



# Posterior analysis of the multiplicative heteroscedasticity model

Tanizaki, Hisashi

Zhang, Xingyuan

---

(Citation)

Communications in Statistics: Theory and Methods, 30(5):855-874

(Issue Date)

2001

(Resource Type)

journal article

(Version)

Accepted Manuscript

(URL)

<https://hdl.handle.net/20.500.14094/90000131>



# POSTERIOR ANALYSIS OF THE MULTIPLICATIVE HETEROSCEDASTICITY MODEL

HISASHI TANIZAKI      and      XINGYUAN ZHANG\*  
(tanizaki@kobe-u.ac.jp)      (zhxy@econ.kobe-u.ac.jp)

Faculty of Economics, Kobe University  
Kobe 657-8501, Japan

(\* Corresponding author.)

Key Words: Multiplicative Heteroscedasticity, Bayesian Approach, the Metropolis-Hastings Algorithm, Sampling Density.

## ABSTRACT

In this paper, we show how to use Bayesian approach in the multiplicative heteroscedasticity model discussed by Harvey (1976). The Gibbs sampler and the Metropolis-Hastings (MH) algorithm are applied to the multiplicative heteroscedasticity model, where some candidate-generating densities are considered in the MH algorithm. We carry out Monte Carlo study to examine the properties of the estimates via Bayesian approach and the traditional counterparts such as the modified two-step estimator (M2SE) and the maximum likelihood estimator (MLE). Our results of Monte Carlo study show that the candidate-generating density chosen in our paper is suitable, and Bayesian approach shows better performance than the traditional counterparts in the criterion of the root mean square error (RMSE) and the interquartile range (IR).

# 1 INTRODUCTION

For the heteroscedasticity model, we have to estimate both the regression coefficients and the heteroscedasticity parameters. In the literature of heteroscedasticity, traditional estimation techniques include the two-step estimator (2SE) and the maximum likelihood estimator (MLE). Harvey (1976) showed that the 2SE has an inconsistent element in the heteroscedasticity parameters and furthermore derived the consistent estimator based on the 2SE, which is called the modified two-step estimator (M2SE) in this paper. These traditional estimators are also examined in Amemiya (1985), Judge et al. (1980) and Greene (1997).

Ohtani (1982) derived the Bayesian estimator for a heteroscedasticity linear model. Using a Monte Carlo experiment, Ohtani (1982) found that among the Bayesian estimator and some traditional estimators, the Bayesian estimator shows the best properties in the mean square error (MSE) criterion. Since Ohtani (1982) obtained the Bayesian estimator by numerical integration, it is not easy to extend to the multi-dimensional cases of both the regression coefficient and the heteroscedasticity parameter.

Recently, Boscardin and Gelman (1996) developed a Bayesian approach in which a Gibbs sampler and the Metropolis-Hastings (MH) algorithm are used to estimate the parameters of heteroscedasticity in the linear model. They argued that through this kind of Bayesian approach, we can average over our uncertainty in the model parameters instead of using a point estimate via the traditional estimation techniques. Their modeling for the heteroscedasticity, however, is very simple and limited. Their choice of the heteroscedasticity is  $\text{Var}(y_i) = \sigma^2 w_i^{-\theta}$ , where  $w_i$  are known “weights” for the problem and  $\theta$  is an unknown parameter. In addition, they took only one candidate for the sampling density used in the MH algorithm and compared it with 2SE.

In this paper, we also consider Harvey’s (1976) model of multiplicative heteroscedasticity. This modeling is very flexible, general, and includes most of the useful formulations for heteroscedasticity as special cases. The Bayesian approach discussed by Ohtani (1982) and Boscardin and Gelman (1996) are extended to the multi-dimensional and more complicated cases in this paper. Our Bayesian approach includes the MH within Gibbs algorithm, where through Monte Carlo studies we examine two kinds of candidates for the sampling density in the MH algorithm and compare the Bayesian approach with the two traditional estimators, i.e., M2SE and MLE, in the criterion of the root mean square error (RMSE) and the interquartile range (IR). We obtain

the result that the Bayesian estimator significantly has smaller MSE and IR than M2SE and MLE at least for the heteroscedasticity parameters. Thus, our results of the Monte Carlo study show that the Bayesian approach performs better than the traditional estimators.

