USING HETEROSCEDASTICITY COVARIANCE CONSISTENT ESTIMATORS IN PARAMETER ESTIMATION OF LINEAR REGRESSION: A BAYESIAN APPROACH

[1]OSENI, B.A., [2]OLUBUSOYE, O.E AND [3]ADEPOJU, A.A
Department of Statistics, University of Ibadan, Nigeria.

Abstract

In this paper, we derived the posterior density function of parameters of linear regression model when various structures of heteroscedasticity consistent covariance estimators developed in the earlier research in frequentist. We specify uniform distribution as out-of-sample information for the models' parameters. We obtained the marginal densities for model's parameter via Gibbs Sampler. A posterior simulator known as Gibbs Sampling and Metropolis-Hasting algorithms was employed to obtain the parameters estimate. As shown in the earlier research where the ordinary least squares (OLS) estimator was used, we recommend that Bayesian estimator is more robust and reliable for making inference for small and large sample sizes.

Keywords: *Heteroscedasticity, Posterior, Bayesian, Gibbs sampling, Noninformative prior.*

Introduction

There exists real-life situation that necessitates the establishment of some forms of relationships among several observations for making reliable and accurate decision. This type of relationship can be formalized mostly through the concept of linear regression model techniques.

Consider the multiple linear regression models that combine a response variable y and a set of covariates observation $x_1, x_2, \cdots, x_k$ of the form;

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + U$, and assume that $E(U_i^2 \mid X) = \sigma^2$ for any given observation known as homoscedasticity or equal variances. The assumption of spherical disturbance $E(UU^1) = \sigma^2 I_n$, involves double assumption that the disturbance variance is constant at each point of observation and that the disturbance covariance at all possible pairs of observation are zero. If this assumption is violated, we have unequal variances, otherwise called heteroscedasticity.

Some of the classical approach in the existing literature are Variance Stabilizing Transformation of the response variable (Box and Cox, 1964), Weighted Least Squares (WLS) by Drapper and Smith (1981) and finally the most highly appealing method of reducing the effect of heteroscedasticity on inference is to employ a heteroscedasticity-consistent standard error (HCSE) estimator of OLS parameter estimates (Hinkley, 1977, Long and Ervin, 2000, Mackinnon and White, 1985 and White, H. 1980). The thrust of this method lies in the fact that, unlike such method as WLS, it requires neither knowledge about nor a model of the functional form of the heteroscedasticity.

Many studies have been found in literature that deals with heteroscedasticity in the context of linear regression model using classical approach. Prominent among them are the work of White, H. (1980) that introduces a heteroscedasticity consistent covariance matrix estimator (HCCME)to draw correct and reliable inferences when the error variances is heteroscedasticity of unknown form, HC0. Mackinnon and White, (1985) developed an estimators HC1, HC2 and HC3 and Cribari-Neto (2004) developed HC4 estimator.

The OLS technique simply minimizes the sum of squares error of the regression model and this method is assumed to be the best linear unbiased estimator (BLUE) amongst the estimators of the regression model in frequentist (Gujarati, 2004). In Bayesian approach, wherever the researchers have relevant, available and adequate prior information about the parameter of interest of the situation being modeled, it may be necessary and sufficient to make use of such information in the estimation of the parameter of the regression model.

However, when this prior information is available, it is expected to improve the efficiency of the estimated regression model.

Furthermore, if the sample data fails to satisfy all the necessary assumptions for the use of ordinary least squares technique, the OLS estimator of variances are biased in this case, thus the usual F and $t-test$ statistics and confidence interval are no longer valid for making inferences. Also, if the functional form of prior information of the model's parameter is available, the use of such information in model's estimation may improve the efficiency of the estimators.

This paper is therefore, intends to examine the performance of Bayesian estimators of linear regression model if the assumption of homoscedasticity is violated and derived the posterior density of parameters of linear regression model by incorporating the various structures of heteroscedasticity consistent covariance estimator developed by White, H.(1980), Mackinnon and White (1985) and Cribari-Neto (2004) in frequentist and that prior information on the functional form of the regression parameters is noninformative.This paper consists of 5 sections. After the introductory section, which is section 1, section 2 discusses the Bayesian estimation of parameter in linear regression model when different structures of HCCE's HC0, HC1, HC2, HC3 and HC4 were injected. Section 3 discusses the Gibbs Sampling and data generating processes procedures with the marginal distributions of the parameters. We present the discussion of results in section 4 and finally concluding remarks are presented in section 5.

The Model Specification and Bayesian estimation

We consider regression model in which a response variable y is related to one or more explanatory or predictor variables $X_1, X_2, \cdots X_{k-1}$ defined by:

$$y = X\beta + U; \qquad U \sim N(0, \Sigma\ ) \qquad (1)$$

where,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X_t = \begin{bmatrix} 1 & X_{1,1} & \cdots & X_{1,k} \\ 1 & X_{2,1} & \cdots & X_{2,k} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n,1} & \cdots & X_{n,k} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad U = \begin{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix} \end{bmatrix} \qquad (2)$$

where y is a $(n \times 1)$ vector of the dependent observations, X is a $n \times (k+1)$ matrix of explanatory variables, β is a $(k+1) \times 1$ vector of parameters. n, k are the number of observations and parameters, respectively while U is a $(n \times 1)$ vector of disturbances, that are homoscedastic and lack of any serial correlation. That is,

$E(U \mid X) = 0$ and $E(UU^T) = \Sigma$

Bayesian Estimation

Bayesian approach to statistical inference differs from traditional frequentist b7 assuming that the data are fixed and model parameters are random, which sets up problems in the form of; what is the probability of a hypothesis (or parameter), given the data at hand? Traditional frequentist inference assumes that the model parameters are fixed (though unknown) and the data are essentially random, for instance, if the null hypothesis is true, what is the probability of this data? In general form, these types of problem can be stated as; what is the probability of the data given the hypothesis.

Bayesian estimations focus on five essential elements. First, the incorporation of prior information, that is, expert opinion, a thorough literature review of the same or similar variables, and / or prior data. Prior information is generally specified quantitatively in the form of a distribution and represents a probability distribution for a coefficient; meaning, the distribution of possible values for a coefficient we are attempting to model. Second, the prior is combined with a likelihood function. The likelihood function represents the data, that is, what is the distribution of the estimate produced by the data. Third, the combination of prior with a likelihood function to form a posterior distribution of coefficient values. Fourth, simulates are drawn from the posterior distribution to create an empirical distribution of likely values for the population parameter. Fifth, basic statistics such as mode, mean, median and so on are used to summarize the empirical distribution of simulates from the posterior represents the maximum likelihood estimate of the true coefficient's population value and credible intervals can capture the true population value with probability attached.

The Likelihood Function

A classical econometrician proceeds by obtaining data y and X and write the likelihood function of the model in (1) given by

$$L(X, y \mid \beta, \Sigma) = (2\pi)^{-\frac{n}{2}} \mid \Sigma \mid^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2}(y - X\hat{\beta})^T \Sigma^{-1}(y - X\hat{\beta}) \right\} \qquad (3)$$

The estimation is carried out using a generalized least squares method (GLS).

Another alternative method to handle the problem of heteroscedasticity when it is suspected to exist is heteroscedasticity consistent covariance estimators (HCCEs). Hinkley (1977), Long and Ervin (2000), White (1980), Mackinnon and White (1985) and Cribari-Neto (2004) among others suggested (HCCEs) of the type $\Sigma = diag(\Sigma_1, \Sigma_2, \cdots, \Sigma_n)$ to overcome the effect of heteroscedasticity on inference. In this work, we incorporate heteroscedasticity consistent covariance estimators (HCCE's) developed by:

White (1980):

$$HC0: \ \Sigma_i = U_i^2 \qquad (4)$$

Mackinnon and White (1985):

$$\begin{cases} HC1: \ \Sigma_i = \dfrac{n}{n-k}\hat{U}_i^2 \\ HC2: \ \Sigma_i = \dfrac{\hat{U}_i^2}{1-h_{ii}} \\ HC3: \ \Sigma_i = \dfrac{\hat{U}_i^2}{(1-h_{ii})^2} \end{cases} \qquad (5)$$

Cribari-Neto (2004):

$$HC4: \ \Sigma_i = \frac{\hat{U}_i^2}{(1-h_{ii})^{\phi_i}} \qquad (6)$$

where, $h_{ii} = H_{ii}$ is defined as the diagonal elements of the hat matrix, $\phi_i = \min\left\{4, \frac{nh_{ii}}{p+1}\right\}$

and Σ_i denotes unequal variance $\Sigma = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_n \end{bmatrix}$ (7)

The likelihood function of the multivariate normal density is defined by,

$$L(X, y \mid \beta, \Sigma) = \mid \Sigma \mid^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} tr(y - X\hat{\beta})^T \Sigma^{-1}(y - X\hat{\beta})\right\} \quad (8)$$

where, the trace of an $(n \times n)$ square matrix is defined to be the sum of the elements on the main diagonal (often abbreviated to "tr".)

Using the method of GLS in classical approach, the estimators of β and $\sigma^2(\Sigma)$ is given as:

$$\hat{\beta}(\Sigma) = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y \quad (9)$$

and

$$\hat{\sigma}^2(\Sigma) = \frac{(y - X\hat{\beta})^T \Sigma^{-1}(y - X\hat{\beta})}{n-k} \quad (10)$$

Based on the model assumption that rows of U are normally and independently distributed, each with zero mean vector and $(n \times n)$ covariance matrix Σ, the likelihood function for β and Σ is

$$L(X, y \mid \beta, \Sigma) = (2\pi)^{-\frac{n}{2}} \mid \Sigma \mid^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} trs^2(\Sigma^{-1}) - \frac{1}{2} tr(y - X\hat{\beta})^T \Sigma^{-1}(y - X\hat{\beta})\right\} \quad (11)$$

It proves convenient to re-write the likelihood is a slightly different way
The exponent on the right hand side in equation (11)

$$(y - X\hat{\beta})^T (y - X\hat{\beta}) = (y - \hat{\beta})^T (y - \hat{\beta}) + (\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})$$

$$= s^2 + (\beta - \hat{\beta})^T (\beta - \hat{\beta})$$

and

$(n-k)s^2 = (y - \hat{\beta})^T (y - \hat{\beta})$ where $\hat{\beta}$ is the estimate of β.

