2, ..., n$$

The likelihood function of the model parameters for a sample of size n is given by-

$$L(\boldsymbol{\beta}, \sigma|\boldsymbol{y}, \boldsymbol{X}) \propto \sigma^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_i (y_i - \boldsymbol{x_i'\beta})^2\right\},\ y_i \leq \boldsymbol{x_i'\beta},\ \forall\, i = 1, 2, ..., n$$

where $\boldsymbol{y} = (y_1, y_2, ..., y_n)' \in \mathbb{R}^n$ and $\boldsymbol{X'} = [\boldsymbol{x_1}|\boldsymbol{x_2}|\boldsymbol{...}|\boldsymbol{x_n}] \in M_{k\times n}$.

Now, let us define a set B as $B = \left\{\boldsymbol{\beta} \in \mathbb{R}^k |\ y_i \leq \boldsymbol{x_i'\beta}\ \forall i = 1, 2, ..., n\right\}$. B is the parametric space of the parameter vector $\boldsymbol{\beta}$ that is compatible with the observed set of data according to the model. As a result , the constraint can be expressed as $\boldsymbol{\beta} \in B$. It is generally assumed that B is non-empty.

As mentioned earlier, our focus will be on the DFM, described by the equation (1). The set of unknown parameters is $(\boldsymbol{\beta}, \sigma)$. In Bayesian analysis the main focus is to obtain the posterior distribution of the parameter vector $(\boldsymbol{\beta}, \sigma)'$ given $\boldsymbol{y}$ and $\boldsymbol{X}$. We will denote the posterior distribution by $\pi(\boldsymbol{\beta}, \sigma|\boldsymbol{y}, \boldsymbol{X})$. But for that, first of all we have to define a suitable prior distribution for the parameters. The priors will be denoted by $\pi(\boldsymbol{\beta}, \sigma)$.

In Bayesian literature often we have seen that a suitable prior of σ is Inverse Gamma and a suitable prior for $\boldsymbol{\beta}$ is Normal, i.e., the joint prior for $(\boldsymbol{\beta}, \sigma)$ will be Gaussian-Inverse Gamma. The reason being that the prior becomes conjugate,so the analysis and the inference become quite easy.

But there are also other priors such as non-informative priors. These are helpful when we do not have any prior knowledge about the distribution of the parameters. There are a quite number of methods to define a non-informative prior. Here we will use the Jeffery's non-informative prior given by-

$$\pi(\boldsymbol{\beta}, \sigma) \propto \sigma^{-1}$$

Thus, the joint posterior distribution is biven by-

$$\pi(\boldsymbol{\beta}, \sigma|\boldsymbol{y}, \boldsymbol{X}) \propto \left(\sigma^2\right)^{-\frac{(n+1)}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_i (y_i - \boldsymbol{x_i'}\boldsymbol{\beta})^2\right\}, \boldsymbol{\beta} \in B$$

which can be written as

$$\pi(\boldsymbol{\beta}, \sigma|\boldsymbol{y}, \boldsymbol{X}) \propto \left(\sigma^2\right)^{-\frac{(n+1)}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\}, \boldsymbol{\beta} \in B$$

Thus we can see that even by using a non informative prior, we get the posterior distribution as Gaussian-Inverse Gamma. The algebraic calculation to obtain the marginal posteriors is not so easy because of the constraint $\boldsymbol{\beta} \in B$. So we will use the Gibbs Sampling.

3 Formulation of the Gibbs Algorithm

We have seen in Section 2 that the kernel of the posterior distribution of $(\boldsymbol{\beta}, \sigma)$ is Gaussian-Inverse Gamma. Thus $\sigma^{-2}|\boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{X}$ will follow a Gamma distribution and $\boldsymbol{\beta}|\sigma, \boldsymbol{X}.\boldsymbol{y}$ will follow a k-dimensional multivariate normal (truncated to the constraint $\boldsymbol{\beta} \in B$).

Now the joint posterior can be written as-

$$\pi(\boldsymbol{\beta}, \sigma|\boldsymbol{y}, \boldsymbol{X}) \propto \left(\sigma^{-2}\right)^{\frac{n+1}{2}} \exp\left\{-\sigma^{-2}\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\right\}, \boldsymbol{\beta} \in B$$

So by integrating over $\boldsymbol{\beta}$, we will get the posterior distribution of σ^{-2} as -

$$\sigma^{-2}|\boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{X} \sim Gamma\left(\frac{n+3}{2}, \frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)'\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)\right)$$

To obtain the distribution of $\boldsymbol{\beta}|\sigma, \boldsymbol{y}, \boldsymbol{X}$, let us consider $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'y}$ and $\hat{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$. Hence, we can write -

$$\left(\boldsymbol{y}-\boldsymbol{X\beta}\right)'\left(\boldsymbol{y}-\boldsymbol{X\beta}\right) = \hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} + \left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\boldsymbol{X'X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right),$$

By using the above equation the joint posterior can be expressed as -

$$\pi\left(\boldsymbol{\beta}, \sigma|\boldsymbol{y}, \boldsymbol{X}\right) \propto \left(\sigma^{2}\right)^{-\frac{(n+1)}{2}} \exp\left\{-\frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\boldsymbol{X'X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\}, \boldsymbol{\beta} \in B$$

which implies -

$$\boldsymbol{\beta}|\sigma, \boldsymbol{X}, \boldsymbol{y} \sim N_k\left(\hat{\boldsymbol{\beta}}, \sigma^2\left(\boldsymbol{X'X}\right)^{-1}\right), \boldsymbol{\beta} \in B$$

which is a k dimensional normal distribution truncated to B.

In general a simple algorithm can be used which is as follows -

At m^{th} step, let $\boldsymbol{\beta}^{(m)}$ and $(\sigma^2)^{(m)}$ be the obtained value of $\boldsymbol{\beta}$ and (σ^2). Then-

1. Simulate $v^{(m+1)}$ from the distribution $\sigma^{-2}|\boldsymbol{\beta}^{(m)}, \boldsymbol{y}, \boldsymbol{X}$. Then take $(\sigma^2)^{(m+1)} = \frac{1}{v^{(m+1)}}$
2. Generate a value $\boldsymbol{\beta}^*$ from $N_k\left(\hat{\boldsymbol{\beta}}, \left(\sigma^2\right)^{(m+1)}\left(\boldsymbol{X'X}\right)^{-1}\right)$.
3. If $\boldsymbol{\beta}^* \in B$, then take $\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^*$. If not, then return to step 2.

However it is seen that in practice the probability of rejection is very high in this procedure. So this method becomes unviable.

So, to overcome this problem, one of the method is to apply Gibbs algorithm. In Gibbs algorithm, we use the marginal distribution of each $\beta_j' s$. We will obtain samples from one dimentional full conditional distributions i.e., from $\sigma|\boldsymbol{\beta, y, X}$ and from $\beta_j|\boldsymbol{\beta_{(j)}}, \sigma, \boldsymbol{y, X}$ where $\boldsymbol{\beta_{(j)}} = (\beta_1, ..., \beta_{j-1}, \beta_{j+1}, ..., \beta_k)' \in \mathbb{R}^{k-1}$. Thus the Gibbs algorithm is as follows-

At m^{th} step, let $\boldsymbol{\beta}^{(m)}$ and $(\sigma^2)^{(m)}$ be the obtained value of $\boldsymbol{\beta}$ and σ^2. Then-

- Simulate $v^{(m+1)}$ from the distribution $\sigma^{-2}|\boldsymbol{\beta}^{(m)}, y, X$. Then take $(\sigma^2)^{(m+1)} = \frac{1}{v^{(m+1)}}$
- For all $j = 1, ..., k$, Generate a value $\beta_j^{(m+1)}$ from the distribution -

$$\beta_j|\beta_1^{(m+1)}, ..., \beta_{j-1}^{(m+1)}, \beta_{j+1}^{(m)}, ..., \beta_k^{(m)}, \left(\sigma^2\right)^{(m+1)}, \boldsymbol{y}, \boldsymbol{X}$$

- Thus the $(m+1)^{th}$ sample of $\boldsymbol{\beta}$ is $\boldsymbol{\beta}^{(m+1)} = \left(\beta_1^{m+1}, ..., \beta_k^{m+1}\right)'$

To obtain the distribution of $\beta_j|\boldsymbol{\beta_{(j)}}, \sigma, \boldsymbol{y, X}$, consider the following partition of $\boldsymbol{\beta}$ and Σ -

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_j \\ \boldsymbol{\beta_{(j)}} \end{pmatrix} \; and \sum = \begin{pmatrix} \sigma_j^2 & \boldsymbol{\sigma'}_{j(j)} \\ \boldsymbol{\sigma}_{j(j)} & \boldsymbol{\Sigma}_{j(j)} \end{pmatrix}$$

where $\sum = \sigma^2 \left(\boldsymbol{X'X}\right)^{-1}$.

If the model has no constraint then we know that

$$\beta_j|\boldsymbol{\beta_{(j)}}, \sigma, \boldsymbol{y}, \boldsymbol{X} \sim N\left(\hat{\beta}_j + \boldsymbol{\sigma'}_{j(j)}\boldsymbol{\Sigma}^{-1}_{j(j)}\left(\boldsymbol{\beta_{(j)}} - \hat{\boldsymbol{\beta}}_{(j)}\right), \sigma_j^2 - \boldsymbol{\sigma'}_{j(j)}\boldsymbol{\Sigma}^{-1}_{j(j)}\boldsymbol{\sigma}_{j(j)}\right) \tag{2}$$

Now, let us assume initially that the explanatory variable takes only non negative values. Thus the constraints $y_i \leq \boldsymbol{x_i'\beta}$,$i = 1, 2, ..., n$ are equivalent to

$$\beta_j \geq \frac{y_i - \boldsymbol{x'}_{i(j)}\boldsymbol{\beta}_{(j)}}{x_{ij}}, i = 1, ..., n \iff \beta_j \geq \max_i \left\{\frac{y_i - \boldsymbol{x'}_{i(j)}\boldsymbol{\beta}_{(j)}}{x_{ij}}\right\} \iff \beta_j \geq b\left(\boldsymbol{\beta_{(j)}}, \boldsymbol{X}, \boldsymbol{y}\right)$$

where $b\left(\boldsymbol{\beta}_{(j)}, \boldsymbol{X}, \boldsymbol{y}\right) = \underset{i}{max}\left\{\frac{y_i - \boldsymbol{x}'_{i(j)}\boldsymbol{\beta}_{(j)}}{x_{ij}}\right\}$. Thus we get-

$$\beta_j|\boldsymbol{\beta}_{(j)}, \sigma, \boldsymbol{y}, \boldsymbol{X} \sim N\left(\hat{\beta}_j + \boldsymbol{\sigma}'_{j(j)}\boldsymbol{\Sigma}^{-1}_{j(j)}\left(\boldsymbol{\beta}_{(j)} - \hat{\boldsymbol{\beta}}_{(j)}\right), \sigma_j^2 - \boldsymbol{\sigma}'_{j(j)}\boldsymbol{\Sigma}^{-1}_{j(j)}\boldsymbol{\sigma}_{j(j)}\right), \beta_j \geq b\left(\boldsymbol{\beta}_{(j)}, \boldsymbol{X}, \boldsymbol{y}\right) \tag{3}$$

- **Note :** There is a typo in the original paper regarding the full conditioal distribution of $\beta_j|\boldsymbol{\beta}_{(j)}, \sigma, \boldsymbol{y}, \boldsymbol{X}$. The term $\boldsymbol{\sigma}'_{j(j)}$ is not multiplied with $\boldsymbol{\Sigma}^{-1}_{j(j)}\left(\boldsymbol{\beta}_{(j)} - \hat{\boldsymbol{\beta}}_{(j)}\right)$ in the mean parameter; which has been corrected here.

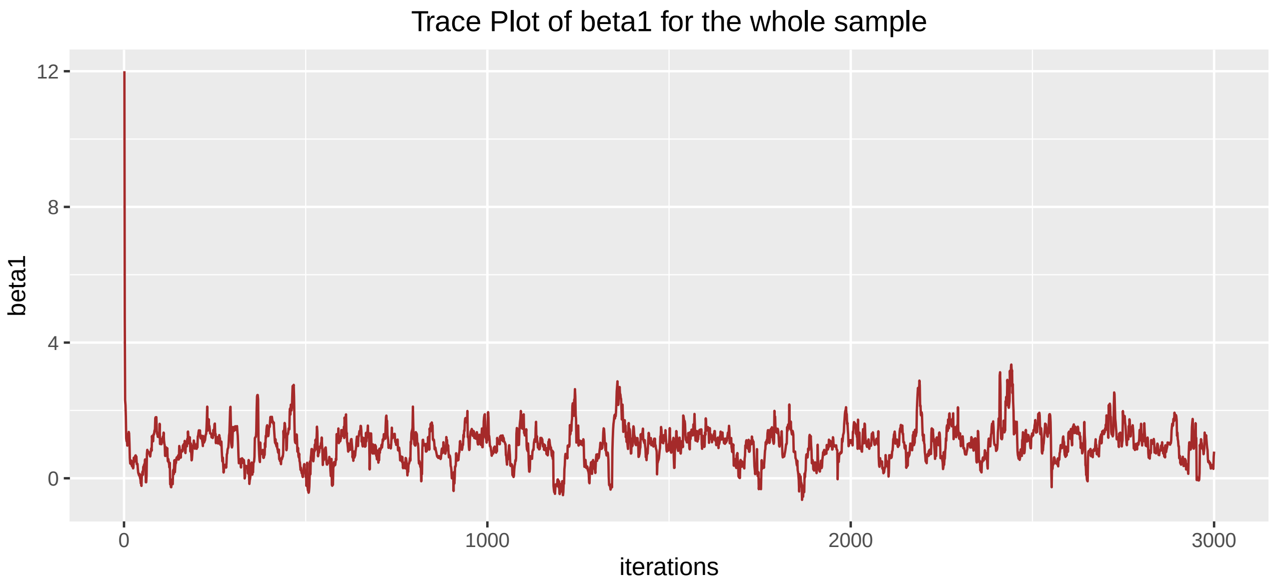
4 Simulation Study

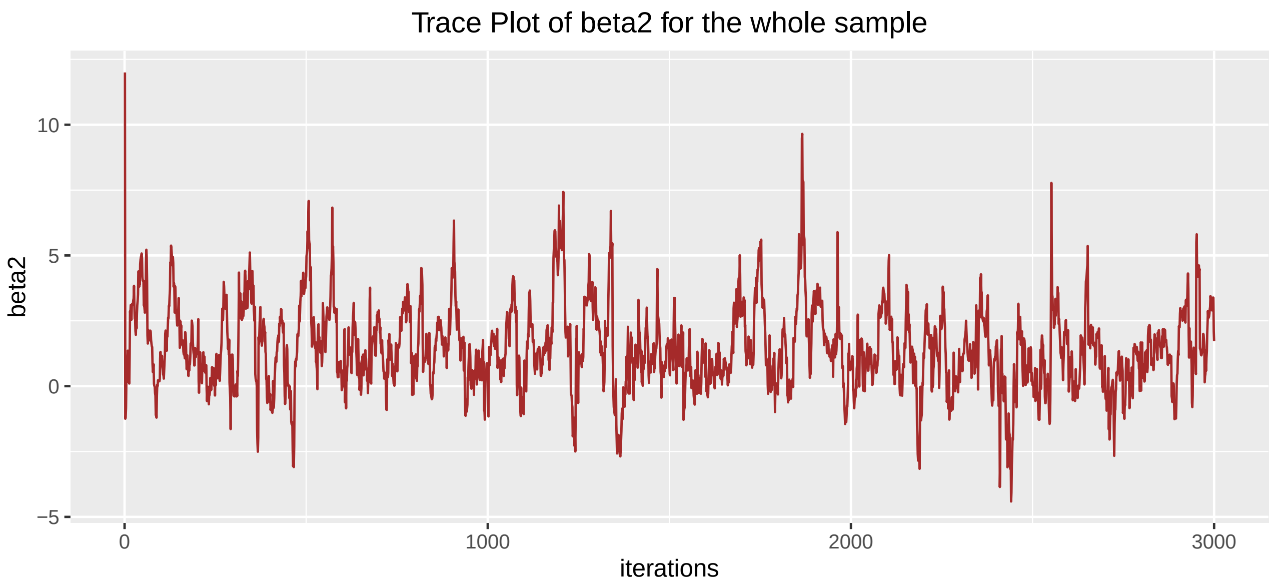
We have applied the Gibb's algorithm using R for our simulation study. We have directly simulated the posterior samples of $\boldsymbol{\beta}$ from the constrained distribution specified by (3), instead of performing any post processing of samples simulated from the unrestricted distribution specified by (2).