The paper is organized as follows. In Section 2, we briefly review the traditional multiplicative heteroscedasticity model. In Section 3, we discuss the Bayesian estimator and two kinds of candidates for the sampling density in the MH algorithm. Section 4 illustrates the data for Monte Carlo study and discusses its results.

## 2 MULTIPLICATIVE HETEROSCEDASTICITY MODEL

The multiplicative heteroscedasticity model discussed by Harvey (1976) can be shown as follows:

$$\mathbf{y}_i | \beta, \sigma_i^2 \sim N(\mathbf{X}_i' \beta, \sigma_i^2), \quad (1)$$

$$\sigma_i^2 = \sigma^2 \exp(\mathbf{q}_i' \alpha), \quad (2)$$

for  $i = 1, 2, \dots, n$ , where  $\mathbf{y}_i$  is the  $i$ -th observation,  $\mathbf{X}_i$  and  $\mathbf{q}_i$  are the  $i$ -th  $K \times 1$  and  $(J - 1) \times 1$  vectors of explanatory variables respectively.  $\beta$  and  $\gamma$  are vectors of unknown parameters.

The model given by (1) and (2) includes several special cases such as the model in Boscardin and Gelman (1996), in which  $\mathbf{q}_i = \ln \mathbf{w}_i$  and  $\theta = -\alpha$ . As shown in Greene (1997), there is a useful simplification of the formulation. Let  $\mathbf{z}_i = (1, \mathbf{q}_i)$  and  $\gamma = (\ln \sigma^2, \alpha)$ . Then we can simply rewrite the model as:

$$\sigma_i^2 = \exp(\mathbf{z}_i' \gamma). \quad (3)$$

Note that  $\exp(\gamma_1)$  provides  $\sigma^2$ , where  $\gamma_1$  denotes the first element of  $\gamma$ . As for the variance of  $\mathbf{u}_i$ , hereafter we use (3), rather than (2).

The generalized least squares (GLS) estimator of  $\beta$ , denoted by  $\hat{\beta}_{\text{GLS}}$ , is given by:

$$\hat{\beta}_{\text{GLS}} = \left( \sum_{i=1}^n \exp(-\mathbf{z}_i' \gamma) \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \sum_{i=1}^n \exp(-\mathbf{z}_i' \gamma) \mathbf{X}_i \mathbf{y}_i, \quad (4)$$

where  $\hat{\beta}_{\text{GLS}}$  depends on  $\gamma$ , which is the unknown parameter vector. To obtain the feasible GLS estimator, we need to replace  $\gamma$  by its consistent estimate. We have two traditional consistent estimators of  $\gamma$ , i.e., M2SE and MLE, which are briefly described as follows.

Modified Two-Step Estimator (M2SE): Let us define the ordinary least squares (OLS) residual by  $e_i = y_i - X_i' \hat{\beta}_{OLS}$ , where  $\hat{\beta}_{OLS}$  represents the OLS estimator, i.e.,  $\hat{\beta}_{OLS} = \left( \sum_{i=1}^n X_i X_i' \right)^{-1} \sum_{i=1}^n X_i y_i$ . For 2SE of  $\gamma$ , we may form the following regression:

$$\ln e_i^2 = z_i' \gamma + u_i.$$

The OLS estimator of  $\gamma$  applied to the above equation leads to the 2SE of  $\gamma$ , because  $e_i$  is obtained by OLS in the first step. Thus, the OLS estimator of  $\gamma$  gives us 2SE, denoted by  $\hat{\gamma}_{2SE}$ , which is given by:

$$\hat{\gamma}_{2SE} = \left( \sum_{i=1}^n z_i z_i' \right)^{-1} \sum_{i=1}^n z_i \ln e_i^2,$$