So that the likelihood function in equation (11) then becomes

$$L(X, y \mid \beta, \Sigma) = \mid \Sigma \mid^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} trs^2(\Sigma^{-1}) - \frac{1}{2} tr(\beta - \hat{\beta})^T X^T X(\beta - \hat{\beta})\Sigma^{-1} \right\} \qquad (12)$$

The Priors and their distributions

The independent priors employ in this paper are as follow;

$$p(\beta) \propto 1, \qquad \beta \in (-\infty, \infty)$$

$$p(\Sigma) \propto \Sigma^{-\frac{1}{2}},$$

The joint noninformative prior is:

$$p(\Sigma, \beta) \propto \mid \Sigma \mid^{-\frac{1}{2}} \qquad (13)$$

The Posterior Density

In Bayesian estimation procedure, probability distribution is used to quantify uncertainty but in frequentist procedure, the distribution of y is specified conditional on the parameters β, Σ;

$$y|\beta, \Sigma \sim N(X\beta, X^T\Sigma^{-1}X) \qquad (14)$$

The researcher's certainty / or uncertainty about the parameters before looking at the data is represented by the prior distribution for the parameters (β, Σ). After observing the sample data (y_i, x_i), the prior distribution is updated by the empirical data using Bayes' theorem;

$$p(\beta, \Sigma \mid y, X) = \frac{L(y, X \mid \beta, \Sigma)\, p(\beta, \Sigma)}{\int_{-\infty}^{+\infty} L(y, X \mid \beta, \Sigma)\, p(\beta, \Sigma)\, d(\beta, \Sigma)} \qquad (15)$$

Equation (15) is then becomes the posterior distribution of $p(\beta, \Sigma \mid y, X)$ of (β, Σ). The denominator of equation above acts as a normalizing constant and simply scales the posterior density which is proportional to the product of likelihood function and the prior.

The uniform priors given above are then combined with the likelihood function to give the posterior distribution given by;

$$p(\beta, \Sigma \mid X, y) \propto \mid \Sigma \mid^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} trs^2(\Sigma^{-1}) - \frac{1}{2} tr(\beta - \hat{\beta})^T X^T X(\beta - \hat{\beta})\Sigma^{-1} \right\} \Sigma^{-\frac{1}{2}}$$

$$\propto \mid \Sigma \mid^{-\frac{(n+1)}{2}} \exp\left\{ -\frac{1}{2} trs^2(\Sigma^{-1}) - \frac{1}{2} tr(\beta - \hat{\beta})^T X^T X(\beta - \hat{\beta})\Sigma^{-1} \right\} \qquad (16)$$

The marginal distribution of β and Σ i. e.; $p(\beta \mid \Sigma, X, y)$ and $p(\Sigma \mid \beta, X, y)$ respectively can be obtained by marginalizing other parameters in the posterior distribution in equation (12).

The marginal density of Σ is obtained by integrating out β, that is

$$p(\beta \mid X, y) \propto |\Sigma|^{-\frac{(n+1)}{2}} \int_{-\infty}^{\infty} \left\{ -\frac{1}{2} trs^2(\Sigma^{-1}) - \frac{1}{2} tr(\beta - \hat{\beta})^T X^T X(\beta - \hat{\beta})\Sigma^{-1} \right\} d\Sigma \quad (17)$$

$$p(\Sigma \mid X, y) \propto |\Sigma|^{-\frac{(n+1)}{2}} \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2} trs^2(\Sigma^{-1}) - \frac{1}{2} tr(\beta - \hat{\beta})^T X^T X(\beta - \hat{\beta})\Sigma^{-1} \right\} d\beta$$

$$p(\Sigma \mid X, y) \propto |\Sigma|^{-\frac{(n+1}{2}} \exp\left\{ -\frac{1}{2} trs^2(\Sigma^{-1}) \right\} \quad (18)$$

The equation (18) indicates that Σ is sampled from Wishart distribution, Σ denotes a positive definite scale matrix $(p \times p)$ which is a variance / covariance matrix from a multivariate normal distribution, n is the parameter that denotes the degree of freedom.

$$p(\beta \mid X, y) \propto |\Sigma|^{-\frac{(n+1)}{2}} \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2} trs^2(\Sigma^{-1}) - \frac{1}{2} tr(\beta - \hat{\beta})^T X^T X(\beta - \hat{\beta})\Sigma^{-1} \right\}$$

$$\propto \left[-\frac{n}{2}(\beta - \hat{\beta})^T X^T \Sigma^{-1} X(\beta - \hat{\beta}) \right] \quad (19)$$

Equation (21) shows that β is sampled from normal distribution.

Marginal Distributions Using Gibbs Sampling

Geman and Geman (1984), Gelfand and Smith (1990) amongst others employed Gibbs sampling to overcome the analytical problem encountered to obtain full posterior simulator in Bayesian estimation. The Gibbs sampler generates a random sample from marginal densities by successively sampling from the full conditional distributions of the model parameters. The full conditional distribution presented in the previous sections are presented below: $\hat{\beta}$ represents the ordinary least squares (OLS) estimate, represented as

$$\beta \sim N[\hat{\beta}, (X^T\Sigma^{-1}X)^{-1}] \quad (20)$$

and Σ indicates that:

$$\Sigma \sim Wishart[n, \Sigma] \quad (21)$$

The interest is to obtain the marginal densities of β and Σ only. The full conditional distribution is usually quite simple via Gibbs sampling proceeds as follows:

i. Set arbitrary initial values for β and Σ.
ii. Generate β from (22), and update β
iii. Generate Σ from(23) and update Σ and
iv. Repeat (i-iii) k- times, with the help of updated values, where k denotes the length of the Gibbs sequence, as $k \to \infty$, the observations from k^{th} iteration are sample observations from the appropriate marginal densities. The convergence of the samples from the marginal densities was established by Gelman and Geman (1984) and restated by Gelfand and Smith (1990) and Tierney (1991). Let the sample points be; $(\beta)_1^k, (\Sigma)_1^k, \quad (i = 1, 2, \cdots, m)$

(iv) The steps (i-iv) was repeated m- times, to generate m Gibbs samples and obtain the required samples $(\beta)_1^k, (\beta)_2^k, (\beta)_3^k, \dots, (\beta)_m^k \sim p(\beta|y)$ and $(\Sigma)_1^k, (\Sigma)_2^k, (\Sigma)_3^k, \dots, (\Sigma)_m^k \sim p(\Sigma|y)$

Data Generating Process

The Mont-Carlo experiment in this section is as follows:
For independent variable (X_0, X_1), where $X_0 = 1$ and simulate $X_1 \sim (2{,}8)$

i. The data generating process is defined as:

$$y_t = \beta_0 + \beta_1 X_t + U_t \quad ; \qquad U_t = \rho U_{t-1} + \varepsilon_t \qquad (22)$$

ii. Given ρ, generate random numbers of U_t ; $t = 1, 2, \cdots, n$ based on the assumptions: $U_t = \rho U_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0{,}1)$
iii. Given β, X_t and U_t ; $t = 1, 2, \cdots, n$ we obtain a set of data y_t, $t = 1, 2, \cdots, n$ from (20), where $(\beta_0, \beta_1) = (1{,}1)$ is assumed.
iv. We perform the experiment for different values of heteroscedasticity-consistent covariance estimators (HCCEs) with varying sample sizes $n = 15, 20, 30, 50, 100$ *and* 200.
v. Given (y_i, X_i) ; $i = 1, 2, \cdots, n$, obtain the Bayes estimates of $\hat{\beta}$ and $\hat{\Sigma}$
vi. Repeat (i)-(iii) B times, $B = 11{,}000$.

Discussion of Results

The point estimates' summary presented in Table 1-5 reflect some performances of Bayesian estimation of parameters of linear regression model in the presence of heteroscedasticity while the posterior distributions of $\hat{\beta}$ for Monte-Carlo experiments for all (HCCEs considered are given in figures 1 to 15. For all the five (5) Heteroscedasticity Consistent Covariance Estimators (HCCEs) in frequentist with varying sample sizes n=15, 20, 30, 50, 100 and 200, the performance was identical in terms of Bayes estimates and standard errors. The values of standard errors of the parameter estimates are given in parentheses. The findings reveal that the $\hat{\beta}_1$'s has positive values for all the odds sample sizes while the values are negative for all the even sample sizes for heteroscedasticity consistent covariance estimators HC0, HC1, HC2, HC3 and HC4. The distribution of these estimates was closer to normal in both small and large samples. As expected as the sample size increases, the standard error decreases. We also noticed that the standard error decreases more rapidly asymptotically as the sample sizes increases.