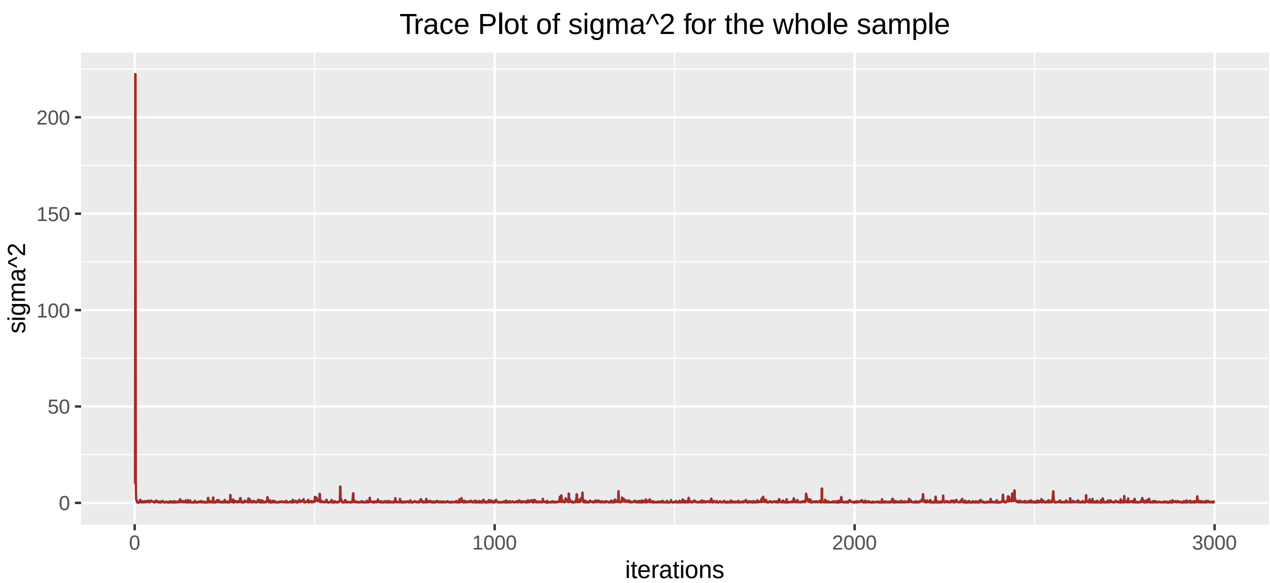
4.1 Simulation 1

- Here the model is $y_i = \beta_1 + \beta_2 x_i - \epsilon_i\,, \forall i = 1, 2, ..., n$ with the constraints on β_1, β_2 given by $y_i \leq \beta_1 + \beta_2 x_i\,, \forall i = 1, 2, ..., n$.
- The true value of the parameters are taken as $\beta_1 = \beta_2 = \sigma^2 = 1$.
- The number of datapoints (n) is taken to be 5.
- $x_i{}'s$ have been sampled from $U(0,1)$ and $\epsilon_i{}'s$ are sampled from $Half - normal(0,1)$ distribution.

First we have simulated 3000 samples from the posterior distribution of β_1, β_2 and σ^2. We have used 12,12 and 10 as the initial values of β_1, β_2 and σ^2 respectively, which are far apart from the true values. We have also checked for other initial values resulting in similar outputs. Next, we have used the traceplot of those samples to find out the point of convergence.

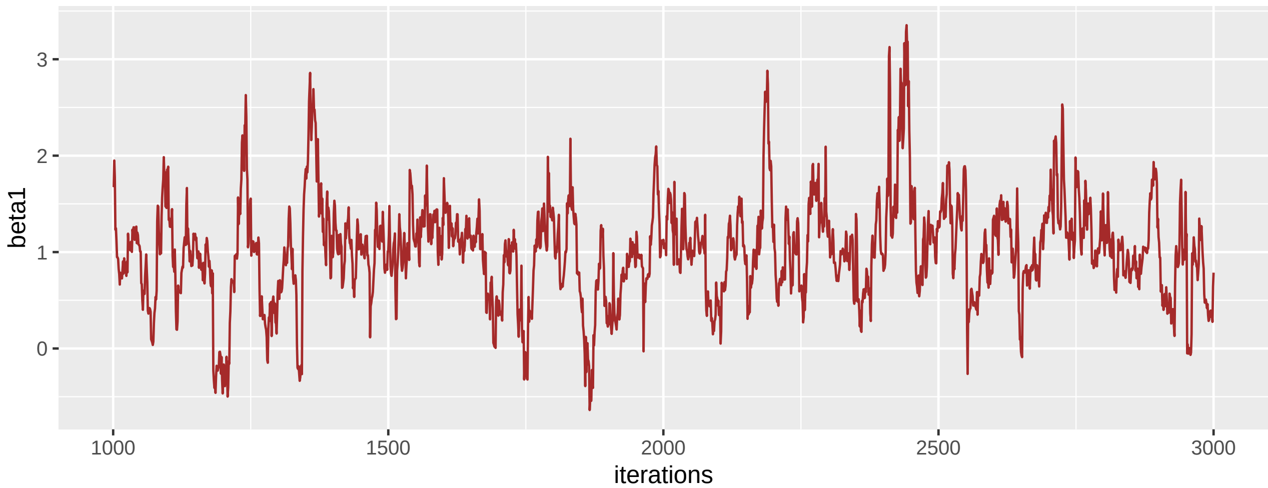

Trace Plot of beta1 for the whole sample
beta1
12
8
4
0
0
1000
2000
3000
iterations


Trace Plot of beta2 for the whole sample
beta2
10
5
0
−5
0
1000
2000
3000
iterations

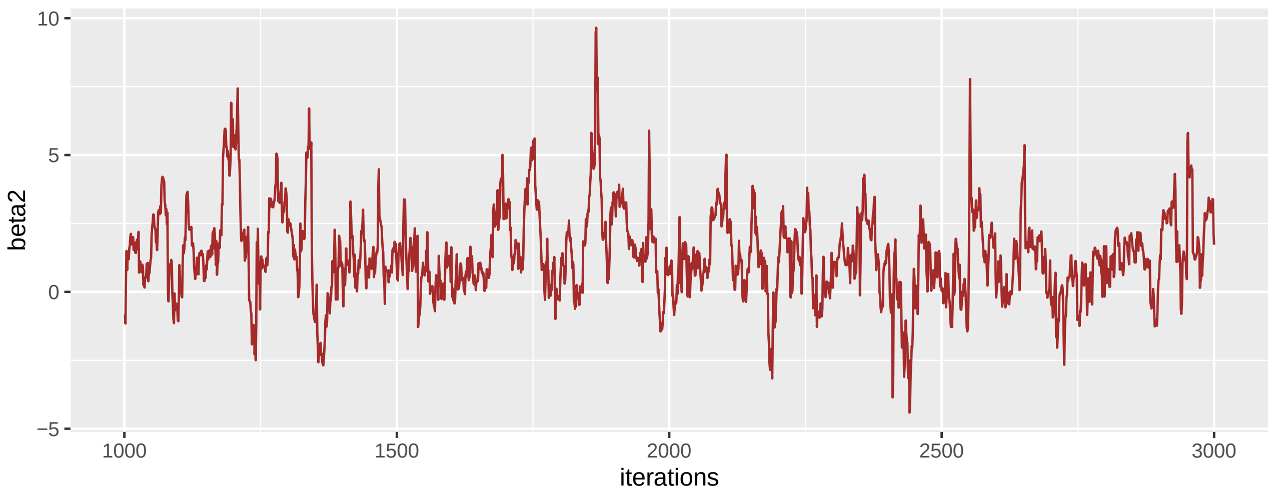

Trace Plot of sigma^2 for the whole sample
sigma^2
200
150
100
50
0
0
1000
2000
3000
iterations

- From the above three plots, we have taken the last 2000 samples as our post burn-in samples and made the traceplots of them for each of the parameters.

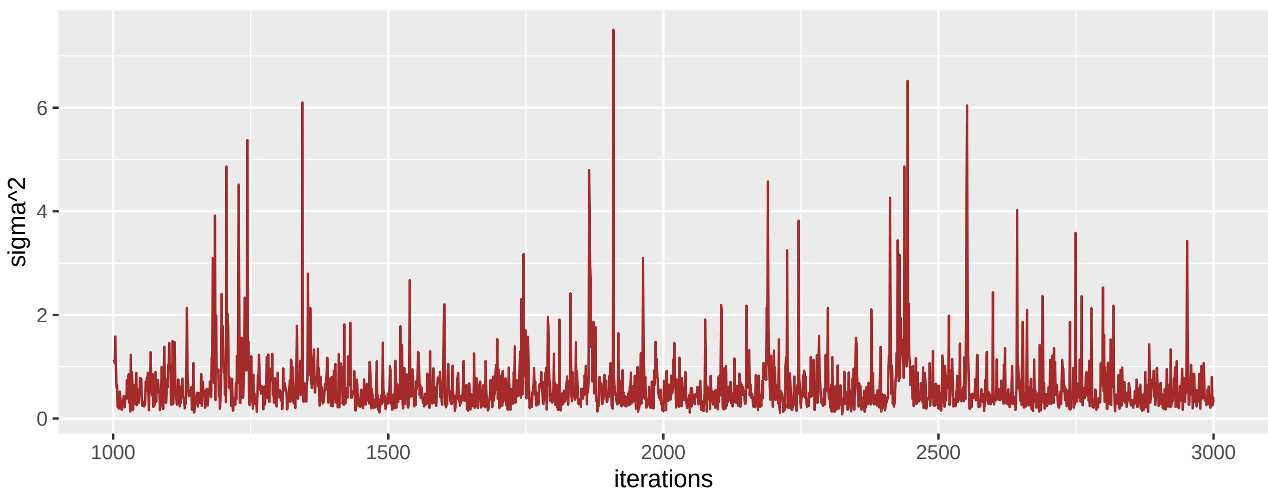
Trace Plot of beta1 for post burn−in samples



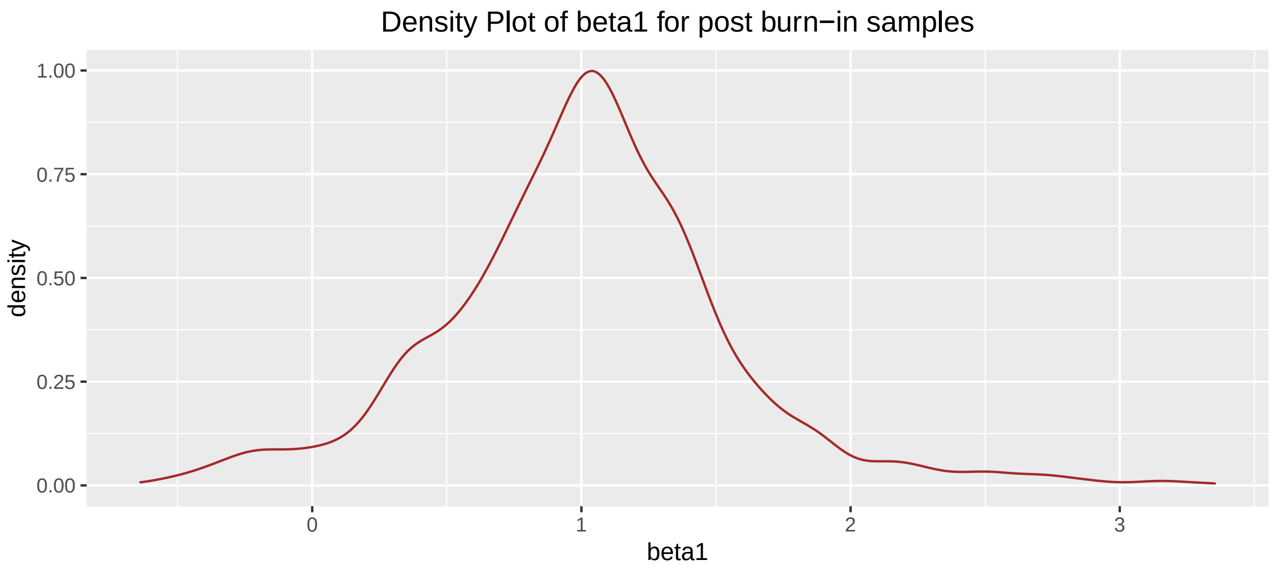
Trace Plot of beta2 for post burn−in samples