A problem with this estimator is that  $u_i$ ,  $i = 1, 2, \dots, n$ , have non-zero means and are heteroscedastic. If  $e_i$  converges in distribution to  $\epsilon_i$ , the  $u_i$  will be asymptotically independent with mean  $E(u_i) = -1.2704$  and variance  $\text{Var}(u_i) = 4.9348$ , which are shown in Harvey (1976). Then, we have the following mean and variance of  $\hat{\gamma}_{2SE}$ :

$$\begin{aligned} E(\hat{\gamma}_{2SE}) &= \gamma - 1.2704 \left( \sum_{i=1}^n z_i z_i' \right)^{-1} \sum_{i=1}^n z_i, \\ \text{Var}(\hat{\gamma}_{2SE}) &= 4.9348 \left( \sum_{i=1}^n z_i z_i' \right)^{-1}. \end{aligned} \tag{5}$$

For the second term in equation (5), the first element is equal to  $-1.2704$  and the remaining elements are zero, which can be obtained by simple calculation. Therefore, the first element of  $\hat{\gamma}_{2SE}$  is biased but the remaining elements are still unbiased. To obtain a consistent estimator of  $\gamma_1$ , we consider M2SE of  $\gamma$ , denoted by  $\hat{\gamma}_{M2SE}$ , which is given by:

$$\hat{\gamma}_{M2SE} = \hat{\gamma}_{2SE} + 1.2704 \left( \sum_{i=1}^n z_i z_i' \right)^{-1} \sum_{i=1}^n z_i.$$

Let  $\Sigma_{M2SE}$  be the variance of  $\hat{\gamma}_{M2SE}$ . Then,  $\Sigma_{M2SE}$  is represented by:

$$\Sigma_{M2SE} \equiv \text{Var}(\hat{\gamma}_{M2SE}) = \text{Var}(\hat{\gamma}_{2SE}) = 4.9348 \left( \sum_{i=1}^n z_i z_i' \right)^{-1}.$$

The first element of  $\hat{\gamma}_{2SE}$  and  $\hat{\gamma}_{M2SE}$  corresponds to the estimate of  $\sigma^2$ , which value does not influence  $\hat{\beta}_{GLS}$ . Since the remaining elements of  $\hat{\gamma}_{2SE}$  are equal to those of  $\hat{\gamma}_{M2SE}$ ,  $\hat{\beta}_{2SE}$  is equivalent to  $\hat{\beta}_{M2SE}$ , where  $\hat{\beta}_{2SE}$  and  $\hat{\beta}_{M2SE}$  denote 2SE and M2SE of  $\beta$ , respectively. Note that  $\hat{\beta}_{2SE}$  and  $\hat{\beta}_{M2SE}$  can be obtained by substituting  $\hat{\gamma}_{2SE}$  and  $\hat{\gamma}_{M2SE}$  into  $\gamma$  in (4).

Maximum Likelihood Estimator (MLE): The density of  $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$  based on (1) and (3) is:

$$f(\mathbf{y}|\boldsymbol{\beta}, \gamma, \mathbf{X}) \propto \exp \left[ -\frac{1}{2} \sum_{i=1}^n \left( \exp(-\mathbf{z}_i' \gamma) (\mathbf{y}_i - \mathbf{X}_i' \boldsymbol{\beta})^2 + \mathbf{z}_i' \gamma \right) \right], \quad (6)$$

which is maximized with respect to  $\boldsymbol{\beta}$  and  $\gamma$ , using the method of scoring. That is, given values for  $\boldsymbol{\beta}^{(j)}$  and  $\gamma^{(j)}$ , the method of scoring is implemented by the following iterative procedure:

$$\begin{aligned} \boldsymbol{\beta}^{(j+1)} &= \left( \sum_{i=1}^n \exp(-\mathbf{z}_i' \gamma^{(j)}) \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \sum_{i=1}^n \exp(-\mathbf{z}_i' \gamma^{(j)}) \mathbf{X}_i \mathbf{y}_i, \\ \gamma^{(j+1)} &= \gamma^{(j)} + 2 \left( \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i' \right)^{-1} \frac{1}{2} \sum_{i=1}^n \mathbf{z}_i \left( \exp(-\mathbf{z}_i' \gamma^{(j)}) \mathbf{e}_i^2 - 1 \right), \end{aligned}$$