The results of the Bayesian regression model for the HCCEs estimators for sample sizes $n = 15, 20, 30, 50, 100$ and 200 are presented as follows:

White, H. (1980): HC0

n= 15: $y = 3.5084 + 0.0768X_1$

n= 20: $y = 2.9730 - 0.0551X_1$

n= 30: $y = 3.0732 + 0.0607X_1$

n= 50: $y = 0.0378 + 0.1638X_1$

n= 100: $y = 3.7345 - 0.1409X_1$

n= 200: $y = 5.4929 - 0.5677X_1$

Mackinnon and White (1985):

For HC1:

n= 15: $y = 3.4964 + 0.0785X_1$

n= 20: $y = 2.9815 - 0.0544X_1$

n= 30: $y = 3.0493 + 0.0643X_1$

n= 50: $y = 2.0537 + 0.1621X_1$

n= 100: $y = 3.7345 - 0.1141X_1$

n= 200: $y = 5.4929 - 0.5677X_1$

For HC2:

n= 15: $y = 3.5025 + 0.0774X_1$

n= 20: $y = 2.9944 - 0.0546X_1$

n= 30: $y = 3.0525 + 0.0639X_1$

n= 50: $y = 2.0544 + 0.1623X_1$

n= 100: $y = 3.7358 - 0.1414X_1$

n= 200: $y = 5.4886 - 0.5669X_1$

For HC3:

n= 15: $y = 3.4751 + 0.0807X_1$

n= 20: $y = 2.9701 - 0.0501X_1$

n= 30: $y = 3.0538 + 0.0647X_1$

n= 50: $y = 2.0544 + 0.1628X_1$

n= 100: $y = 2.0550 - 0.1418X_1$

n= 200: $y = 3.7372 - 0.5661X_1$

Cribari-Neto (2004): HC4

n= 15: $y = 3.442 + 0.0799X_1$

n= 20: $y = 2.9877 - 0.0544X_1$

n= 30: $y = 3.0509 + 0.0641X_1$

n= 50: $y = 2.0540 + 0.1623X_1$

n= 100: $y = 3.7394 - 0.1424X_1$

n= 200: $y = 5.4929 - 0.5664X_1$

Table 1: Summary of Monte Carlo Experiment: HC0

Sample	$\hat{\beta}_0$	$\hat{\beta}_1$
$n = 15$	3.5084 (1.0587)	0.0768 (0.1718)
$n = 20$	2.9730 (0.9012)	–0.0551 (0.1483)
$n = 30$	3.0732 (0.8190)	0.0607 (0.1375)
$n = 50$	2.0378 (0.6858)	0.1638 (0.1221)
$n = 100$	3.7345 (0.5073)	–0.1409 (0.0877)
$n = 200$	5.4929 (0.3458)	–0.5677 (0.0592)

The values of standard deviation (S.D) are given in the parenthesis

Table 2: Summary of Monte Carlo Experiment: HC1

Sample	$\hat{\beta}_0$	$\hat{\beta}_1$
$n = 15$	3.4964 (1.1298)	0.0785 (0.1830)
$n = 20$	2.9815 (0.9619)	–0.0544 (0.1578)

$n=30$	3.0493 (0.8365)	0.0643 (0.1411)
$n=50$	2.0537 (0.7008)	0.1621 (0.1229)
$n=100$	3.7345 (0.5124)	–0.1141 (0.0889)
$n=200$	5.4929 (0.3476)	–0.5677 (0.0600)

The values of standard deviation (S.D) are given in the parenthesis

Table 3: Summary of Monte Carlo Experiment: HC2

Sample	$\hat{\beta}_0$	$\hat{\beta}_1$
$n=15$	3.5025 (1.2012)	0.0774 (0.1967)
$n=20$	2.9944 (1.0255)	–0.0546 (0.11691)
$n=30$	3.0525 (0.88752)	0.0639 (0.1480)
$n=50$	2.0544 (0.7210)	0.1623 (0.1269)
$n=100$	3.7358 (0.5196)	–0.1414 (0.0900)
$n=200$	5.4886 (0.3501)	–0.5669 (0.0608)

The values of standard deviation (S.D) are given in the parenthesis

Table 4: Summary of Monte Carlo Experiment: HC3

Sample	$\hat{\beta}_0$	$\hat{\beta}_1$
$n=15$	3.4751 (1.3111)	0.0807 (0.2156)
$n=20$	2.9701 (1.0939)	–0.0501 (0.1814)
$n=30$	3.0538 (0.9179)	0.0647 (0.1562)
$n=50$	2.0544 (0.7210)	0.1628 (0.1304)
$n=100$	2.0550 (0.7421)	–0.1418 (0.0917)
$n=200$	3.7372 (0.5269)	–0.5661 (0.0608)

The values of standard deviation (S.D) are given in the parenthesis

Table 5: Summary of Monte Carlo Experiment: HC4

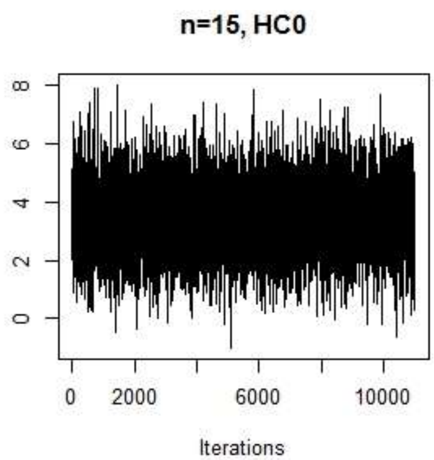
Sample	$\hat{\beta}_0$	$\hat{\beta}_1$

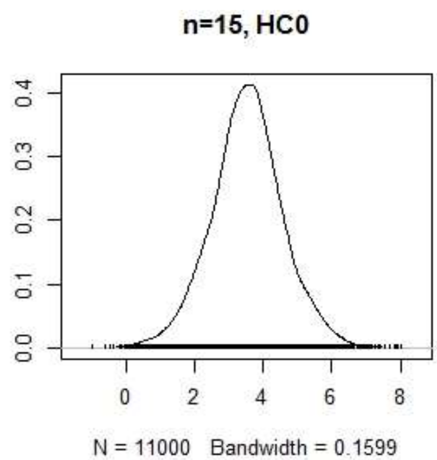
$n = 15$	3.4842 (1.7990)	0.0799 (0.1918)
$n = 20$	2.9877 (0.9931)	–0.0544 (0.1631)
$n = 30$	3.0509 (0.8556)	0.0641 (0.1446)
$n = 50$	2.0540 (0.7109)	0.1623 (0.1249)
$n = 100$	3.7394 (0.5111)	–0.1424 (0.0877)
$n = 200$	5.4929 (0.3531)	–0.5664 (0.0616)

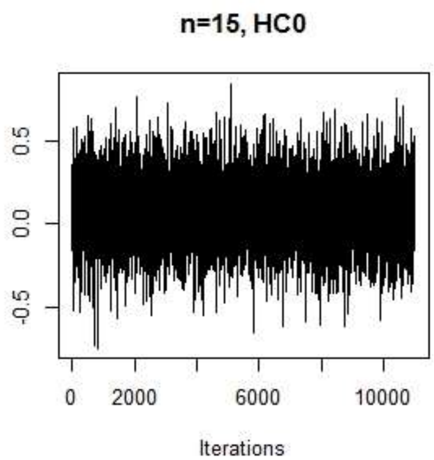
The values of standard deviation (S.D) are given in the parenthesis

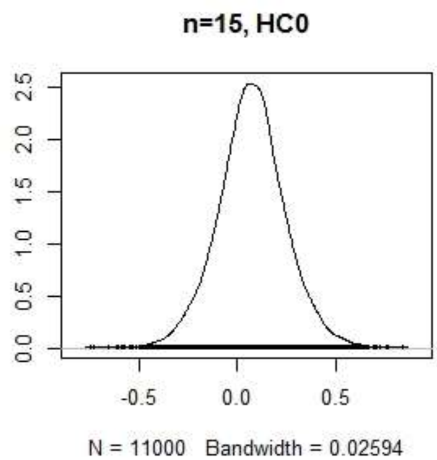
Conclusions

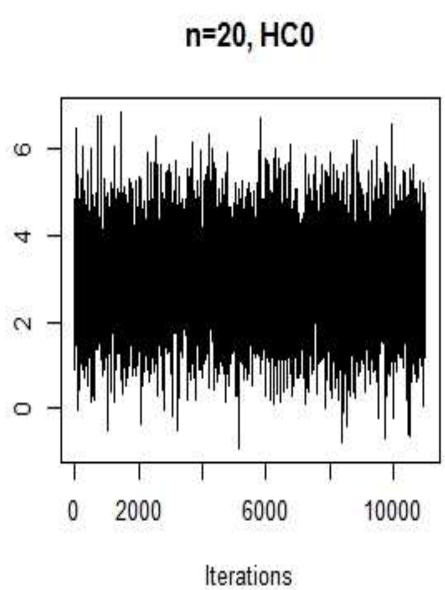
In this paper we have provided Bayesian approach as an alternative method of handling parameter estimation of linear regression model in the presence of heteroscedasticity. We incorporate heteroscedasticity consistent standard error estimators developed by White, H. (1980), Mackinnon and White (1985) and Cribari-Neto (2004) into the likelihood using noninformative prior. We observed that the estimates provide better estimates in terms of standard errors, which is line with the work of Andrew and Li (2007) on using heteroscedasticity-consistent standard error estimators in OLS regression. We therefore suggest that Bayesian approach to parameter estimation of linear regression should be routinely used when the appropriate priors information about the parameters of interest are known and the assumptions of OLS are violated.