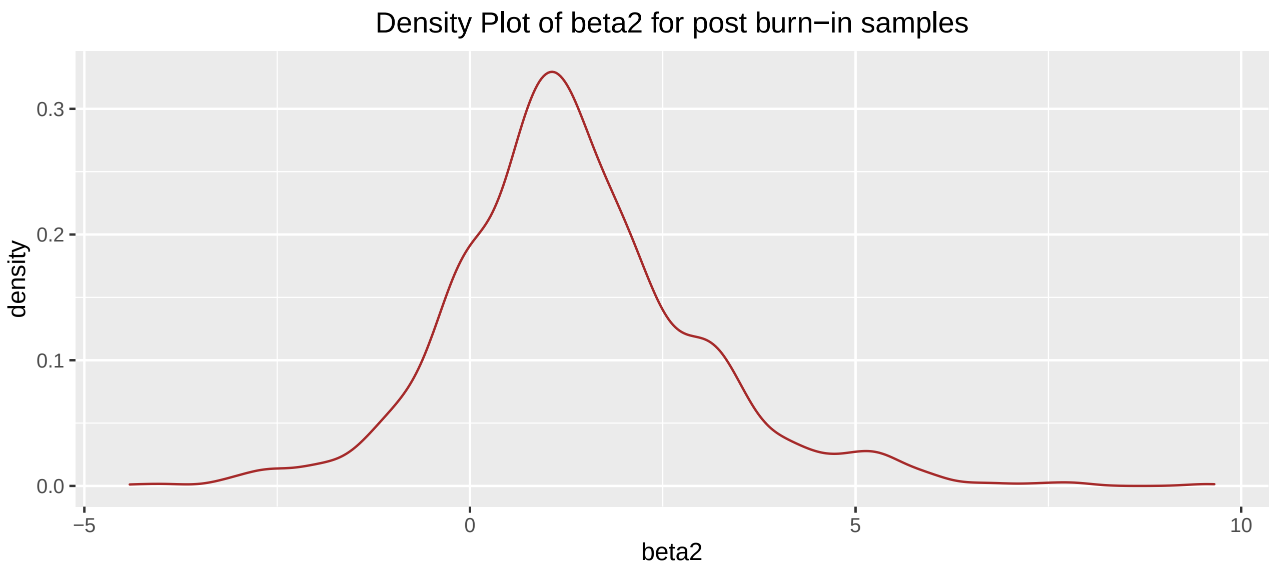


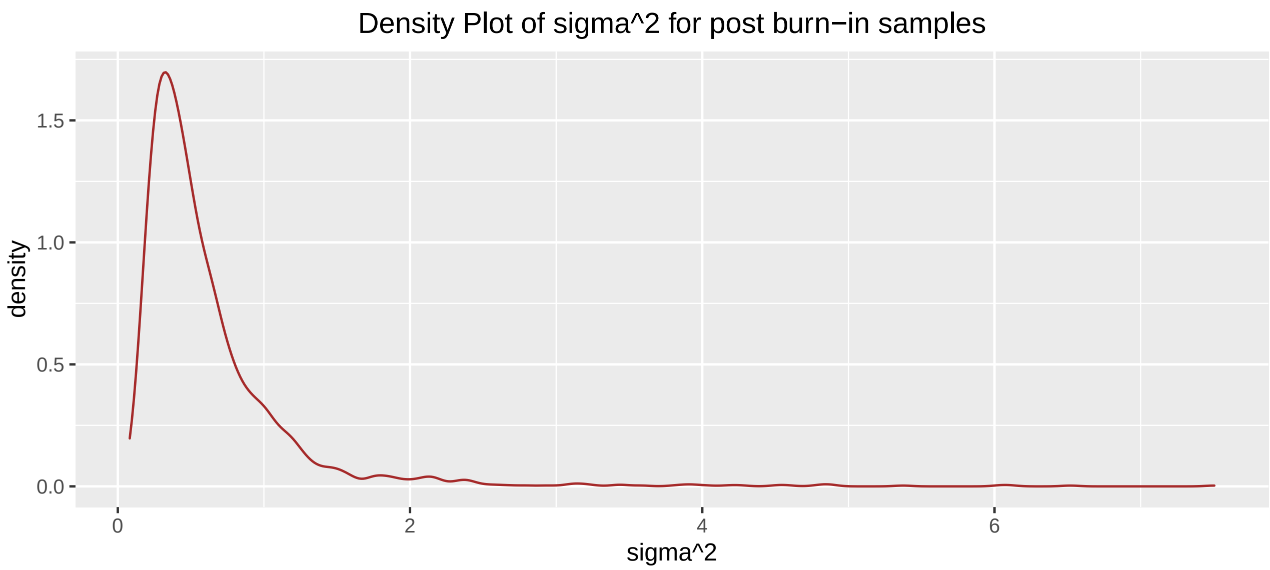
Trace Plot of sigma^2 for post burn−in samples



- The density plots of post burn-in sample observations for each of the parameters are shown below :







- Now the following R output shows the posterior summaries :

```
mean_beta1_S1

[1] 1.01171

mean_beta2_S1

[1] 1.351719

mean_sigma2_S1

[1] 0.6254774

standarderror_beta1_S1

[1] 0.5338785

standarderror_beta2_S1

[1] 1.589315

standarderror_sigma2_S1

[1] 0.5886883

CI_beta1_S1

      2.5%      97.5%
-0.1337823  2.1999268

CI_beta2_S1

     2.5%     97.5%
-1.533510  5.131628

CI_sigma2_S1

     2.5%     97.5%
0.1670719 2.1045048
```

- Now, the above R output is summarised in the following table :

Parameter	**Posterior Mean**	**Standard Error**	95% **Credible Interval**
β_1	1.01171	0.5338785	$[-0.1337823, 2.1999268]$
β_2	1.351719	1.589315	$[-1.533510, 5.131628]$
σ^2	0.6254774	0.5886883	$[0.1670719, 2.1045048]$

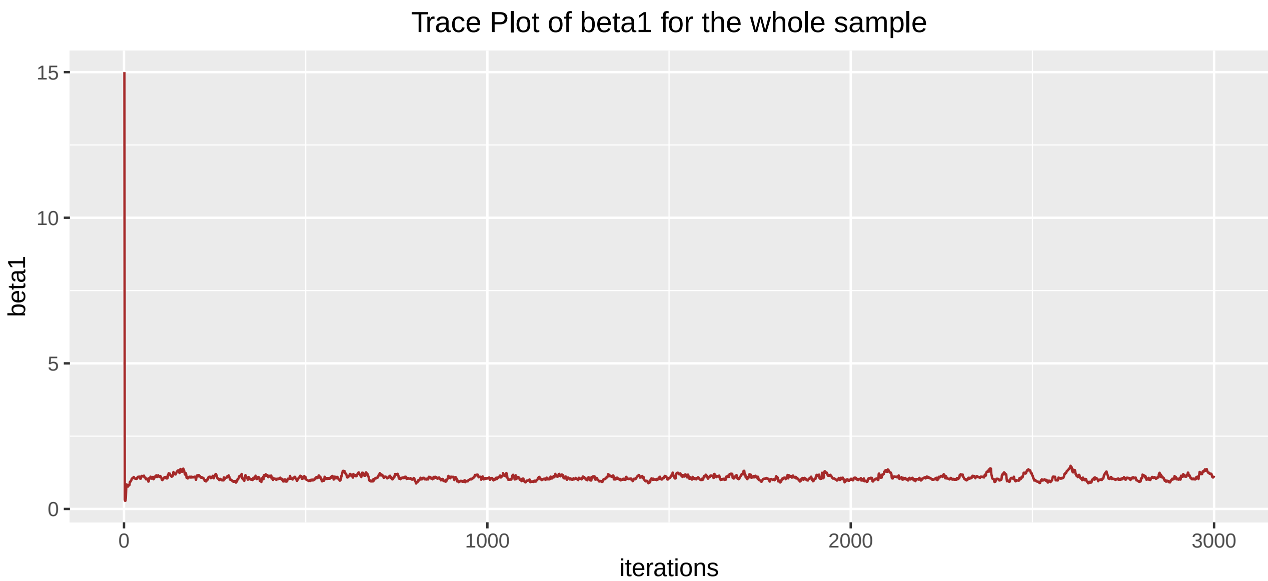
Observations :

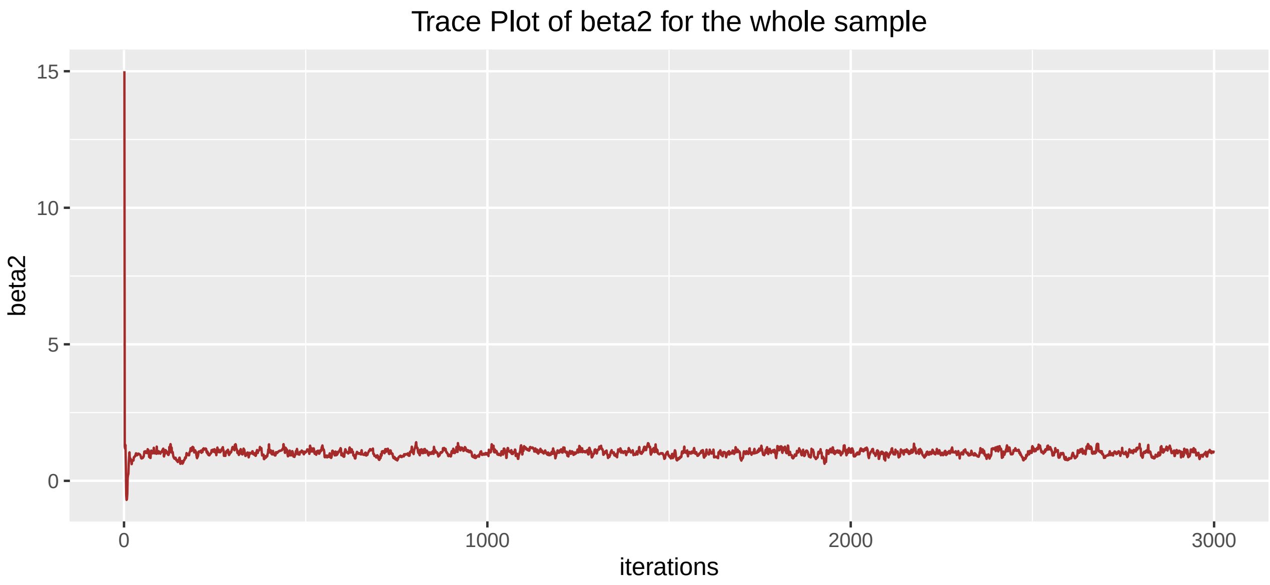
1. Observe that, the traceplots of the post burn-in samples of the three parameters look like caterpillar.
2. The density plots corresponding to the parameters β_1 and β_2 look like normal curves and the density plot corresponding to σ^2 is positively skewed and looks like Inverse Gamma pdf.
3. Note that the value of the posterior mean of β_1 is very close to 1, whereas for the other two parameters, the posterior means are not that close.
4. For $\beta_1$and σ^2, the standard error is moderately low, but in case of β_2, it is a bit high.
5. The 95% credible intervals have been calculated using the 0.025^{th} and 0.975^{th} sample quantiles. For each case, the intervals contains the true value of the parameters, though the length of the intervals are not that small.
6. It should be emphasized that, the estimates are not very accurate, because we have used only 5 datapoints.

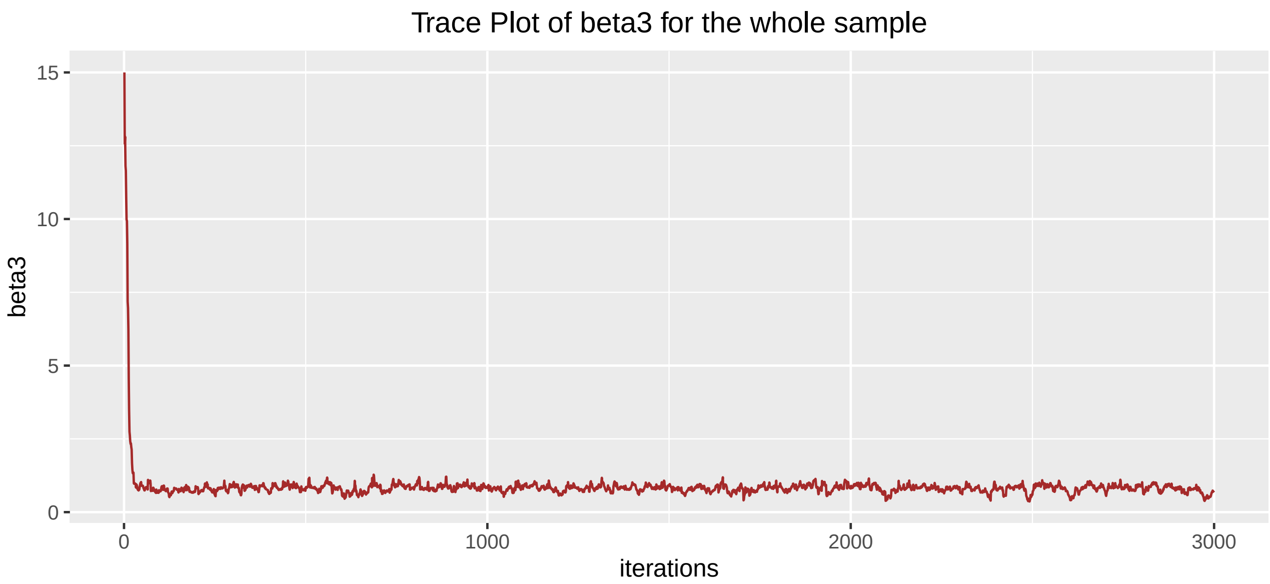
4.2 Simulation 2

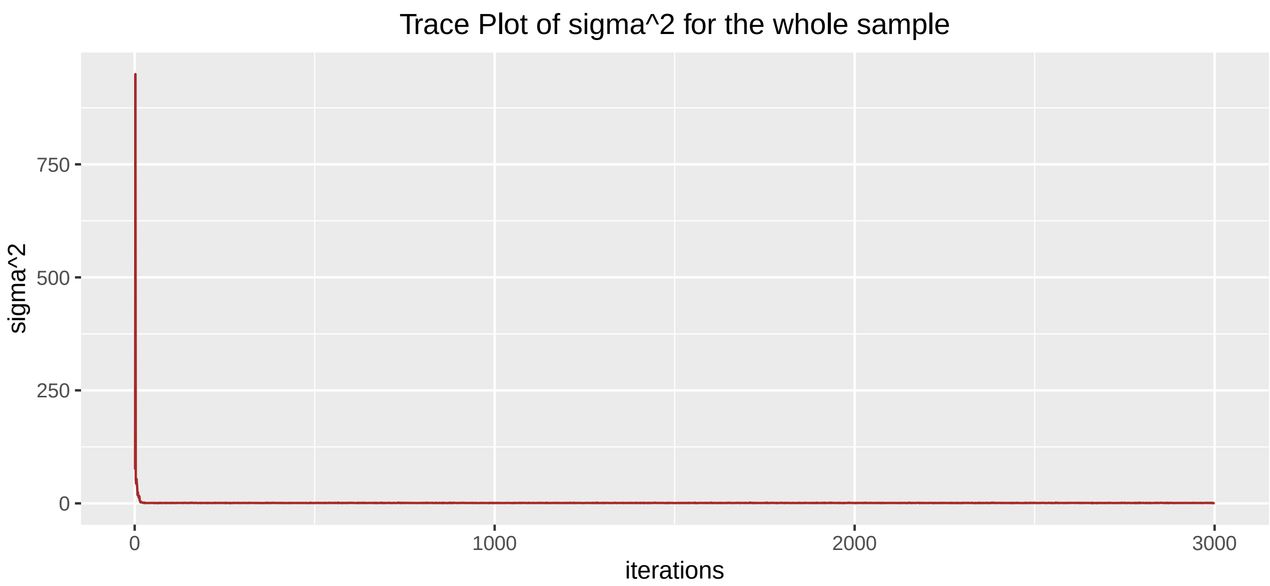
- Here the model is $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} - \epsilon_i \,, \forall i = 1, 2, ..., n$ with the constraints on $\beta_1, \beta_2, \beta_3$ given by $y_i \leq \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} \,, \forall i = 1, 2, ..., n.$
- The true value of the parameters are taken as $\beta_1 = \beta_2 = \beta_3 = \sigma^2 = 1.$
- The number of datapoints (n) is taken to be 50.
- $x_{i2}{'s}$ and $x_{i3}{'s}$ have been sampled from $U(0, 1)$ independently and $\epsilon_i{'s}$ are sampled from $Half - normal(0, 1)$ distribution.