for  $j = 0, 1, 2, \dots$ , where  $\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i' \boldsymbol{\beta}^{(j)}$ . The starting value for the above iteration may be taken as  $(\boldsymbol{\beta}^{(0)}, \gamma^{(0)}) = (\hat{\boldsymbol{\beta}}_{\text{OLS}}, \hat{\gamma}_{2\text{SE}})$ ,  $(\hat{\boldsymbol{\beta}}_{2\text{SE}}, \hat{\gamma}_{2\text{SE}})$  or  $(\hat{\boldsymbol{\beta}}_{\text{M2SE}}, \hat{\gamma}_{\text{M2SE}})$ .

Let  $\boldsymbol{\delta} = (\boldsymbol{\beta}, \gamma)$ . The limit of  $\boldsymbol{\delta}^{(j)} = (\boldsymbol{\beta}^{(j)}, \gamma^{(j)})$  gives us the MLE of  $\boldsymbol{\delta}$ , which is denoted by  $\hat{\boldsymbol{\delta}}_{\text{MLE}} = (\hat{\boldsymbol{\beta}}_{\text{MLE}}, \hat{\gamma}_{\text{MLE}})$  in this paper. Based on the information matrix, the asymptotic covariance matrix of  $\hat{\boldsymbol{\delta}}_{\text{MLE}}$  is represented by:

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\delta}}_{\text{MLE}}) &= \left[ -\text{E} \left( \frac{\partial^2 \ln f(\mathbf{y}|\boldsymbol{\delta}, \mathbf{X})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'} \right) \right]^{-1} \\ &= \begin{pmatrix} \left( \sum_{i=1}^n \exp(-\mathbf{z}_i' \gamma) \mathbf{X}_i \mathbf{X}_i' \right)^{-1} & 0 \\ 0 & 2 \left( \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i' \right)^{-1} \end{pmatrix}. \end{aligned} \quad (7)$$

Thus, from (7), asymptotically there is no correlation between  $\hat{\boldsymbol{\beta}}_{\text{MLE}}$  and  $\hat{\gamma}_{\text{MLE}}$ , and furthermore the asymptotic variance of  $\hat{\gamma}_{\text{MLE}}$  is represented by:  $\boldsymbol{\Sigma}_{\text{MLE}} \equiv \text{Var}(\hat{\gamma}_{\text{MLE}}) = 2 \left( \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i' \right)^{-1}$ , which implies that  $\hat{\gamma}_{\text{M2SE}}$  is asymptotically inefficient because  $\boldsymbol{\Sigma}_{\text{M2SE}} - \boldsymbol{\Sigma}_{\text{MLE}}$  is positive definite.

### 3 BAYESIAN ESTIMATION

We assume that the prior distributions of the parameters  $\boldsymbol{\beta}$  and  $\gamma$  are diffuse, which are represented by:

$$P_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \text{const.}, \quad P_{\gamma}(\gamma) = \text{const.} \quad (8)$$

Combining the prior distributions (8) and the likelihood function (6), the posterior distribution  $P_{\boldsymbol{\beta}\gamma}(\boldsymbol{\beta}, \gamma|\mathbf{y}, \mathbf{X})$  is obtained. We utilize the Gibbs sampler to sample from the posterior distribution. Then, from the joint density

$P_{\beta|\gamma}(\beta, \gamma | \mathbf{y}, \mathbf{X})$ , we can derive the following two conditional densities:

$$P_{\gamma|\beta}(\gamma | \beta, \mathbf{y}, \mathbf{X}) \propto \exp \left[ -\frac{1}{2} \sum_{i=1}^n \left( \exp(-z_i' \gamma) (\mathbf{y}_i - \mathbf{X}_i' \beta)^2 + z_i' \gamma \right) \right], \quad (9)$$

$$P_{\beta|\gamma}(\beta | \gamma, \mathbf{y}, \mathbf{X}) = \mathbf{N}(\mathbf{B}_1, \mathbf{H}_1), \quad (10)$$

where

$$\mathbf{H}_1^{-1} = \sum_{i=1}^n \exp(-z_i' \gamma) \mathbf{X}_i \mathbf{X}_i', \quad \mathbf{B}_1 = \mathbf{H}_1 \sum_{i=1}^n \exp(-z_i' \gamma) \mathbf{X}_i \mathbf{y}_i.$$