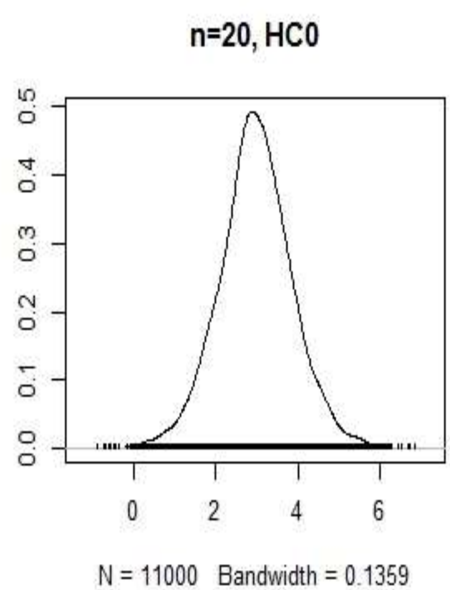

n=15, HC0
Iterations


n=15, HC0
N = 11000 Bandwidth = 0.1599


n=15, HC0
Iterations


n=15, HC0
N = 11000 Bandwidth = 0.02594


n=20, HC0
Iterations


n=20, HC0
N = 11000 Bandwidth = 0.1359

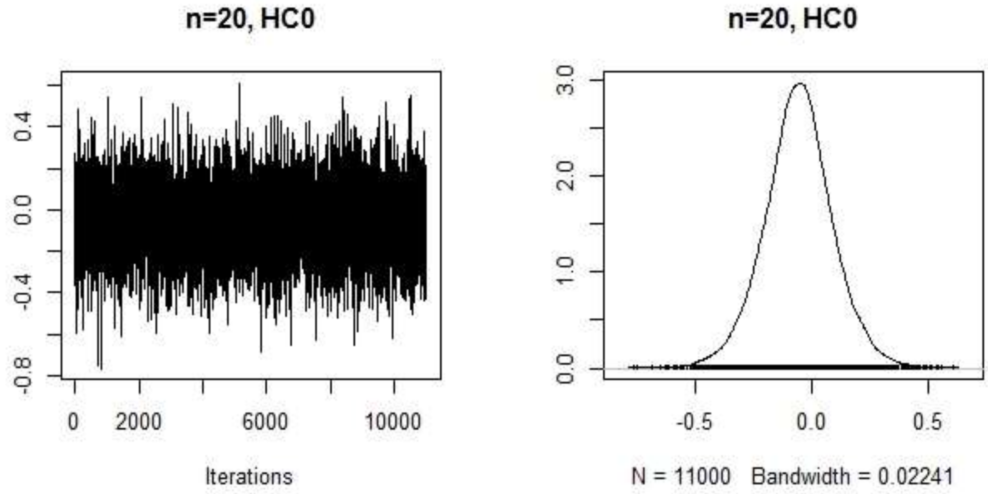
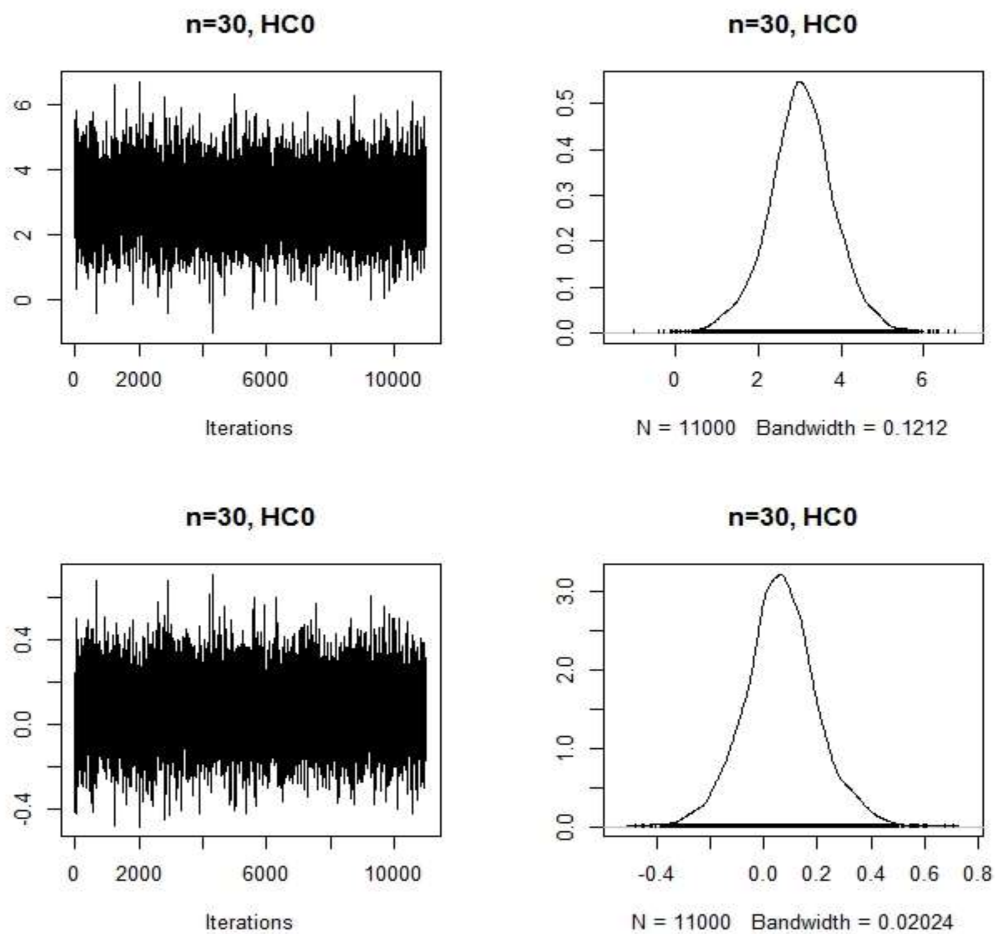


Figure 1: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC0, N=15 and N=20



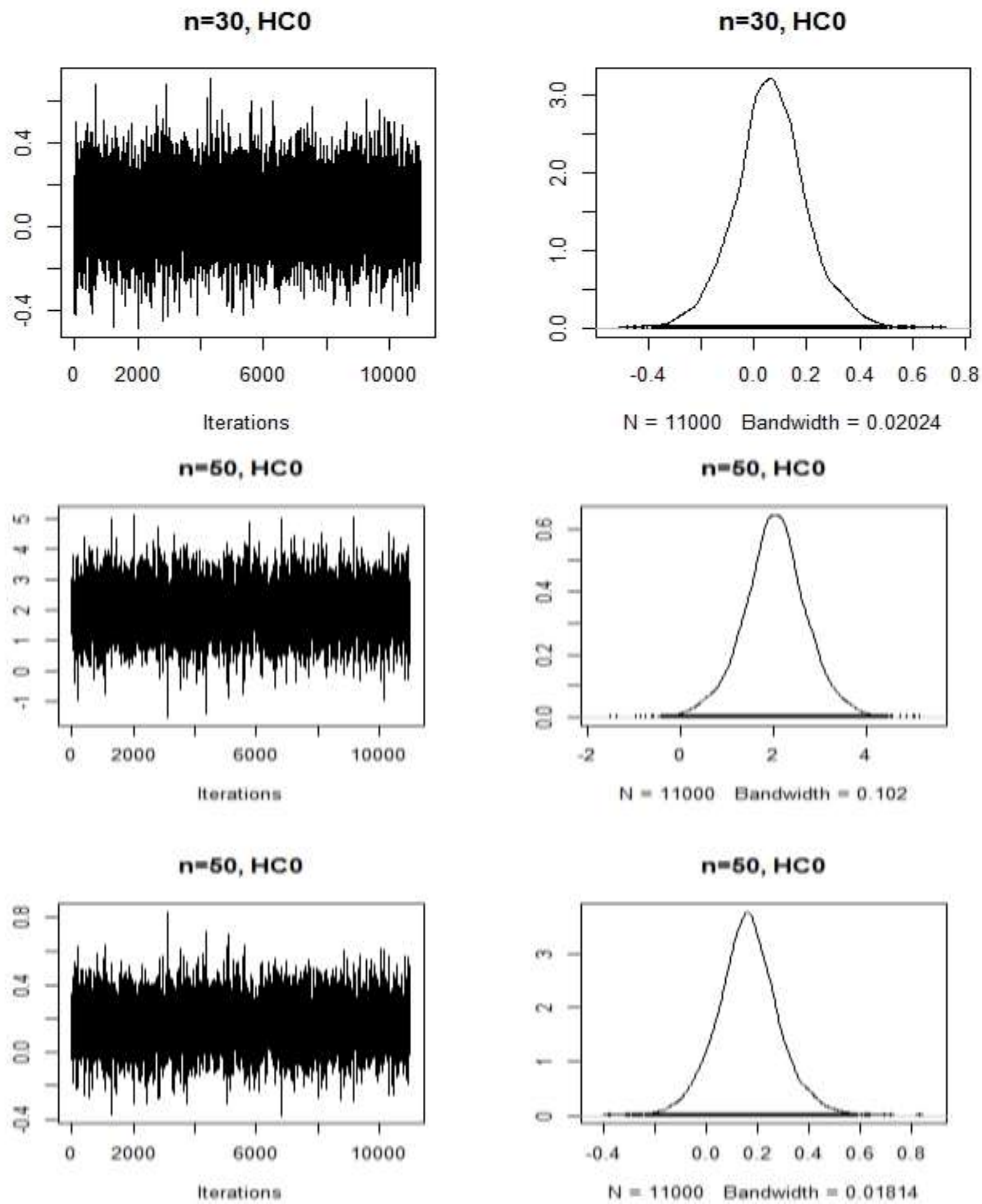
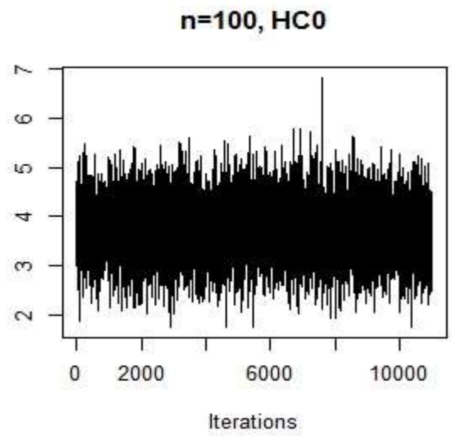
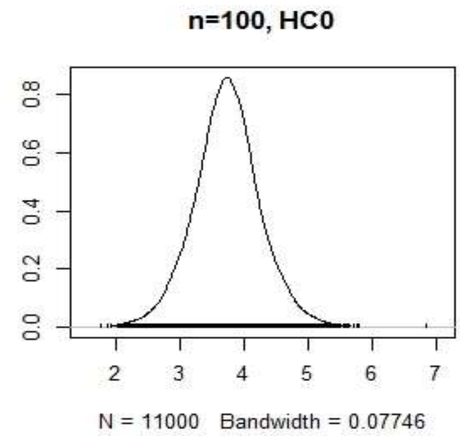
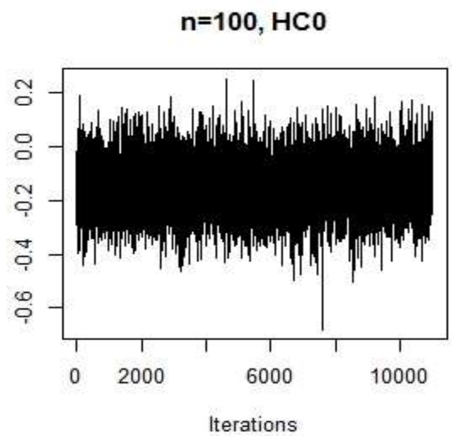
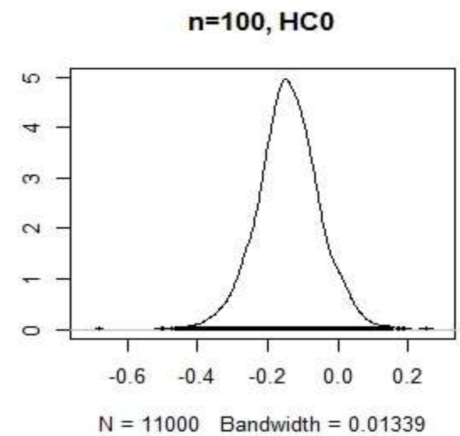