First we have simulated 3000 samples from the posterior distribution of $\beta_1 \ , \beta_2 \ , \beta_3 \ and \ \sigma^2$. We have used 15 as the initial guess for all the three model coefficients and 75 for σ^2, which are far apart from the true values. We have also checked for other initial values resulting in similar outputs. Next, we have used the traceplot of those samples to find out the point of convergence.

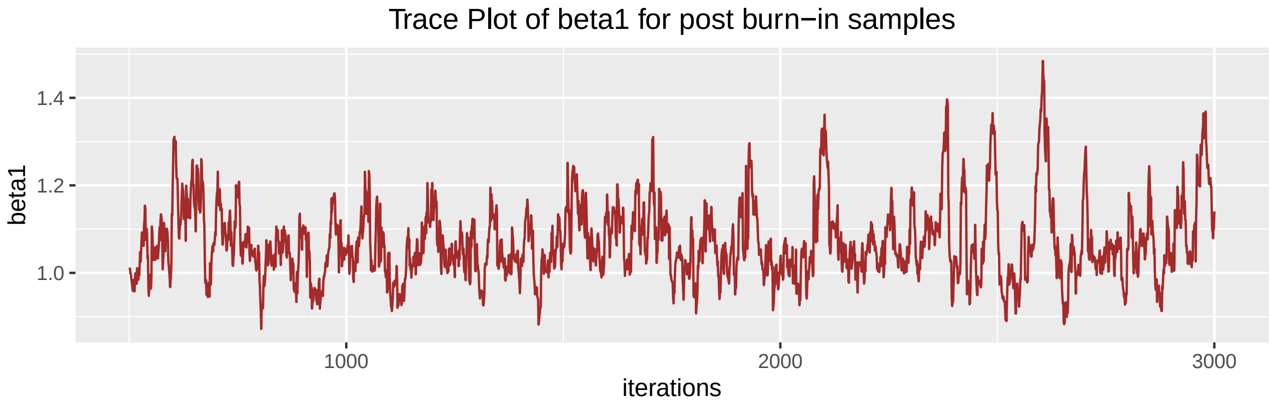


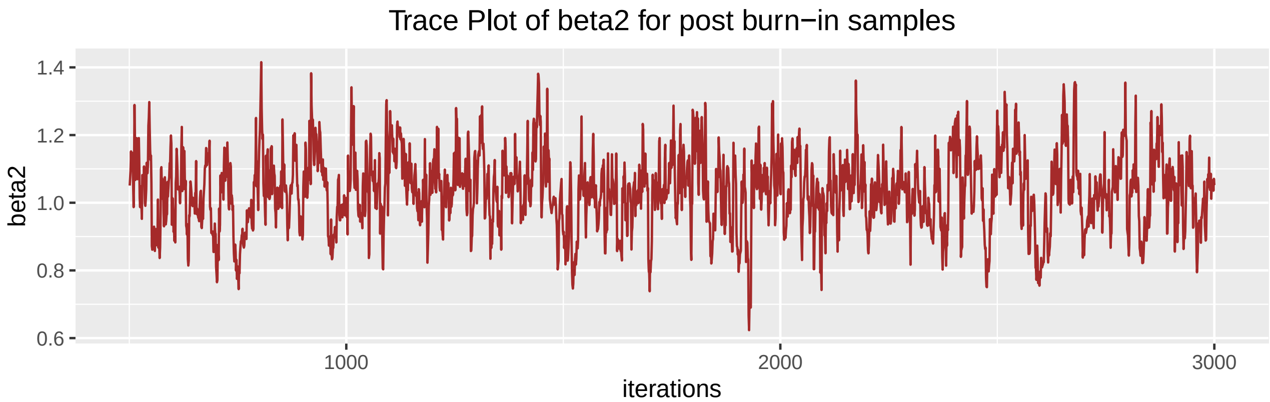

Trace Plot of beta2 for the whole sample
beta2
15
10
5
0
0
1000
2000
3000
iterations

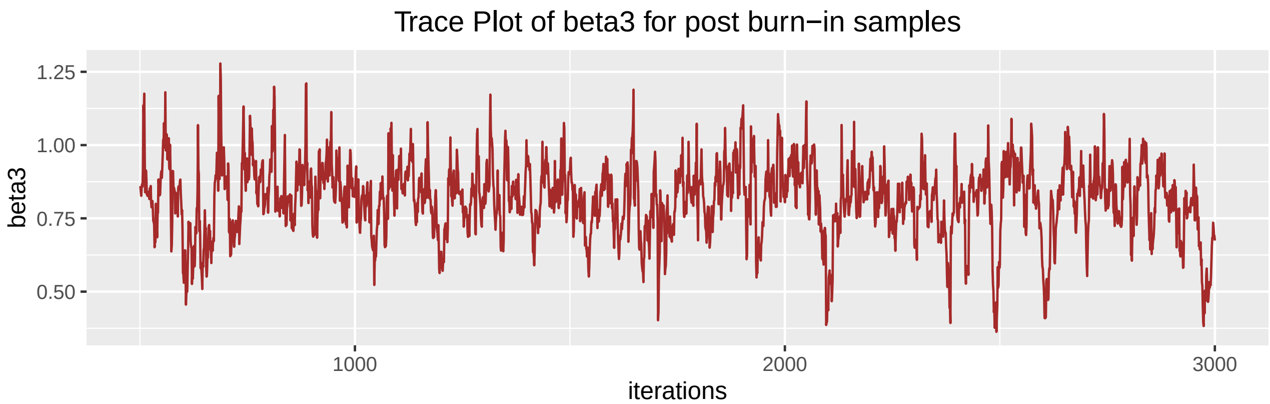

Trace Plot of beta3 for the whole sample
beta3
15
10
5
0
0
1000
2000
3000
iterations

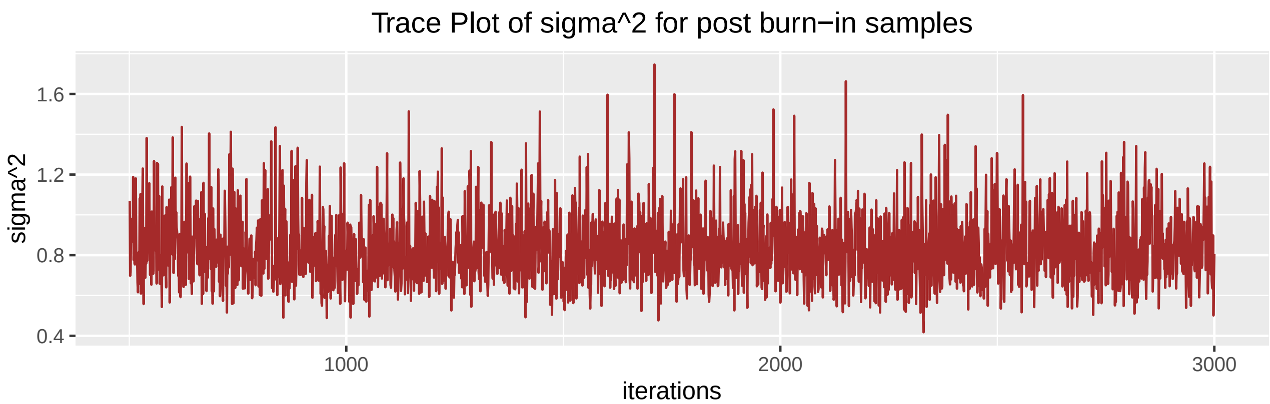

Trace Plot of sigma^2 for the whole sample
sigma^2
750
500
250
0
0
1000
2000
3000
iterations

- From the above four plots, we have taken the last 2500 samples as our post burn-in samples and made the traceplots of them for each of the parameters.

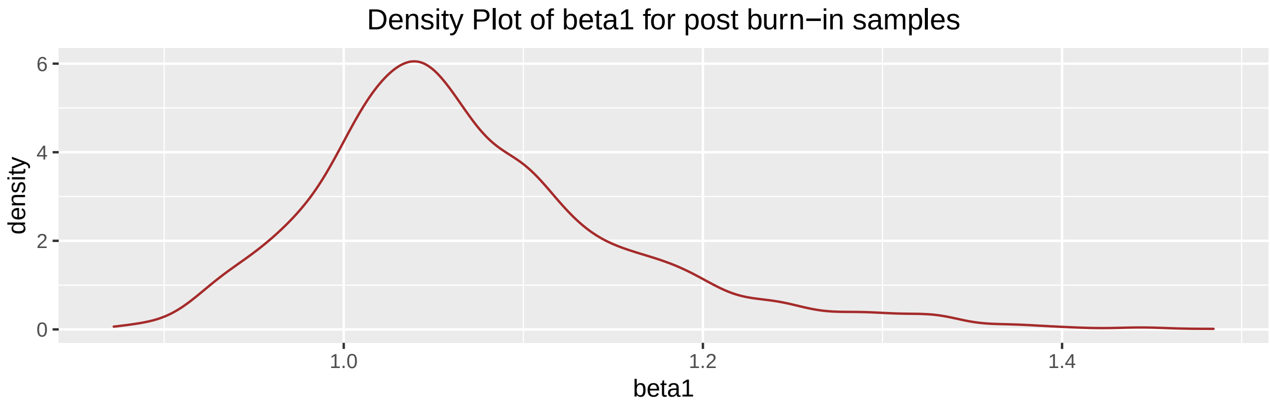


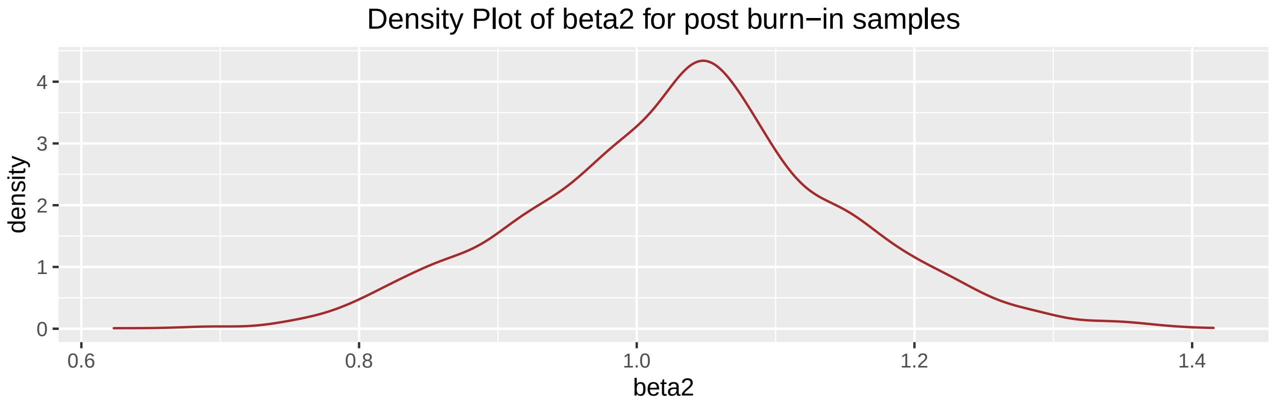


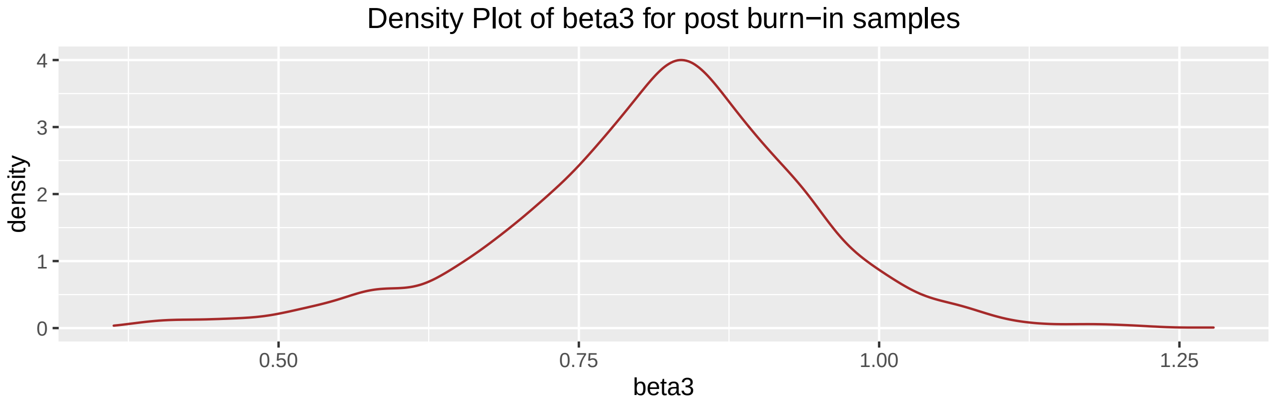


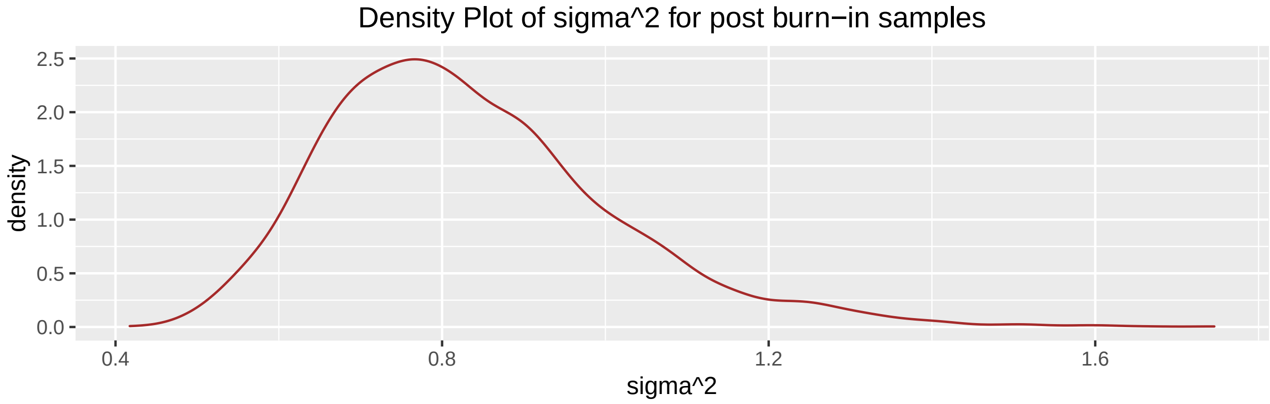


- The density plots of post burn-in sample observations for each of the parameters are shown below :









- Now the following R output shows the posterior summaries :

```
mean_beta1_S2

[1] 1.070286

mean_beta2_S2

[1] 1.035668

mean_beta3_S2

[1] 0.8150019

mean_sigma2_S2

[1] 0.8338915

standarderror_beta1_S2

[1] 0.08752781

standarderror_beta2_S2

[1] 0.1108057

standarderror_beta3_S2

[1] 0.1244527

standarderror_sigma2_S2

[1] 0.1745717
```

```
CI_beta1_S2

     2.5%     97.5%
0.9329349 1.2957241

CI_beta2_S2

    2.5%    97.5%
0.817099 1.253235

CI_beta3_S2

     2.5%     97.5%
0.5287583 1.0451003

CI_sigma2_S2

     2.5%     97.5%
0.5618761 1.2553461
```

- Now, the above R output is summarised in the following table :

Parameter	**Posterior Mean**	**Standard Error**	95% **Credible Interval**
β_1	1.070286	0.08752781	$[0.9329349, 1.2957241]$
β_2	1.035668	0.1108057	$[0.817099, 1.253235]$
β_3	0.8150019	0.1244527	$[0.5287583, 1.0451003]$
σ^2	0.8338915	0.1745717	$[0.5618761, 1.2553461]$

Observations :

1. Observe that, the traceplots of the post burn-in samples of all the four parameters look like caterpillar.

2. The density plots corresponding to the parameters β_1, β_2 and β_3 look like normal curves and the density plot corresponding to σ^2 is positively skewed and looks like Inverse Gamma pdf.