Sampling from (10) is simple since it is a  $K$ -variate normal distribution with mean  $\mathbf{B}_1$  and variance  $\mathbf{H}_1$ . However, since the  $J$ -variate distribution (9) does not take the form of any standard density, it is not easy to sample from (9). In this case, the MH algorithm can be used within the Gibbs sampler. See Tierney (1994) and Chib and Greeberg (1995) for a general discussion.

Let  $\gamma^{(t)}$  be the  $t$ -th random draw of  $\gamma$  and  $\mathbf{z}$  be a candidate of the  $(t+1)$ -th random draw of  $\gamma$ . The MH algorithm utilizes another appropriate distribution function  $P_*(\mathbf{z} | \gamma^{(t)})$ , which is called the sampling density or the proposal density. Let us define the acceptance rate  $\omega(\gamma^{(t)}, \mathbf{z})$  as:

$$\omega(\gamma^{(t)}, \mathbf{z}) = \begin{cases} \min \left( \frac{P_{\gamma|\beta}(\mathbf{z} | \beta^{(t)}, \mathbf{y}, \mathbf{X}) P_*(\gamma^{(t)} | \mathbf{z})}{P_{\gamma|\beta}(\gamma^{(t)} | \beta^{(t)}, \mathbf{y}, \mathbf{X}) P_*(\mathbf{z} | \gamma^{(t)})}, 1 \right), & \text{if } P_{\gamma|\beta}(\gamma^{(t)} | \beta^{(t)}, \mathbf{y}, \mathbf{X}) P_*(\mathbf{z} | \gamma^{(t)}) > 0, \\ 1, & \text{otherwise.} \end{cases}$$

The sampling procedure based on the MH algorithm within Gibbs sampling is as follows:

- (i) Set the initial value  $\beta^{(0)}$ , which may be taken as  $\hat{\beta}_{\text{M2SE}}$  or  $\hat{\beta}_{\text{MLE}}$ .
- (ii) Given  $\beta^{(t)}$ , generate a random draw of  $\gamma$ , denoted by  $\gamma^{(t+1)}$ , from the conditional density  $P_{\gamma|\beta}(\gamma | \beta^{(t)}, \mathbf{y}, \mathbf{X})$ , where the MH algorithm is utilized for random number generation because it is not easy to generate random draws of  $\gamma$  from (9). The MH algorithm is implemented as follows:
  - (a) Given  $\gamma^{(t)}$ , generate a random draw  $\mathbf{z}$  from  $P_*(\cdot | \gamma^{(t)})$  and compute the acceptance rate  $\omega(\gamma^{(t)}, \mathbf{z})$ .
  - (b) Set  $\gamma^{(t+1)} = \mathbf{z}$  with probability  $\omega(\gamma^{(t)}, \mathbf{z})$  and  $\gamma^{(t+1)} = \gamma^{(t)}$  otherwise,

- (iii) Given  $\gamma^{(t+1)}$ , generate a random draw of  $\beta$ , denoted by  $\beta^{(t+1)}$ , from the conditional density  $P_{\beta|\gamma}(\beta|\gamma^{(t+1)}, \mathbf{y}, \mathbf{X})$ , which is  $\beta|\gamma^{(t+1)}, \mathbf{y}, \mathbf{X} \sim N(\mathbf{B}_1, \mathbf{H}_1)$  as shown in (10).
- (iv) Repeat (ii) and (iii) for  $t = 0, 1, 2, \dots, N$ .