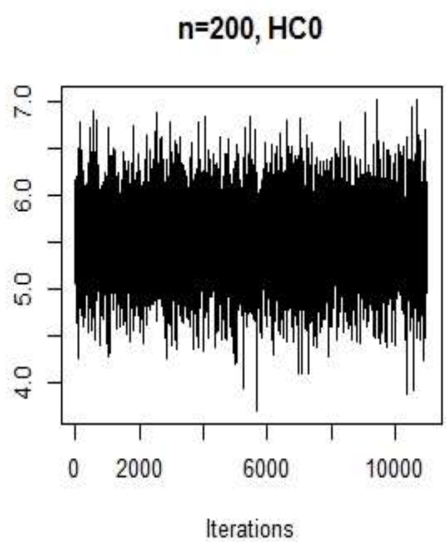
Figure 2: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC0, N=30 and N=50

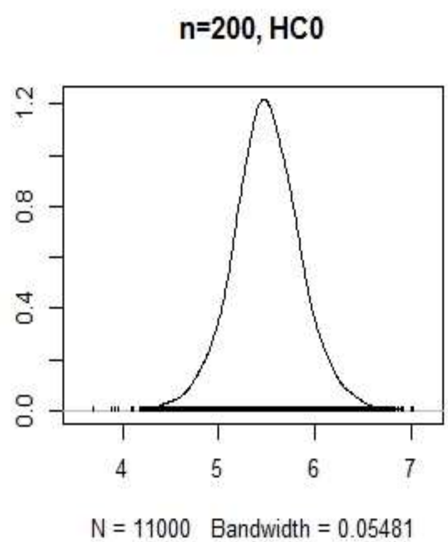

n=100, HC0
Iterations


n=100, HC0
N = 11000 Bandwidth = 0.07746


n=100, HC0
Iterations


n=100, HC0
N = 11000 Bandwidth = 0.01339


n=200, HC0
Iterations


n=200, HC0
N = 11000 Bandwidth = 0.05481

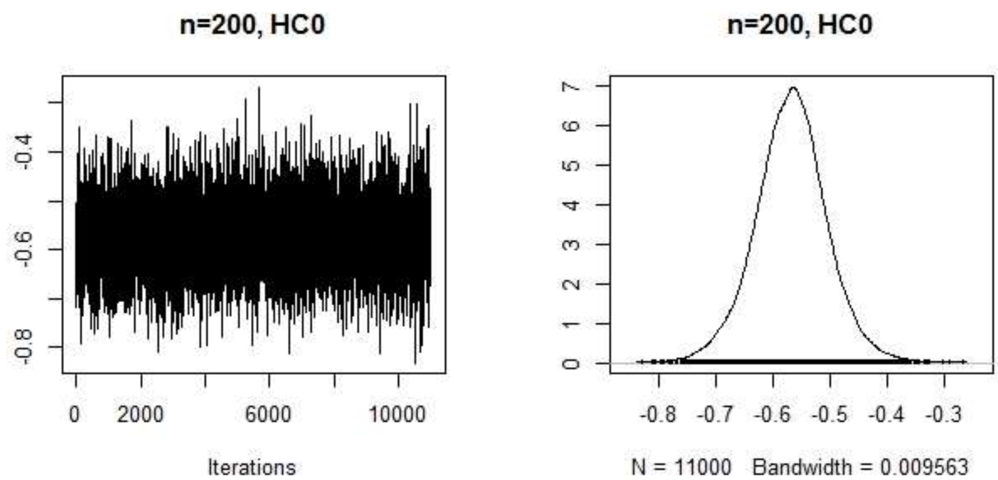


Figure 3: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC0, N=100 and N=200

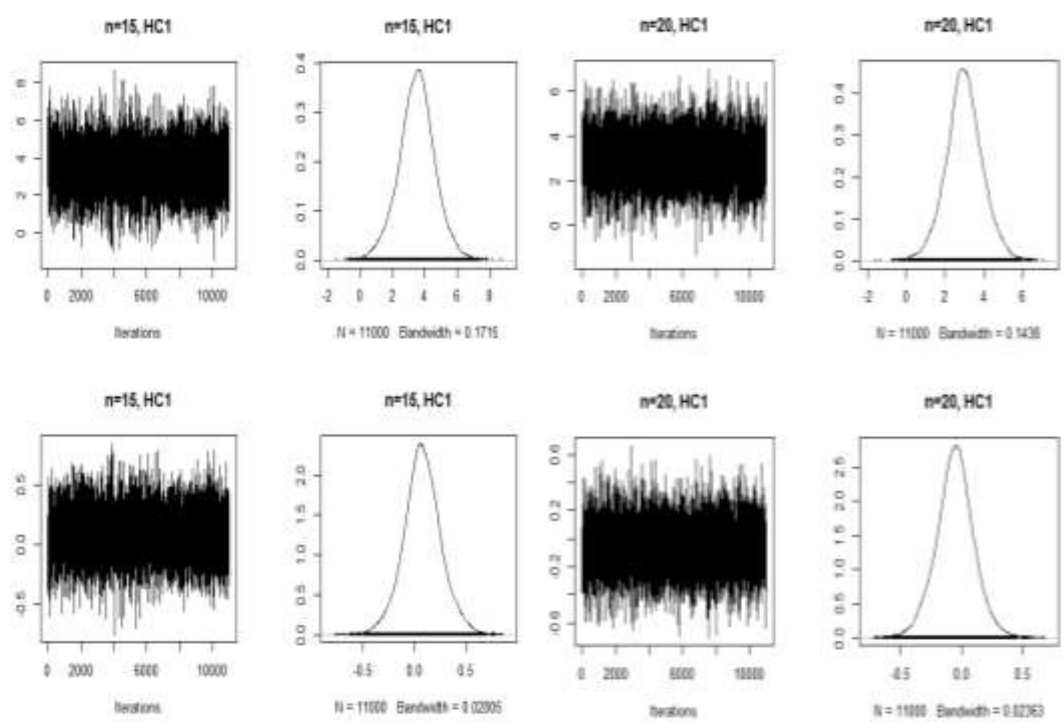


Fig 4: Posterior distributions of β for Monte-Carlo experiments- HC1, N=15 and N20

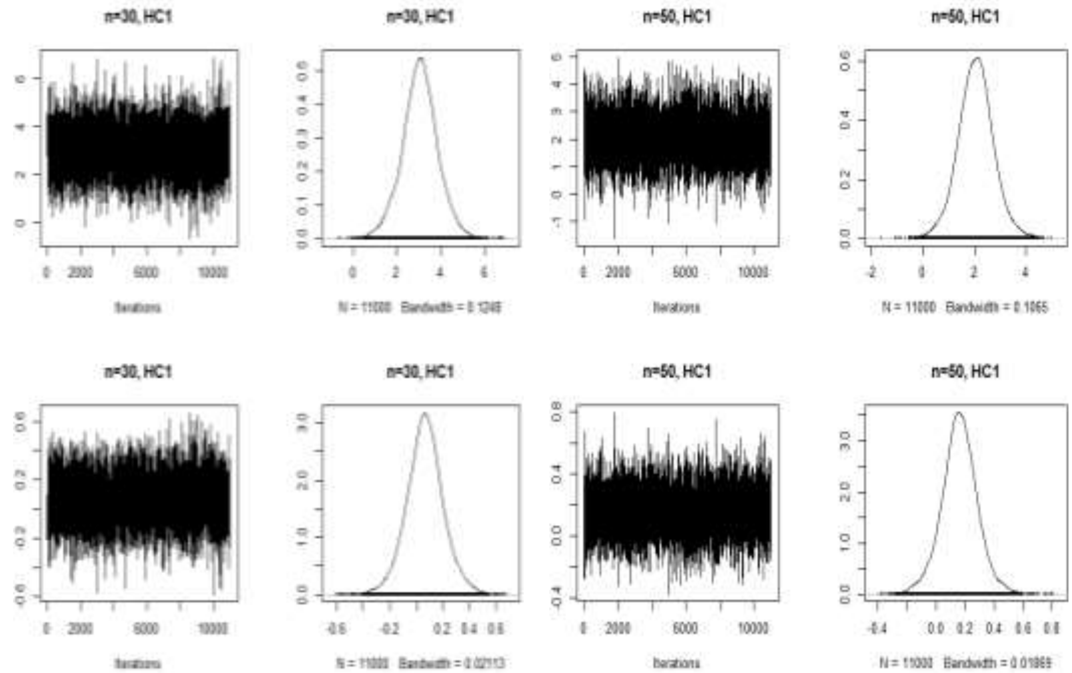


Figure 5: Posterior distribution of $\hat{\beta}$ for Monte-Carlo experiment: HC1, N=30 and N=50