3. Note that the value of the posterior mean of β_1 and β_2 are very close to 1, whereas for the other two parameters, the posterior means are not that close.

4. For all the four parameters, the standard errors are low.

5. The 95% credible intervals have been calculated using the 0.025^{th} and 0.975^{th} sample quantiles. For each case, the intervals contains the true value of the parameters. The length of the intervals are small.

6. It should be noted that, the estimates are more accurate compared to simulation 1, because we have used more ($n = 50$) datapoints.

Further, we will study different properties (bias, MSE and the length of 95% credible intervals) of the estimators of the parameters for different number of datapoints $n = 10,\ 20,\ 50,\ 100$. We will simulate 3000 samples for each of the parameters and take the last 2000 samples as post burn-in samples for the next three simulation studies.

- In the first simulation study we will keep the same setup as in simulation 1.

- Next, we will keep the same setup as in simulation 1 except for the fact that we will generate the data from a different distribution.

- Finally, in the third simulation we will change the true values of the parameters keeping the model same as in simulation 1.

4.3 Simulation 3

- We will use the same setup as in simulation 1.
- We will study the properties of the estimators for different values of n as mentioned earlier.
- The following R output shows the results of the simulation study.

```
print(bias_b1_S3)

[1] 0.02627282 0.11309644 0.02426090 0.01997344

print(bias_b2_S3)

[1] -0.23364338 -0.12194114  0.01076527 -0.05591327

print(bias_s2_S3)

[1] -0.07952005 -0.23395595  0.34024644 -0.11341034

print(MSE_b1_S3)

[1] 0.3329384805 0.0484861709 0.0040491783 0.0008536237

print(MSE_b2_S3)

[1] 1.270857702 0.084126865 0.016193769 0.004828626

print(MSE_s2_S3)

[1] 0.22087654 0.13011067 0.19768223 0.02999356

print(matrix(CI_b1_S3,4,2,byrow=T))
```

```
          [,1]      [,2]
[1,] -0.1716239 2.170336
[2,]  0.8803573 1.601451
[3,]  0.9299623 1.166709
[4,]  0.994112  1.074009

print(matrix(CI_b2_S3,4,2,byrow=T))

          [,1]     [,2]
[1,] -1.402276 3.049227
[2,] 0.2408165 1.322035
[3,] 0.8048071 1.315568
[4,] 0.8693199 1.039311

print(matrix(CI_s2_S3,4,2,byrow=T))

          [,1]     [,2]
[1,] 0.371576  2.065102
[2,] 0.4009405 1.445798
[3,] 0.8910845 1.995362
[4,] 0.6659788 1.176438

print(len_b1_S3)

[1] 2.34196000 0.72109324 0.23674702 0.07989681

print(len_b2_S3)

[1] 4.4515032 1.0812187 0.5107606 0.1699907

print(len_s2_S3)

[1] 1.6935262 1.0448571 1.1042776 0.5104587
```

- The next table will summarise the result for the parameter β_1 :

n	**Bias**	**MSE**	95% **Credible Interval**	**Length of CI**
10	0.02627282	0.3329384805	$[-0.1716239, 2.170336]$	2.34196000
20	0.11309644	0.0484861709	$[0.8803573, 1.601451]$	0.72109324
50	0.02426090	0.0040491783	$[0.9299623, 1.166709]$	0.23674702
100	0.01997344	0.0008536237	$[0.994112, 1.074009]$	0.07989681

- The next table will summarise the result for the parameter β_2 :

n	**Bias**	**MSE**	95% **Credible Interval**	**Length of CI**
10	-0.23364338	1.270857702	$[-1.402276, 3.049227]$	4.4515032
20	-0.12194114	0.084126865	$[0.2408165, 1.322035]$	1.0812187
50	0.01076527	0.016193769	$[0.8048071, 1.315568]$	0.5107606
100	-0.05591327	0.004828626	$[0.8693199, 1.039311]$	0.1699907

- The next table will summarise the result for the parameter σ^2 :

n	**Bias**	**MSE**	95% **Credible Interval**	**Length of CI**
10	-0.07952005	0.22087654	$[0.371576, 2.065102]$	1.6935262
20	-0.23395595	0.13011067	$[0.4009405, 1.445798]$	1.0448571
50	0.34024644	0.19768223	$[0.8910845, 1.995362]$	1.1042776
100	-0.11341034	0.02999356	$[0.6659788, 1.176438]$	0.5104587

Observations :

- For β_1, there is an overall decreasing trend in bias (except for the value $n = 20$). The MSE drops drastically as number of datapoints increases. The credible intervals also get shrinked significantly as n increases. All the credible intervals contain the true value of β_1.
- For β_2, the absolute value of the bias decreases except for $n = 100$. But note that, the MSE shows a sharp decline. All the credible intervals contain the true value of the parameter and get shorter as n increases.
- Though there is no such decreasing trend in the bias of σ^2, the values are small. But the overall patterns for MSE and the length of the credible intervals are decreasing.
- So, we can conclude that as the number of datapoints increases, the estimation of the parameters gets better.

4.4 Simulation 4

- We will use the same setup as in simulation 1 except for the fact that we have sampled $x_i's$ from $U(0, 5)$.
- We will study the properties of the estimators for different values of n as mentioned earlier.
- The R output is shown below.

```
print(bias_b1_S4)

[1] -0.233897067 -0.008809642  0.038778454 -0.011028198

print(bias_b2_S4)

[1]  0.055828392  0.017140084 -0.005983789  0.008903726
```

```r
print(bias_s2_S4)
```

```
[1] -0.509956690  0.004934989  0.084185018 -0.132903347
```

```r
print(MSE_b1_S4)
```

```
[1] 0.082367214 0.031946645 0.010114098 0.001740979
```

```r
print(MSE_b2_S4)
```

```
[1] 0.0081947695 0.0043563854 0.0014590130 0.0004123185
```

```r
print(MSE_s2_S4)
```

```
[1] 0.33651411 0.13821705 0.05955174 0.03392900
```

```r
print(matrix(CI_b1_S4,4,2,byrow=T))
```

```
     [,1]      [,2]
[1,] 0.525332  1.133439
[2,] 0.6673099 1.420754
[3,] 0.8944848 1.230859
[4,] 0.8999658 1.06315
```

```r
print(matrix(CI_b2_S4,4,2,byrow=T))
```

```
     [,1]      [,2]
[1,] 0.9339806 1.208912
[2,] 0.9003239 1.173106
[3,] 0.9261233 1.068774
[4,] 0.9839941 1.052037
```

```r
print(matrix(CI_s2_S4,4,2,byrow=T))
```

```
     [,1]      [,2]
[1,] 0.1917123 1.183437
[2,] 0.5198301 1.955441
[3,] 0.7186309 1.606113
[4,] 0.6501339 1.146219

print(len_b1_S4)

[1] 0.6081075 0.7534436 0.3363742 0.1631840

print(len_b2_S4)

[1] 0.27493091 0.27278248 0.14265063 0.06804323

print(len_s2_S4)

[1] 0.9917244 1.4356111 0.8874824 0.4960851
```

- The next table will summarise the result for the parameter β_1 :

n	**Bias**	**MSE**	95% **Credible Interval**	**Length of CI**
10	-0.233897067	0.082367214	$[0.525332, 1.133439]$	0.6081075
20	-0.008809642	0.031946645	$[0.6673099, 1.420754]$	0.7534436
50	0.038778454	0.010114098	$[0.8944848, 1.230859]$	0.3363742
100	-0.011028198	0.001740979	$[0.8999658, 1.06315]$	0.1631840

- The next table will summarise the result for the parameter β_2 :

n	**Bias**	**MSE**	95% **Credible Interval**	**Length of CI**
10	0.055828392	0.0081947695	$[0.9339806, 1.208912]$	0.27493091
20	0.017140084	0.0043563854	$[0.9003239, 1.173106]$	0.27278248
50	-0.005983789	0.0014590130	$[0.9261233, 1.068774]$	0.14265063
100	0.008903726	0.0004123185	$[0.9839941, 1.052037]$	0.06804323

- The next table will summarise the result for the parameter σ^2 :

n	**Bias**	**MSE**	95% **Credible Interval**	**Length of CI**
10	-0.509956690	0.33651411	$[0.1917123, 1.183437]$	0.9917244
20	0.004934989	0.13821705	$[0.5198301, 1.955441]$	1.4356111
50	0.084185018	0.05955174	$[0.7186309, 1.606113]$	0.8874824
100	-0.132903347	0.03392900	$[0.6501339, 1.146219]$	0.4960851

Observations :

- For β_1, there is an overall decreasing trend in bias (except for the value $n = 20$). The MSE drops as number of datapoints increases. The values of both bias and MSE for each n are overall very small. The credible intervals also get shrinked as n increases (except for the value $n = 20$). All the credible intervals contain the true value of β_1.

- For β_2, the absolute values of the bias are too small and they decreases as n increases. The values of MSE are close to zero and they show a decresing pattern. All the credible intervals contain the true value of the parameter and get shorter as n increases.

- Though there is no such decreasing trend in the bias of σ^2, the values are small. But the overall pattern for MSE is decreasing. The credible intervals contain the true value of the parameter, and they show an overall decreasing trend except for $n = 20$.

- So, we can conclude that as the number of datapoints increases, the estimation of the parameters gets better.

4.5 Simulation 5

- We will use the same setup as in simulation 1 except for the fact that we have taken the true values of the parameters as $\beta_1 = \beta_2 = \sigma^2 = 9$
- As in the previous two simulation studies, the R output for different values of n is shown below.

```
print(bias_b1_S5)

[1] -0.51770215 -1.33760870  0.08411165  0.04493632

print(bias_b2_S5)

[1] 1.459696279 2.159028659 0.040743820 0.009549966

print(bias_s2_S5)

[1] -1.7817198 -4.0225273  2.5660477 -0.1454908

print(MSE_b1_S5)

[1] 1.48272151 1.90212185 0.03969815 0.01044956

print(MSE_b2_S5)

[1] 5.89629685 5.00758219 0.14650288 0.02294995

print(MSE_s2_S5)

[1] 17.336910 19.557347 12.785690  1.709038

print(matrix(CI_b1_S5,4,2,byrow=T))
```

```
     [,1]     [,2]
[1,] 6.476929 10.7048
[2,] 7.279576 8.517385
[3,] 8.809376 9.534622
[4,] 8.925771 9.293437

print(matrix(CI_b2_S5,4,2,byrow=T))

     [,1]     [,2]
[1,] 6.945267 14.26427
[2,] 9.661759 12.16614
[3,] 8.398146 9.957266
[4,] 8.662247 9.312999

print(matrix(CI_s2_S5,4,2,byrow=T))

     [,1]     [,2]
[1,] 2.850854 16.88375
[2,] 2.6522   9.475357
[3,] 7.684804 17.30762
[4,] 6.668713 11.77307

print(len_b1_S5)

[1] 4.2278684 1.2378087 0.7252463 0.3676663

print(len_b2_S5)

[1] 7.3190050 2.5043784 1.5591199 0.6507519

print(len_s2_S5)

[1] 14.032894  6.823157  9.622815  5.104362
```

- The next table will summarise the result for the parameter β_1 :

n	**Bias**	**MSE**	95% **Credible Interval**	**Length of CI**
10	-0.51770215	1.48272151	$[6.476929, 10.7048]$	4.2278684
20	-1.33760870	1.90212185	$[7.279576, 8.517385]$	1.2378087
50	0.08411165	0.03969815	$[8.809376, 9.534622]$	0.7252463
100	0.04493632	0.01044956	$[8.925771, 9.293437]$	0.3676663

- The next table will summarise the result for the parameter β_2 :

n	**Bias**	**MSE**	95% **Credible Interval**	**Length of CI**
10	1.459696279	5.89629685	$[6.945267, 14.26427]$	7.3190050
20	2.159028659	5.00758219	$[9.661759, 12.16614]$	2.5043784
50	0.040743820	0.14650288	$[8.398146, 9.957266]$	1.5591199
100	0.009549966	0.02294995	$[8.662247, 9.312999]$	0.6507519

- The next table will summarise the result for the parameter σ^2 :

n	**Bias**	**MSE**	95% **Credible Interval**	**Length of CI**
10	-1.7817198	17.336910	$[2.850854, 16.88375]$	14.032894
20	-4.0225273	19.557347	$[2.6522, 9.475357]$	6.823157
50	2.5660477	12.785690	$[7.684804, 17.30762]$	9.622815
100	-0.1454908	1.709038	$[6.668713, 11.77307]$	5.104362

Observations :

- For β_1, there is an overall decreasing trend in absolute bias (except for the value $n = 20$). The MSE drops significantly as number of datapoints increases. The values of both bias and MSE for each n are very small. The credible intervals also get shorter as n increases. All the credible intervals contain the true value of β_1(except for the value $n = 20$).

- For β_2, the values of the bias are small and they decreases as n increases (except for the value $n = 20$). The values of MSE drops significantly as number of datapoints increases. All the credible intervals contain the true value of the parameter and get shorter as n increases.