Note that the iteration of Steps (ii) and (iii) corresponds to the Gibbs sampler, which iteration yields random draws of  $\beta$  and  $\gamma$  from the joint density  $P_{\beta\gamma}(\beta, \gamma|\mathbf{y}, \mathbf{X})$  when  $t$  is large enough. It is well known that convergence of the Gibbs sampler is slow when  $\beta$  is highly correlated with  $\gamma$ . That is, a large number of random draws have to be generated in this case. Therefore, depending on the underlying joint density, we have the case where the Gibbs sampler does not work at all. For example, see Chib and Greenberg (1995) for convergence of the Gibbs sampler. In the model represented by (1) and (2), however, there is asymptotically no correlation between  $\hat{\beta}_{\text{MLE}}$  and  $\hat{\gamma}_{\text{MLE}}$ , as shown in (7). It might be expected that correlation between  $\hat{\beta}_{\text{MLE}}$  and  $\hat{\gamma}_{\text{MLE}}$  is not too high even in the small sample. Therefore, it might be appropriate to consider that the Gibbs sampler works well in our model.

In Step (ii), the sampling density  $P_*(\mathbf{z}|\gamma^{(t)})$  is utilized. In this paper, we consider the multivariate normal density function for the sampling distribution, which is discussed as follows.

Choice of the Sampling Density in Step (ii): Several generic choices of the sampling density are discussed by Tierney (1994) and Chib and Greenberg (1995). In this paper, we take  $P_*(\gamma|\gamma^{(t)}) = P_*(\gamma)$  as the sampling density, which is called the independence chain because the sampling density is not a function of  $\gamma^{(t)}$ . We consider taking the multivariate normal sampling density in our independence MH algorithm, because of its simplicity. Therefore,  $P_*(\gamma)$  is taken as follows:

$$P_*(\gamma) = N(\gamma_*, \mathbf{d}\Sigma_*), \quad (11)$$

which represents the  $J$ -variate normal distribution with mean  $\gamma_*$  and variance  $\mathbf{d}\Sigma_*$ . The tuning parameter  $\mathbf{d}$  is introduced into the sampling density (11). We choose the tuning parameter  $\mathbf{d}$  which maximizes the acceptance rate in the MH algorithm. A detail discussion is given in Section 4.

Thus, the sampling density of  $\gamma$  is normally distributed with mean  $\gamma_*$  and variance  $\Sigma_*$ . As for  $(\gamma_*, \Sigma_*)$ , in the next section we choose one of  $(\hat{\gamma}_{\text{M2SE}}, \Sigma_{\text{M2SE}})$  and  $(\hat{\gamma}_{\text{MLE}}, \Sigma_{\text{MLE}})$  from the criterion of the acceptance rate. As shown in Section 2, both of the two estimators  $\hat{\gamma}_{\text{M2SE}}$  and  $\hat{\gamma}_{\text{MLE}}$  are consistent estimates

of  $\gamma$ . Therefore, it might be very plausible to consider that the sampling density is distributed around the consistent estimates.

**Bayesian Estimator:** From the convergence theory of the Gibbs sampler and the MH algorithm, as  $t$  goes to infinity we can regard  $\gamma^{(t)}$  and  $\beta^{(t)}$  as random draws from the target density  $P_{\beta\gamma}(\beta, \gamma | y, X)$ . Let  $M$  be a sufficiently large number.  $\gamma^{(M)}$  and  $\beta^{(M)}$  are taken as the random draws from the posterior density  $P_{\beta\gamma}(\beta, \gamma | y, X)$ . Therefore, the Bayesian estimators  $\hat{\gamma}_{BZZ}$  and  $\hat{\beta}_{BZZ}$  are given by:

$$\hat{\gamma}_{BZZ} = \frac{1}{N - M} \sum_{t=M+1}^N \gamma^{(t)}, \quad \hat{\beta}_{BZZ} = \frac{1}{N - M} \sum_{t=M+1}^N \beta^{(t)},$$

where we read BZZ as the Bayesian estimator which uses the multivariate normal sampling density with mean  $\hat{\gamma}_{ZZ}$  and variance  $\Sigma_{ZZ}$ . ZZ takes M2SE or MLE. In this paper we consider two kinds of candidates of the sampling density for the Bayesian estimator, which are denoted by BM2SE and BMLE. Thus, in Section 4, we compare the two Bayesian estimators (i.e, BM2SE and BMLE) with the two traditional estimators (i.e., M2SE and MLE).