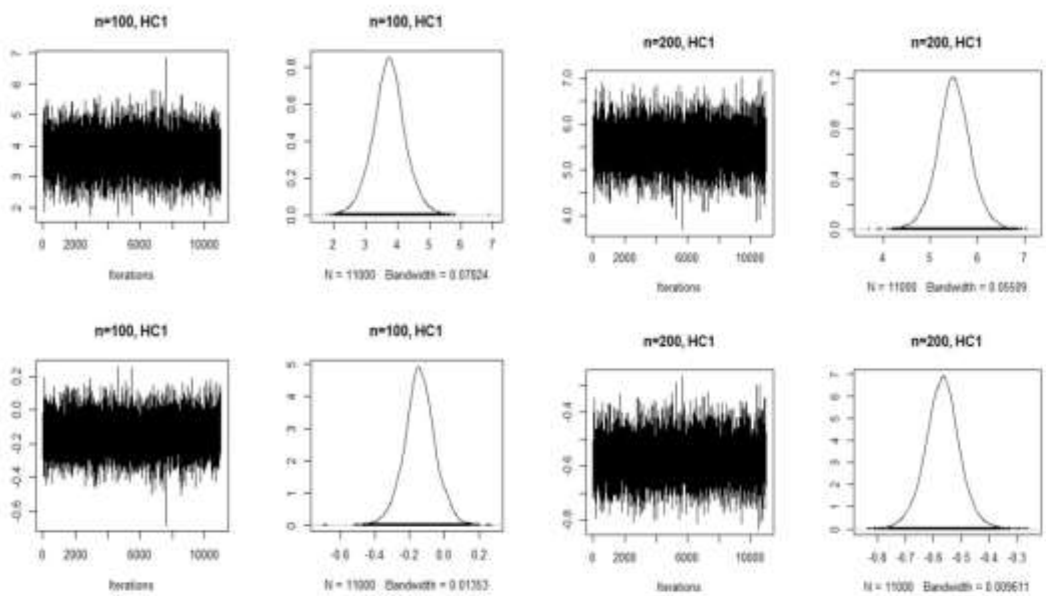


Figure 6: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC1, N=100 and N=200

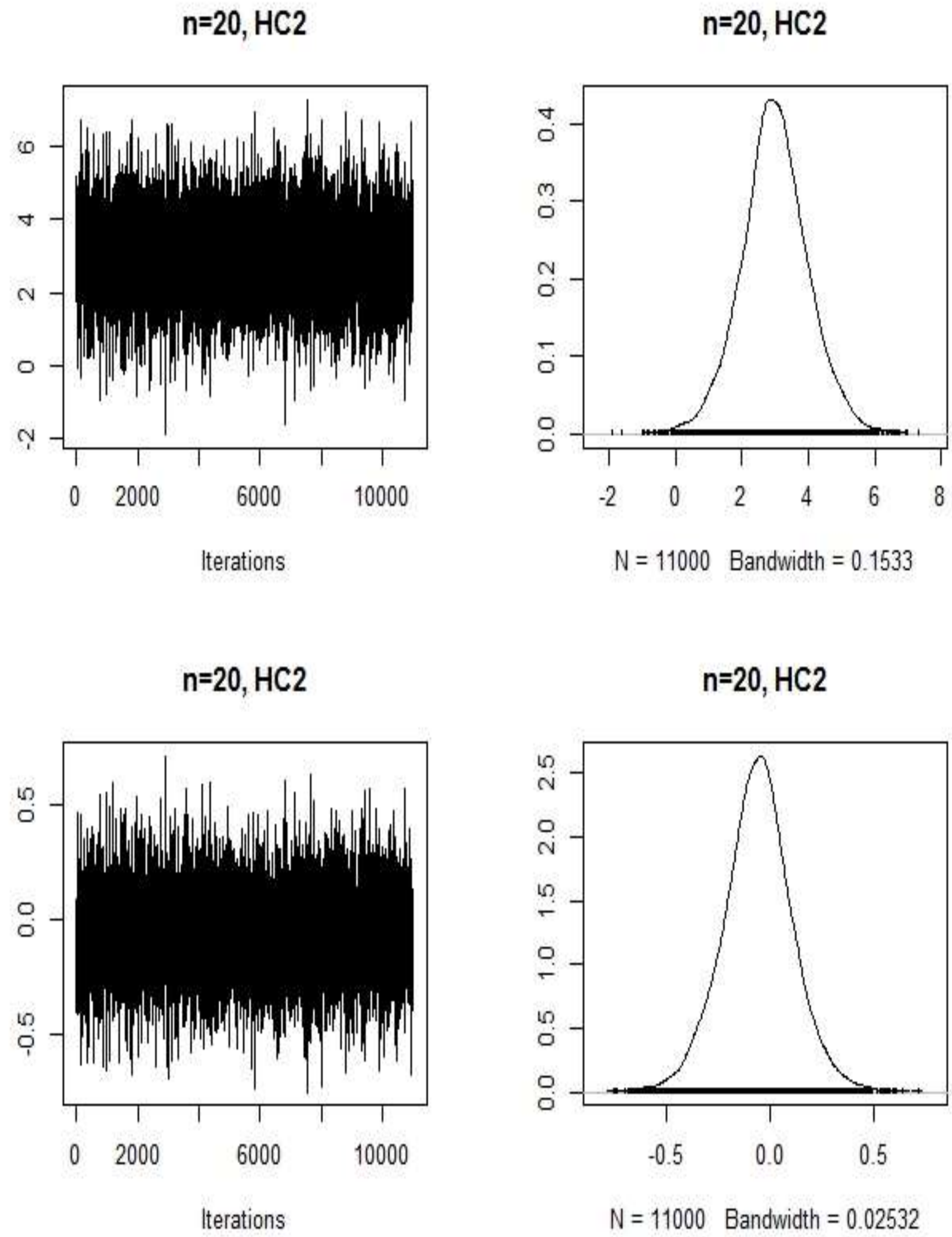
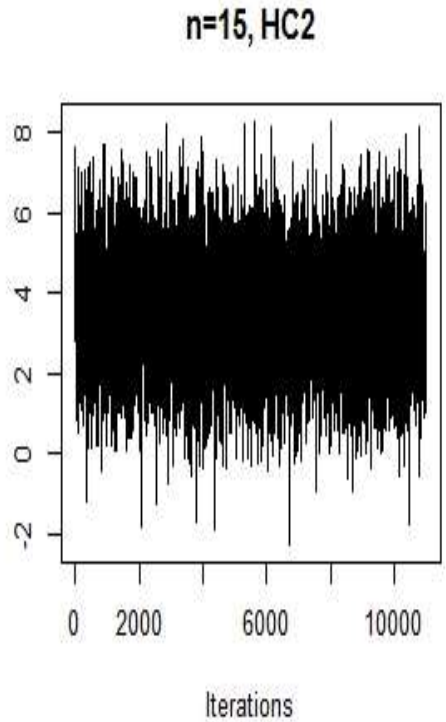
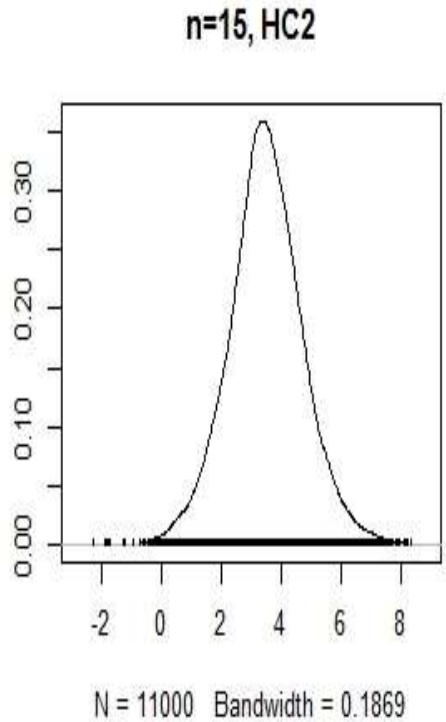
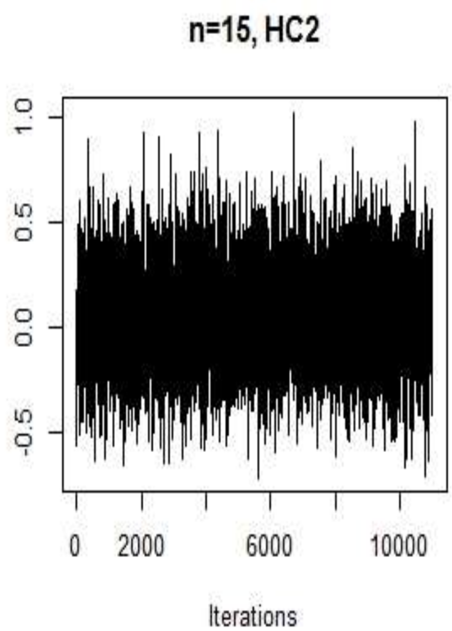
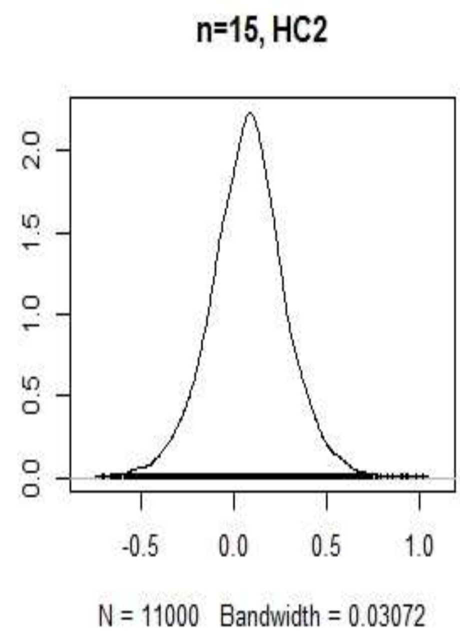