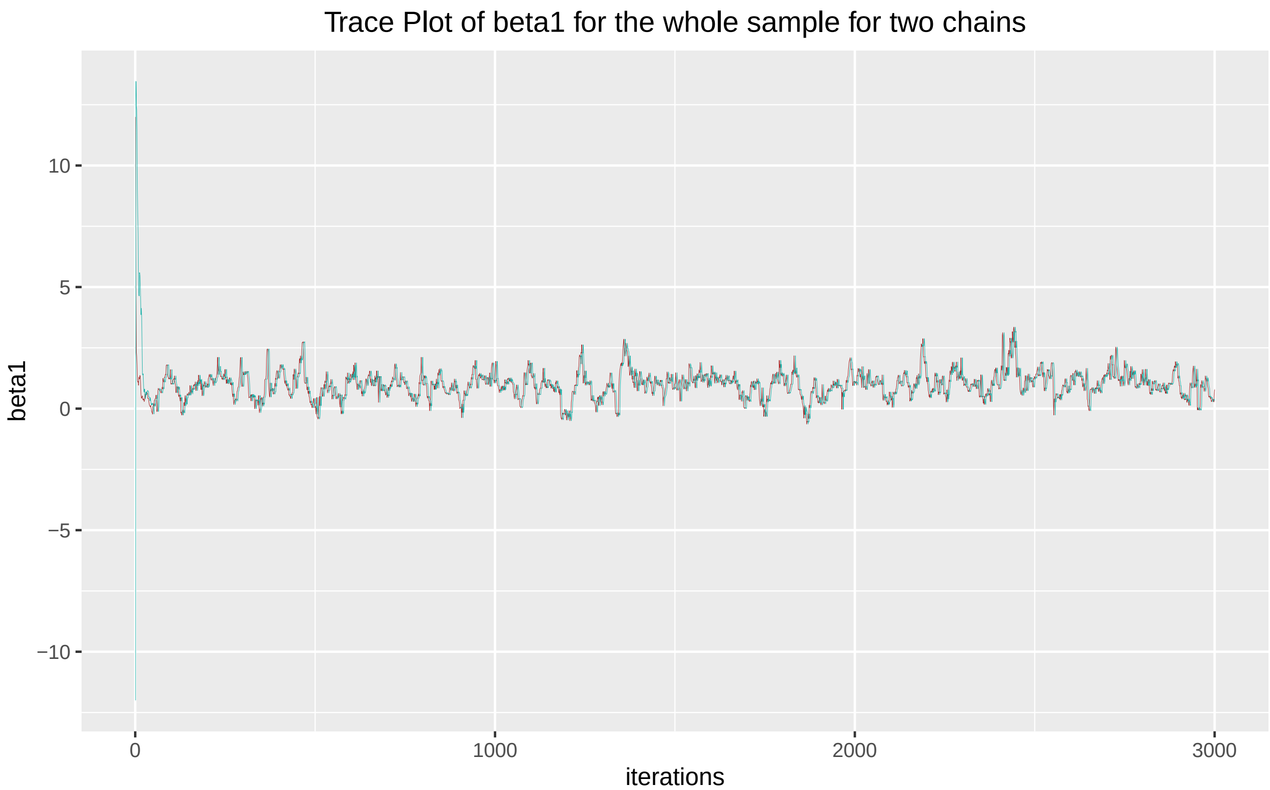
- There is no such decreasing trend in the bias of σ^2. The MSE drops strikingly for $n = 100$ compared to $n = 50$. The credible intervals contain the true value of the parameter, and they show a decreasing trend.

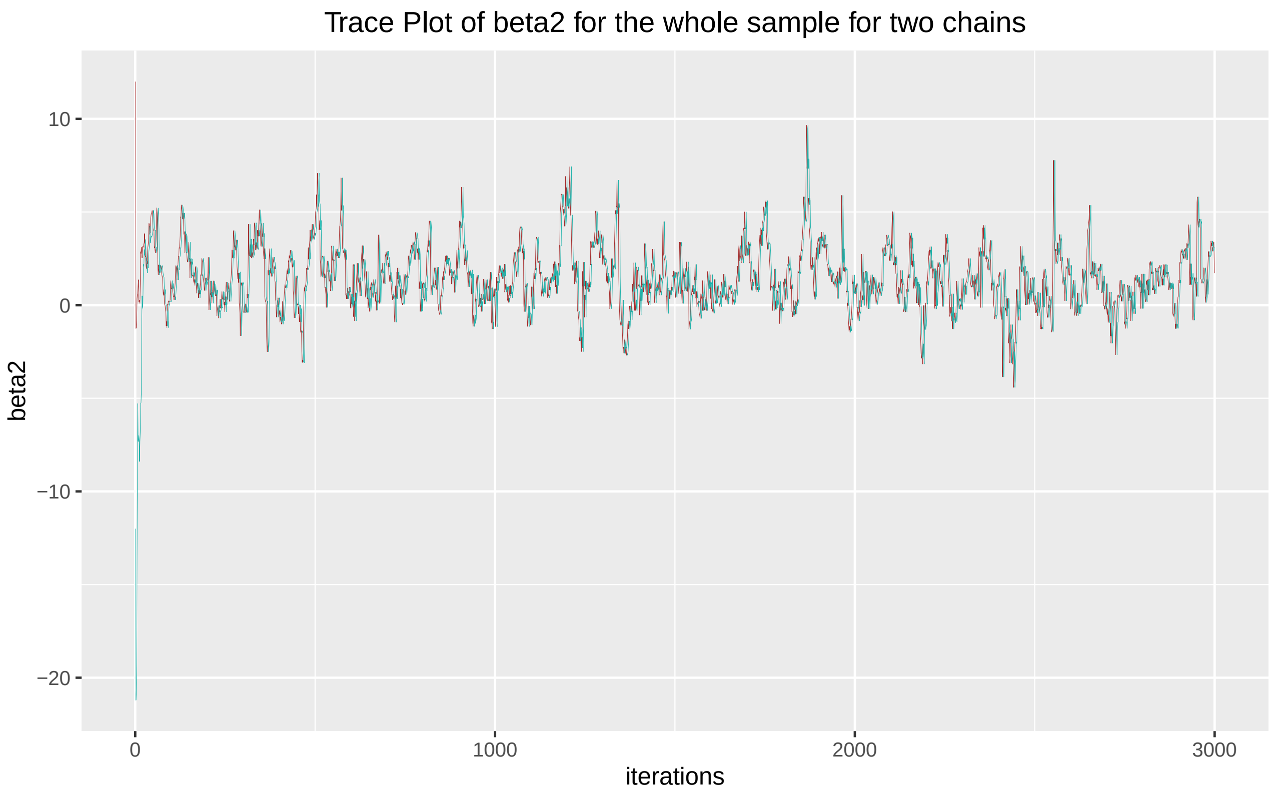
- So, as the number of datapoints increases, the estimation of the parameters gets better in this case too.

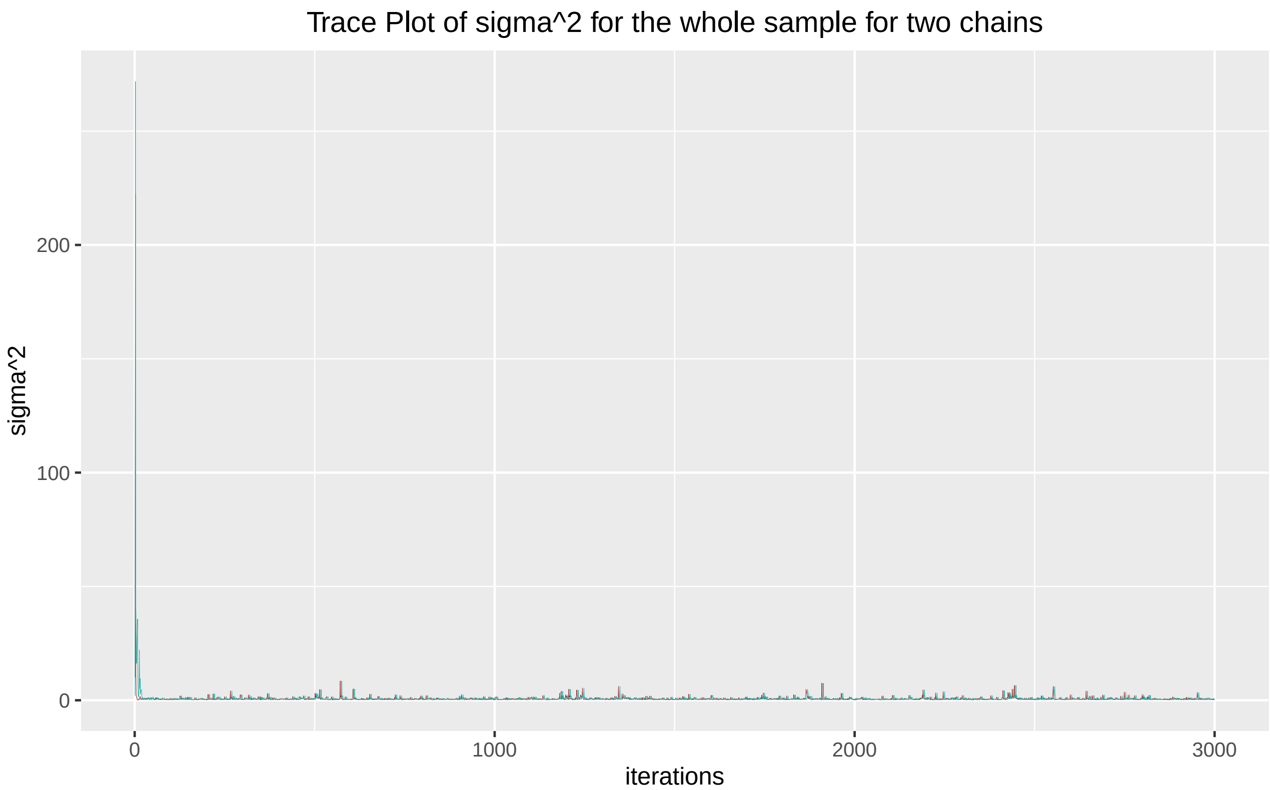
4.6 Multiple Chains

In this section we will show that for two different set of initial values of the parameters, the chains converge to the same place.Here we will consider the setup for simulation 1.

The trace plots for the whole samples are shown below. The two chains are shown by the brown and green colours.

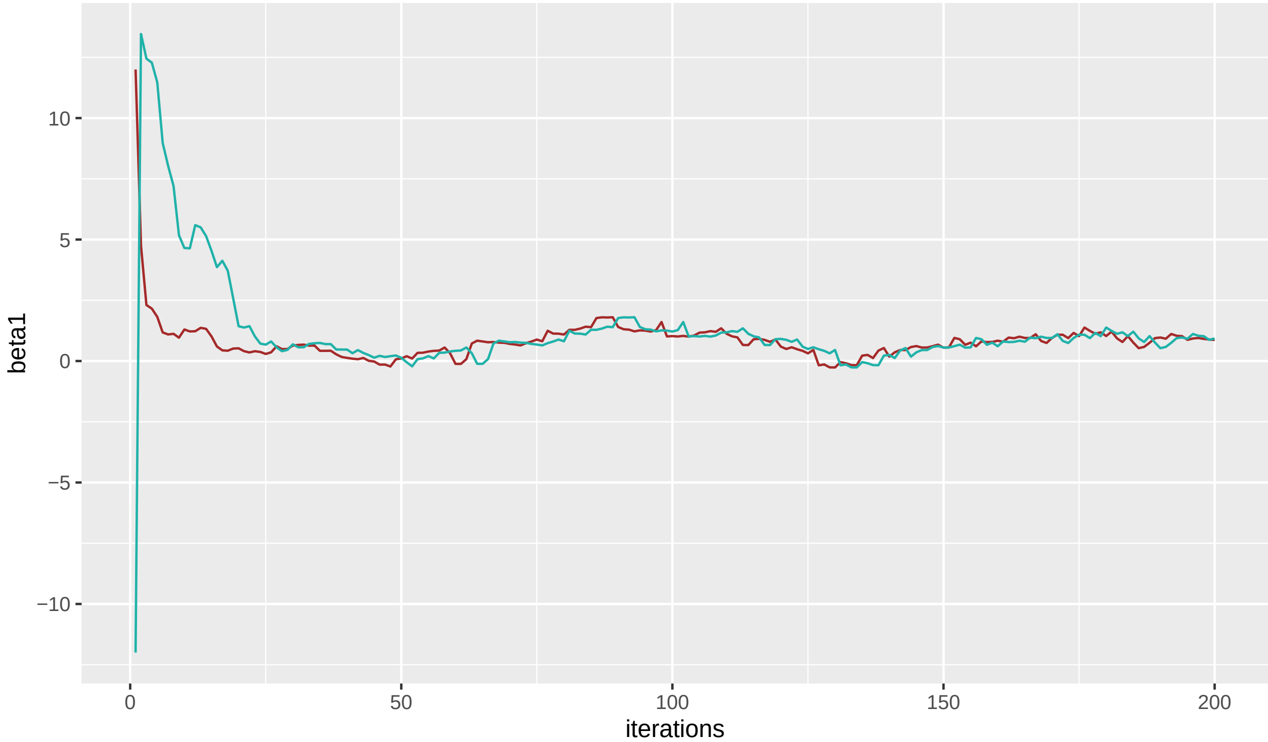



Trace Plot of beta2 for the whole sample for two chains
beta2
10
0
−10
−20
0
1000
2000
3000
iterations

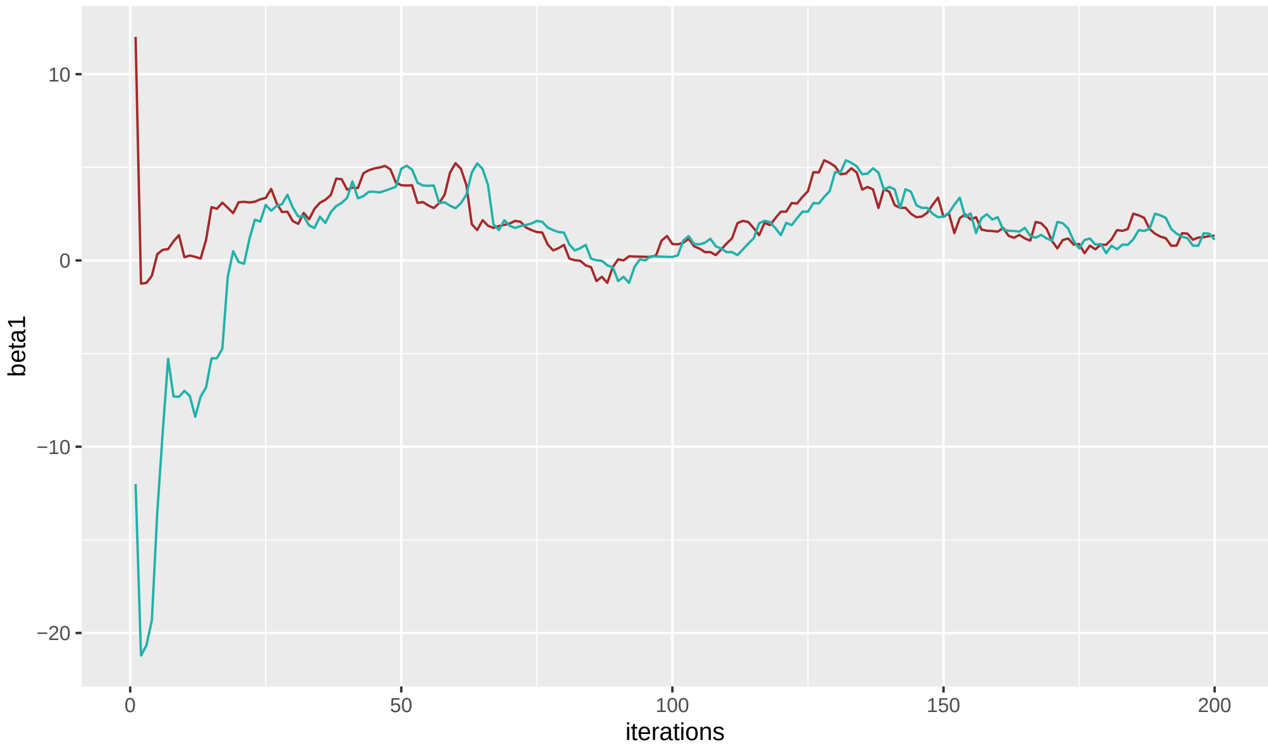

Trace Plot of sigma^2 for the whole sample for two chains
sigma^2
200
100
0
0
1000
2000
3000
iterations