## 4 MONTE CARLO STUDY

### 4.1 Setup of the Model

In our Monte Carlo study, we consider using the simulated data, in which the true data generating process (DGP) is presented in Judge et al. (1980, pp.156). The DGP is defined as:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i, \quad (12)$$

where the  $\epsilon_i$ 's are normally and independently distributed with  $E(\epsilon_i) = 0$ ,  $E(\epsilon_i^2) = \sigma_i^2$  and,

$$\sigma_i^2 = \exp(\gamma_1 + \gamma_2 x_{2i}), \quad \text{for } i = 1, 2, \dots, n. \quad (13)$$

As in Judge et al. (1980), the parameter values are set to be  $(\beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2) = (10, 1, 1, -2, 0.25)$ . From (12) and (13), Judge et al. (1980, pp.160-165) generated one hundred samples of  $y$  with  $n = 20$ . In this paper, we utilize  $x_{2i}$  and  $x_{3i}$  given in Judge et al. (1980, pp.156), which is shown in Table 1, and generate  $G$  samples of  $y$  given the  $X$ . That is, we perform  $G$  simulation runs for each estimator, where  $G = 10000$  is taken in this paper.

The simulation procedure in this section is as follows:

Table 1: The Exogenous Variables  $\mathbf{x}_{1i}$  and  $\mathbf{x}_{2i}$

i	1	2	3	4	5	6	7	8	9	10
$\mathbf{x}_{2i}$	14.53	15.30	15.92	17.41	18.37	18.83	18.84	19.71	20.01	20.26
$\mathbf{x}_{3i}$	16.74	16.81	19.50	22.12	22.34	17.47	20.24	20.37	12.71	22.98
i	11	12	13	14	15	16	17	18	19	20
$\mathbf{x}_{2i}$	20.77	21.17	21.34	22.91	22.96	23.69	24.82	25.54	25.63	28.73
$\mathbf{x}_{3i}$	19.33	17.04	16.74	19.81	31.92	26.31	25.93	21.96	24.05	25.66

- (i) Given  $\gamma$  and  $\mathbf{x}_{2i}$  for  $i = 1, 2, \dots, n$ , generate random numbers of  $\epsilon_i$  for  $i = 1, 2, \dots, n$ , based on the assumptions:  $\epsilon_i \sim N(0, \exp(\gamma_1 + \gamma_2 \mathbf{x}_{2,i}))$ , where  $(\gamma_1, \gamma_2) = (-2, 0.25)$  is taken.
- (ii) Given  $\beta$ ,  $(\mathbf{x}_{2i}, \mathbf{x}_{3i})$  and  $\epsilon_i$  for  $i = 1, 2, \dots, n$ , we obtain a set of data  $\mathbf{y}_i$ ,  $i = 1, 2, \dots, n$ , from equation (12), where  $(\beta_1, \beta_2, \beta_3) = (10, 1, 1)$  is assumed.
- (iii) Given  $(\mathbf{y}, \mathbf{X})$ , perform M2SE, MLE, BM2SE and BMLE discussed in Sections 2 and 3 in order to obtain the estimates of  $\delta = (\beta, \gamma)$ , denoted by  $\hat{\delta}$ . Note that  $\hat{\delta}$  takes  $\hat{\delta}_{\text{M2SE}}$ ,  $\hat{\delta}_{\text{MLE}}$ ,  $\hat{\delta}_{\text{BM2SE}}$  and  $\hat{\delta}_{\text{BMLE}}$ .
- (iv) Repeat (i) – (iii)  $G$  times, where  $G = 10000$  is taken as mentioned above.
- (v) From  $G$  estimates of  $\delta$ , compute the arithmetic average (AVE), the root mean square error (RMSE), the first quartile (25%), the median (50%), the 3rd quartile (75%) and the interquartile range (IR) for each estimator.