Figure7: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC2, N=15 and N=20


n=15, HC2
-2
0
2
4
6
8
0
2000
6000
10000
Iterations


n=15, HC2
0.00
0.10
0.20
0.30
-2
0
2
4
6
8
N = 11000 Bandwidth = 0.1869


n=15, HC2
-0.5
0.0
0.5
1.0
0
2000
6000
10000
Iterations


n=15, HC2
0.0
0.5
1.0
1.5
2.0
-0.5
0.0
0.5
1.0
N = 11000 Bandwidth = 0.03072

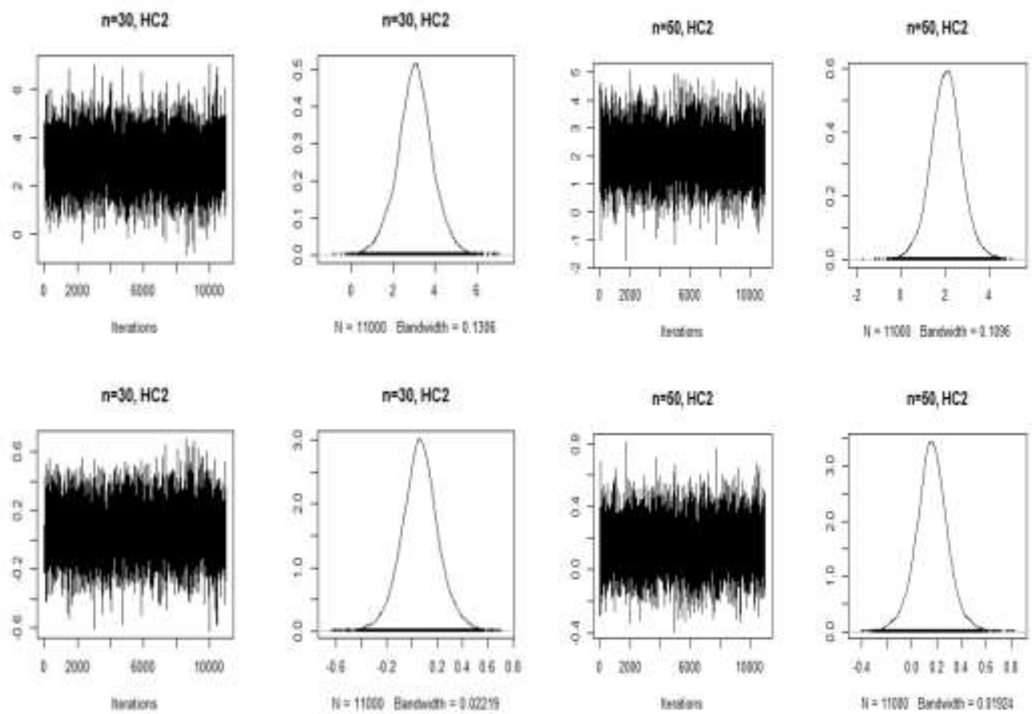


Figure 8: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC2, N=30 and N=50

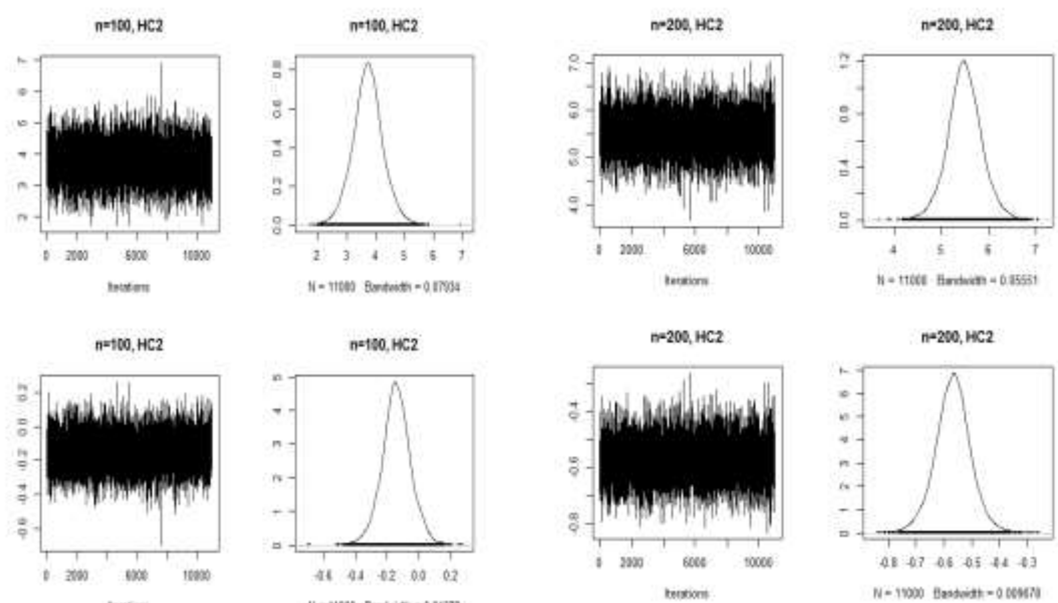


Figure 9: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC2, N=100 and N=200