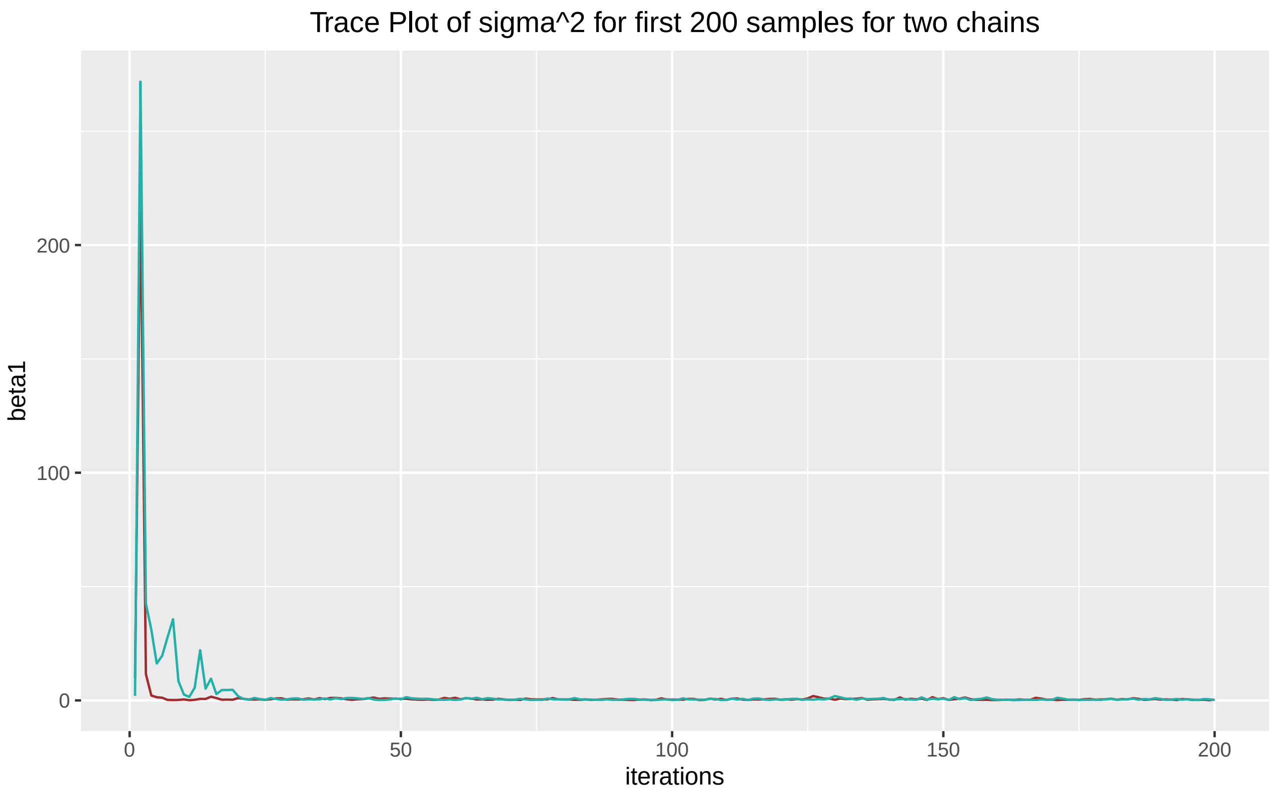
To get the enlarged view, we have plotted for the first 200 samples.

Trace Plot of beta1 for first 200 samples for two chains



Trace Plot of beta2 for first 200 samples for two chains





Observation :

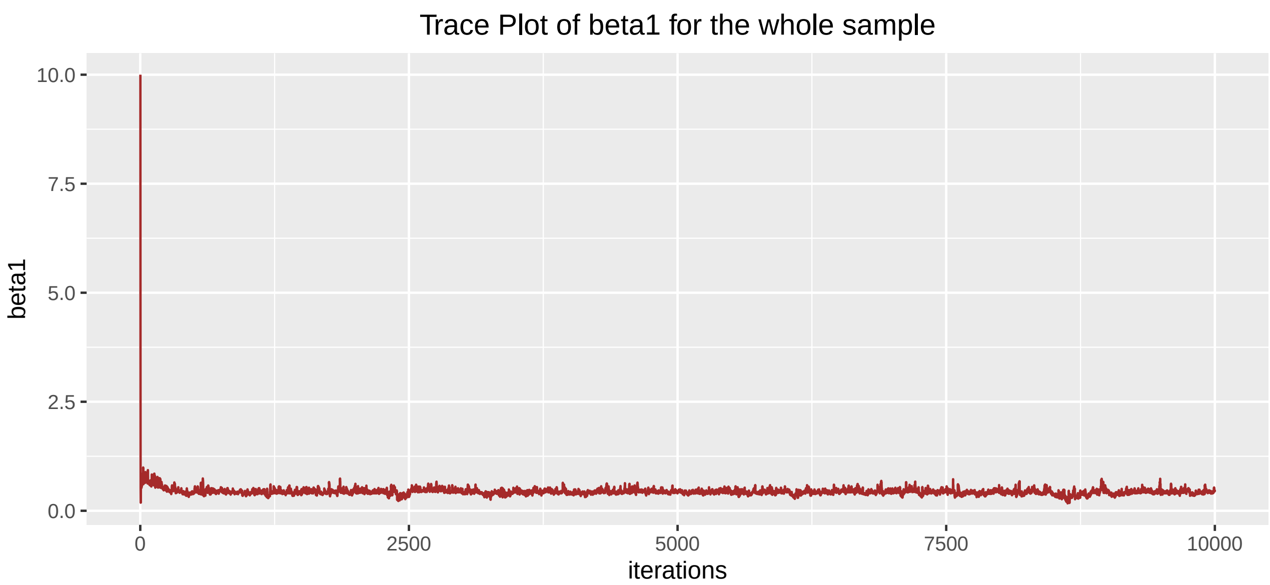
From the above plots we can see that for two different set of initial values of the parameters, the chains converge to the same value.

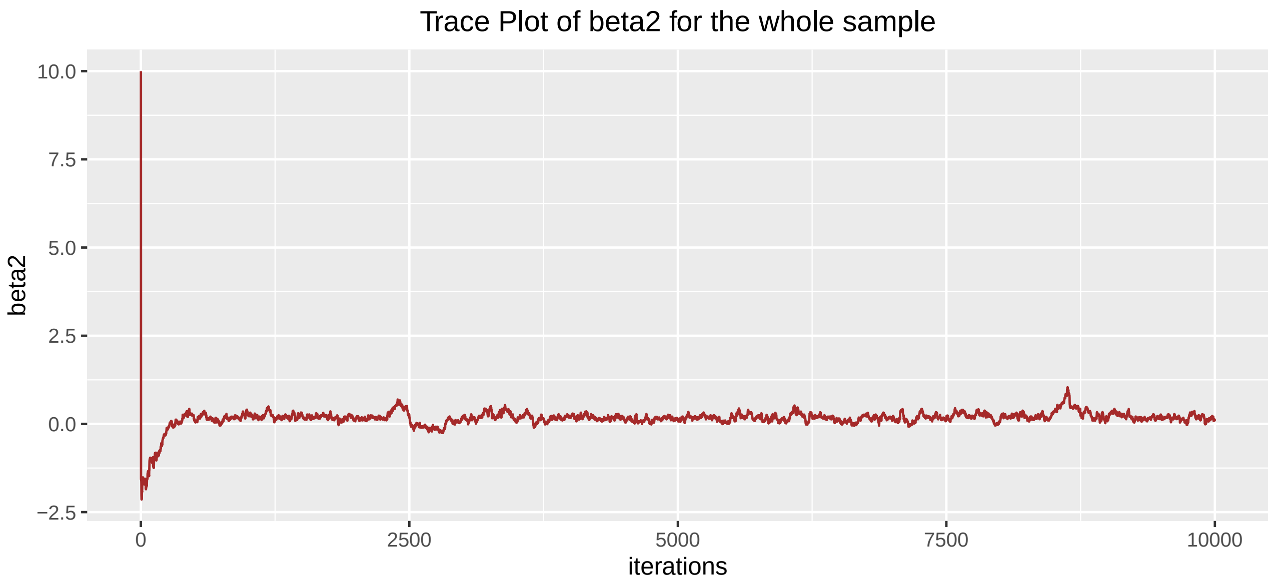
5 An Application on Telecommunication Data

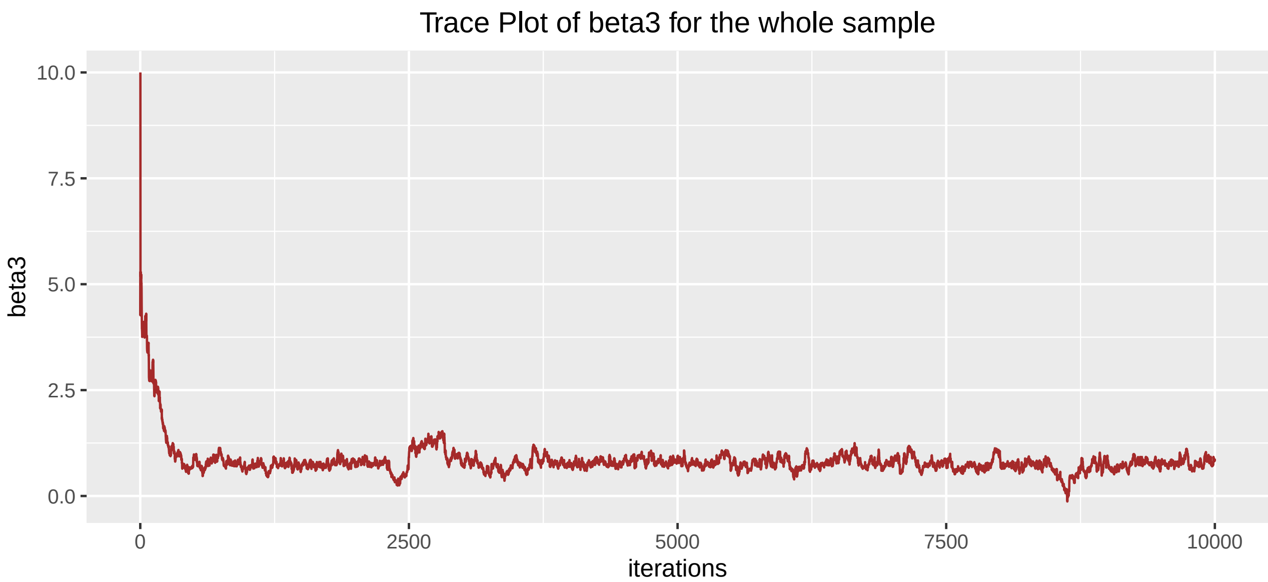
In this section, we will apply the Bayesian estimation method on the telecommunication data provided in the paper.The data refers to telecommunication companies belonging to 21 countries in 1990. The output y is is an index based on incomes and the two inputs are : the capital factor x_1 (measured through the kilometres of the lines) and the labour factor x_2 (number of employees).

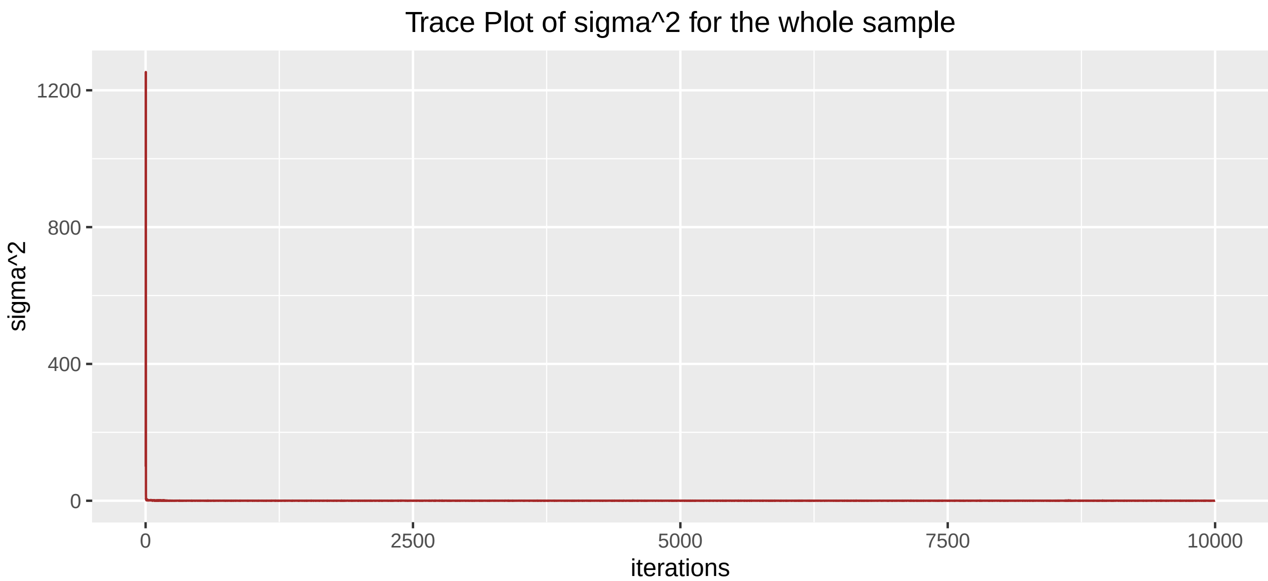
The model is $\boldsymbol{y} = \beta_1 + \beta_2\boldsymbol{x_1} + \beta_3\boldsymbol{x_2} - \boldsymbol{\epsilon}$ where the random perturbation follows $half - normal(0, \sigma^2)$ with unknown $\sigma(> 0)$.

First we have simulated 10000 samples from the posterior distribution of $\beta_1, \beta_2, \beta_3 and\ \sigma^2$. We have used 10 as the initial guess for all the three model coefficients and 100 for σ^2. Next, we have used the traceplot of those samples to find out the point of convergence.