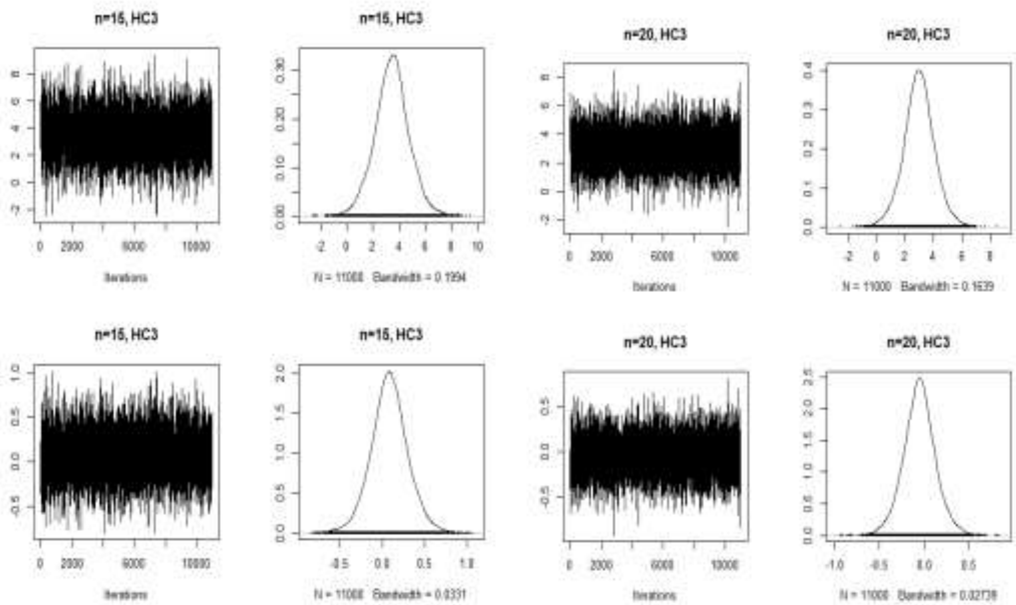


Figure 10: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC3, N=15 and N=20

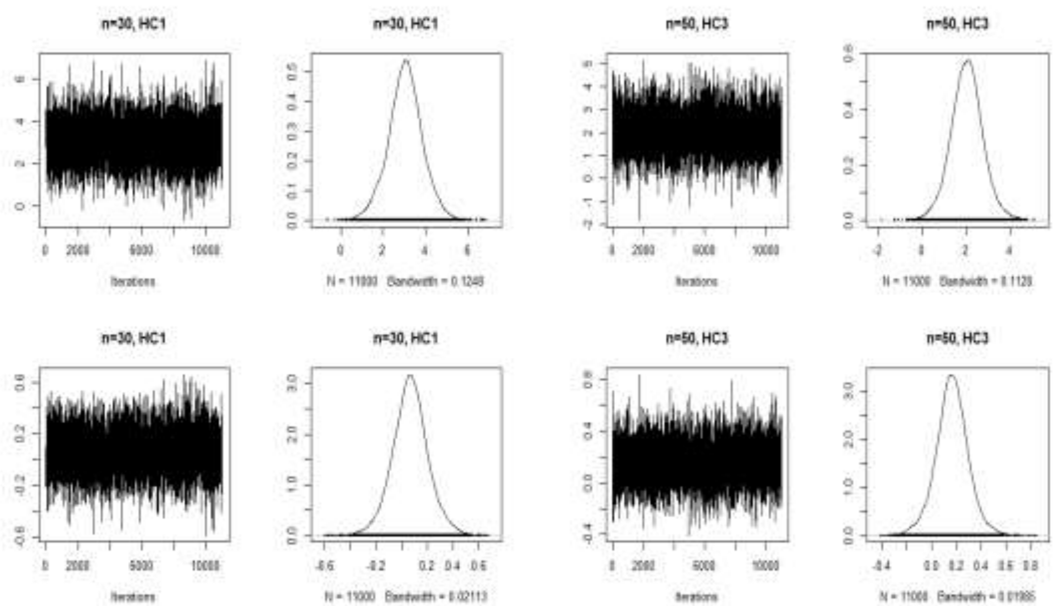


Figure 11: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC3, N=30 and N=50

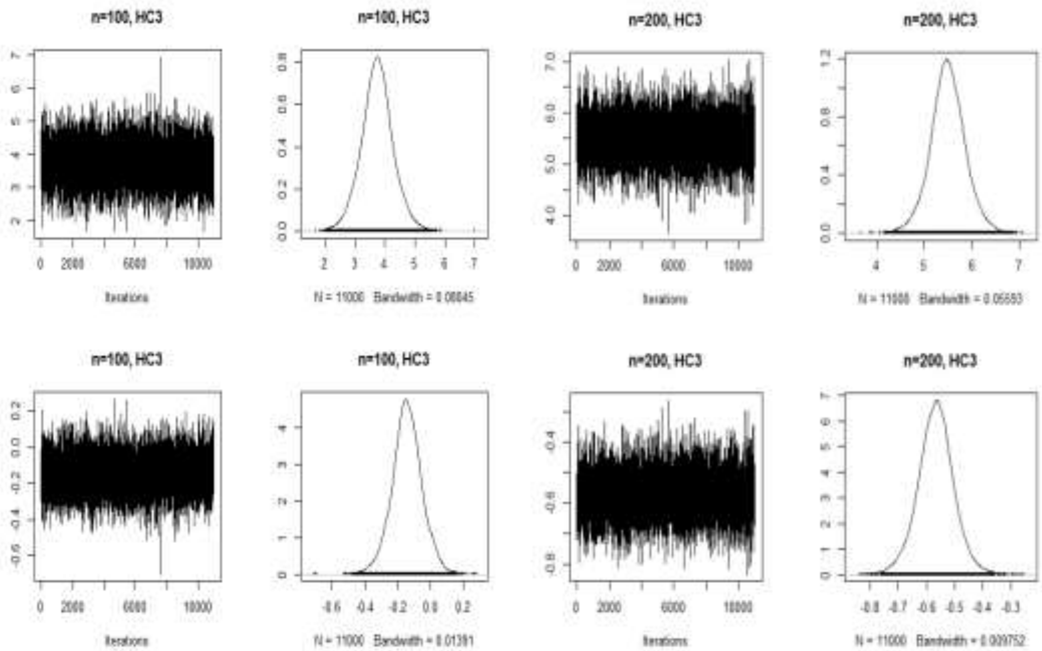


Figure 12: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC3, N=100 and N=200

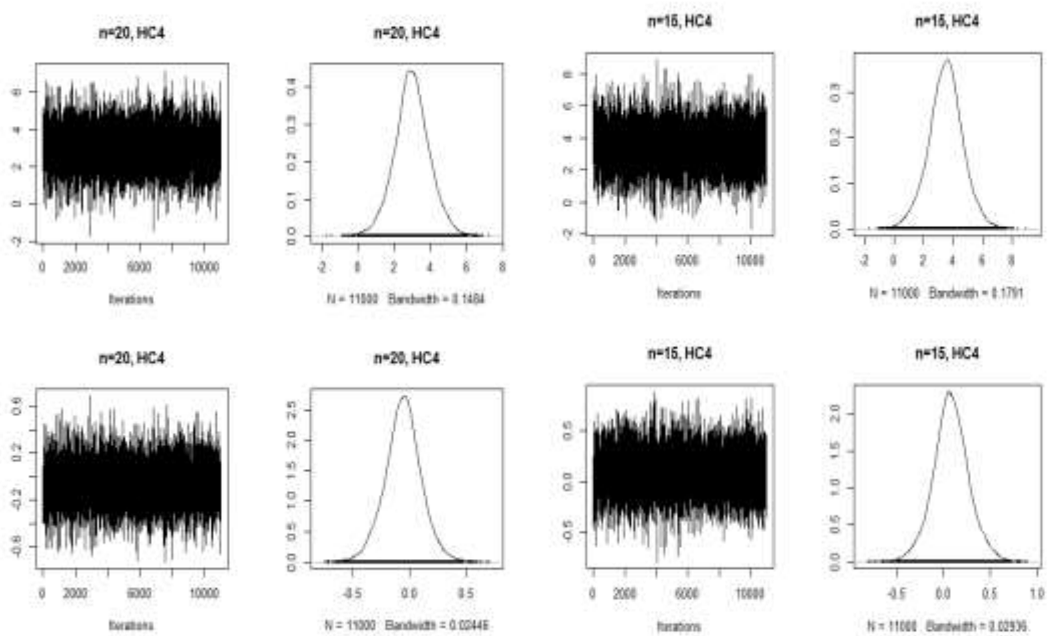


Figure 13: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC4, N=15 and N=20

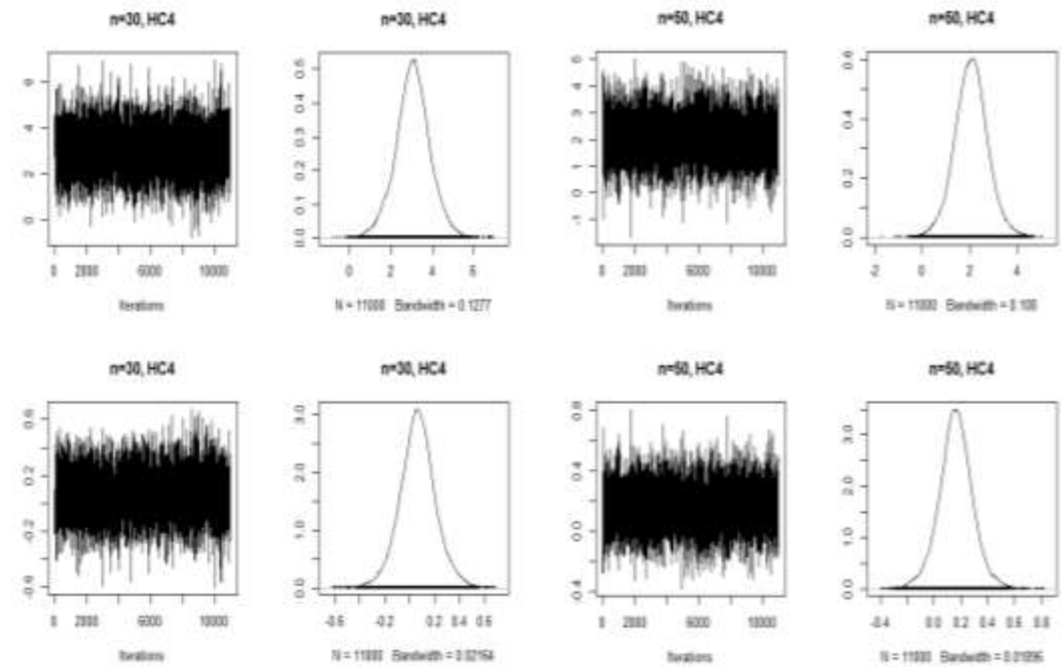


Figure 14: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC4, N=30 and N=50

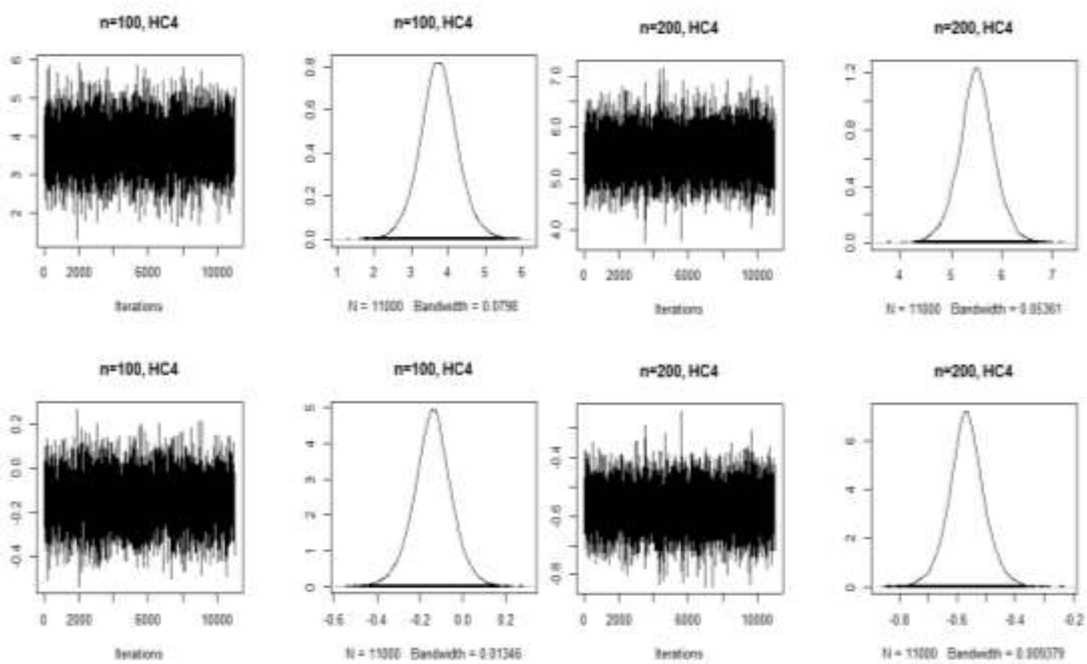


Figure 15: Posterior distributions of $\hat{\beta}$ for Monte-Carlo experiment: HC4, N=100 and N=200

References

Box, G.E.P.,and Cox, D.R. (1964). An analysis of transformations. Journal of the Royal Statistical Society B, 26,211-243.

Cribari-Neto, F. (2004). Asymptotic inference under heteroscedasticity of unknown form. Computational Statistics & Data Analysis, 45, 215-233.

Draper,N.R. and Smith, H. (1981). Applied regression analysis (2nd ed.). New York. Wiley.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities, Journal of the American Statistical Association 85, 398-409.

Geman, S. and Geman, D.G. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721-741.

Gujarati, D.N. (2004). Basic Econometrics, 4th Ed., McGraw-Hill, New York.

Hinkley, D.V. (1977). Jackknifing in unbalanced situations. Technometrics, 19,285-292.

Long, J.S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. American Statistician, 54 217-224.

Mackinnon J. G. and White H. (1985). Some Heteroscedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties, Journal of Econometrics.

Tierney, L. (1991). Markov chains for exploring posterior distributions, Manuscript.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48, 817-838.