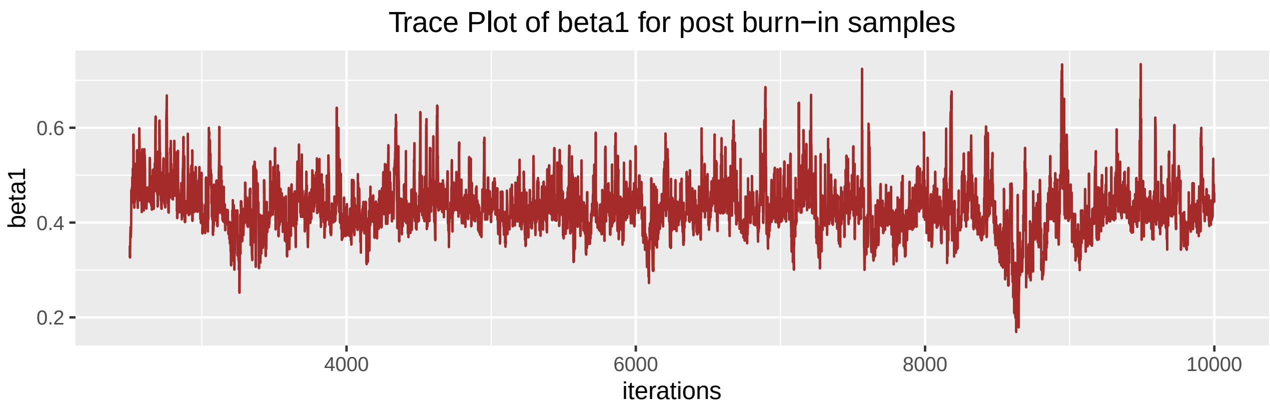


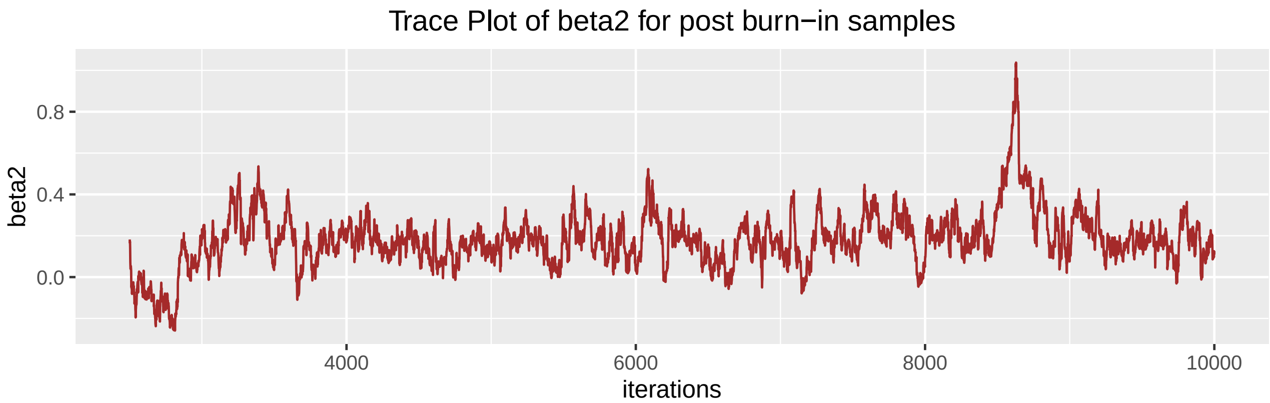

Trace Plot of beta2 for the whole sample
10.0
7.5
5.0
2.5
0.0
−2.5
beta2
0
2500
5000
7500
10000
iterations

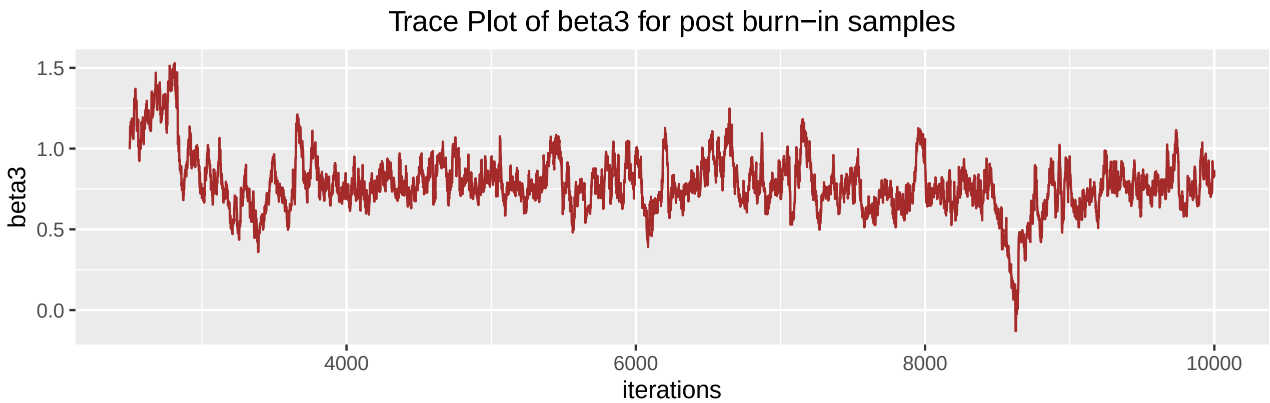

Trace Plot of beta3 for the whole sample
10.0
7.5
5.0
2.5
0.0
beta3
0
2500
5000
7500
10000
iterations

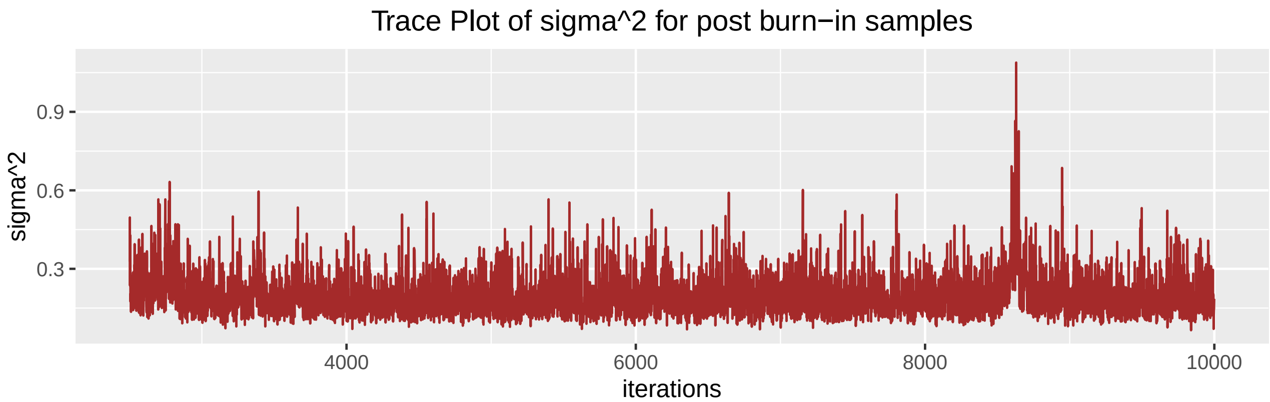

Trace Plot of sigma^2 for the whole sample
1200
800
400
0
sigma^2
0
2500
5000
7500
10000
iterations

- From the above four plots, we have taken the last 7500 samples as our post burn-in samples and made the traceplots of them for each of the parameters.

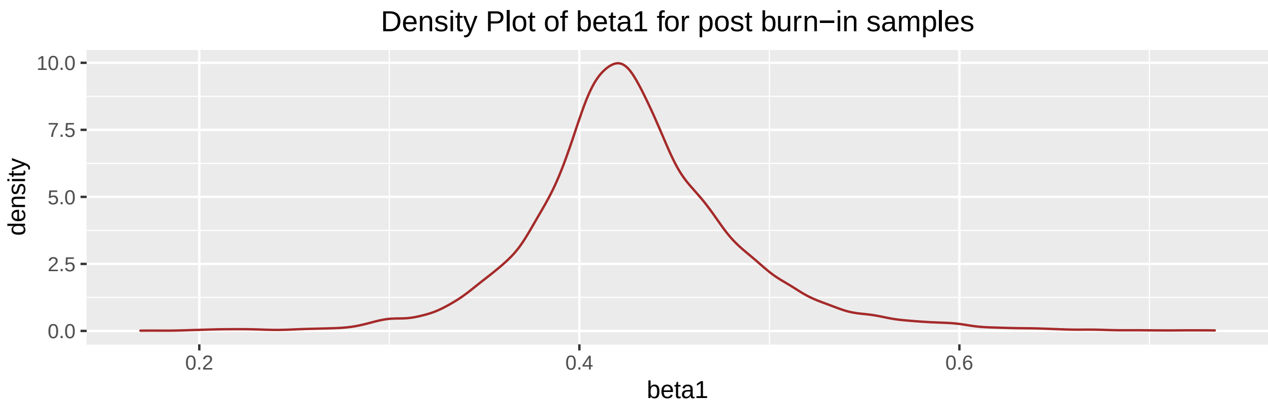


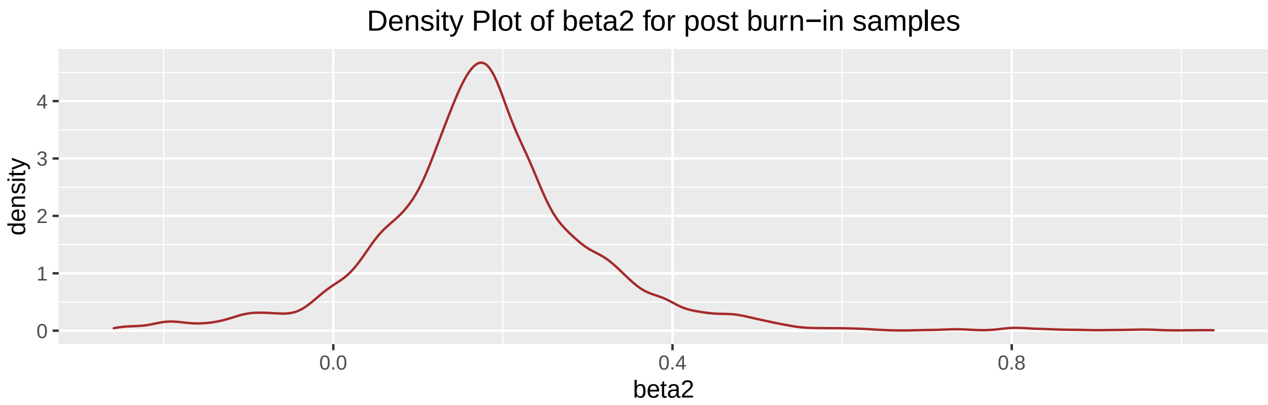


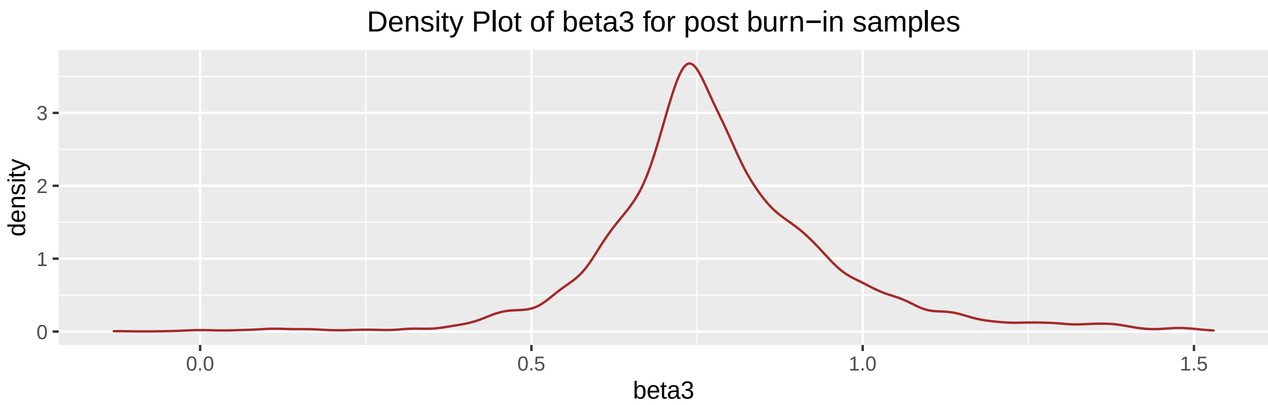


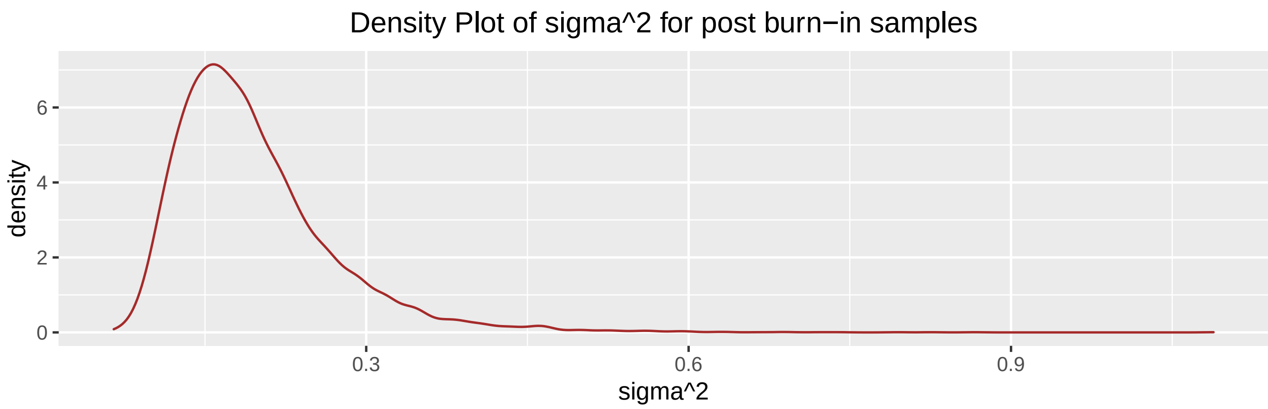


- The density plots for all the our parameters are shown below :









- Now the following R output shows the posterior summaries :

```
mean_beta1_data

[1] 0.4291172

mean_beta2_data

[1] 0.180565

mean_beta3_data

[1] 0.7825684

mean_sigma2_data

[1] 0.1961639

standarderror_beta1_data

[1] 0.05575912

standarderror_beta2_data

[1] 0.1334866

standarderror_beta3_data

[1] 0.1783307

standarderror_sigma2_data

[1] 0.07547814
```

```
CI_beta1_data

     2.5%     97.5%
0.3260002 0.5570012

CI_beta2_data

       2.5%       97.5%
-0.09559379  0.46821528

CI_beta3_data

     2.5%     97.5%
0.4586907 1.2205964

CI_sigma2_data

     2.5%     97.5%
0.1009142 0.3861886
```

- Now, the above R output is summarised in the following table :

Parameter	**Posterior Mean**	**Standard Error**	95% **Credible Interval**
β_1	0.4291172	0.05575912	$[0.3260002, 0.5570012]$
β_2	0.180565	0.1334866	$[-0.09559379, 0.46821528]$
β_3	0.7825684	0.1783307	$[0.4586907, 1.2205964]$
σ^2	0.1961639	0.07547814	$[0.1009142, 0.3861886]$

Observations :

- The standard errors of the estimates corresponding to all the four parameters are quite low.

- All the 95% credible intervals are of very short lengths.
- Here also, density plots corresponding to the parameters β_1, β_2 and β_3 look like normal curves and the density plot corresponding to σ^2 is positively skewed and looks like Inverse Gamma pdf.

6 Conclusions

- The sampling from full posterior distributions using Gibbs algorithm is straightforward and gives reasonably well estimations. Simulating directly from the posterior distribuion with the constraints (which involve both the parameters and the observations), instead of performing any post processing of samples obtained from unrestricted distribution gives sufficiently good results.
- All the simulation studies and the real data analysis show that the method works well.
- As the number of datapoints increases, the estimations are getting better. For $n = 100$, the bias, MSE are very small and credible intervals are very precise.

7 Contributions

- Paper Finding : Krishnendu Paul, Rahul Ghosh Dastidar, Saprativa Bhowal, Souraj Mazumdar
- Theory and Derivation : Krishnendu Paul, Rahul Ghosh Dastidar, Souraj Mazumdar
- Code : Souraj Mazumdar
- Project Report : Krishnendu Paul, Rahul Ghosh Dastidar, Saprativa Bhowal, Souraj Mazumdar

References

[1] Original Paper : **Bayesian estimation of the half-normal regression model with deterministic frontier** authored by **Francisco J. Ortega** and **Jose M. Gavilan**

[2] Classnotes of the course **MTH535A: AN INTRODUCTION TO BAYESIAN ANALYSIS** prepared by **Dr. Arnab Hazra**

[3] R-website