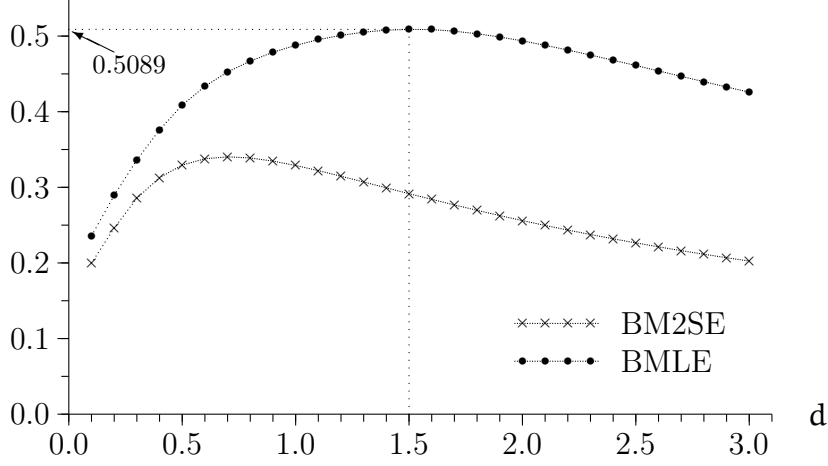
$$\text{AVE} = \frac{1}{G} \sum_{g=1}^G \hat{\delta}_j^{(g)}, \quad \text{RMSE} = \left( \frac{1}{G} \sum_{g=1}^G (\hat{\delta}_j^{(g)} - \delta_j)^2 \right)^{1/2},$$

for  $j = 1, 2, \dots, 5$ , where  $\delta_j$  denotes the  $j$ -th element of  $\delta$  and  $\hat{\delta}_j^{(g)}$  represents the  $j$ -element of  $\hat{\delta}$  in the  $g$ -th simulation run.  $\hat{\delta}$  denotes the estimate of  $\delta$ .

In this section, we compare the Bayesian estimator with the two traditional estimators through our Monte Carlo study.

**Choice of  $(\gamma_*, \Sigma_*)$  and  $\mathbf{d}$ :** For the Bayesian approach, depending on  $(\gamma_*, \Sigma_*)$  we have BM2SE and BMLE, which denote the Bayesian estimators using the

Figure 1: Acceptance Rates in Average



multivariate normal sampling density whose mean and covariance matrix are calibrated on the basis of M2SE and MLE. In this paper, thus, we consider the following sampling density:  $P_*(\gamma) = N(\gamma_*, d\Sigma_*)$ , where  $d$  denotes the tuning parameter and  $(\gamma_*, \Sigma_*)$  takes  $(\gamma_{M2SE}, d\Sigma_{M2SE})$  or  $(\gamma_{MLE}, d\Sigma_{MLE})$ . Generally, for choice of the sampling density, the sampling density should not have too large variance and too small variance. Chib and Greenberg (1995) pointed out that if standard deviation of the sampling density is too low, the Metropolis steps are too short and move too slowly within the target distribution; if it is too high, the algorithm almost always rejects and stays in the same place. Note that  $d$  in (11) represents the variance in the sampling density. As mentioned above, the sampling density should be chosen so that the chain travels over the support of the target density. Since we adopt the independence chain for the sampling density, the acceptance rate in the MH algorithm should be as large as possible. In the case of the independence chain, the large acceptance rate implies that the chain travels over the support of the target density. Therefore, we choose  $(\gamma_*, \Sigma_*)$  and  $d$  which maximizes the acceptance rate. In this section, we consider maximizing the arithmetic average of the acceptance rates obtained from  $G$  simulation runs. The results are in Figure 1, where  $n = 20$ ,  $M = 1000$ ,  $N = 6000$ ,  $G = 10000$  and  $d = 0.1, 0.2, \dots, 3.0$  are taken (choice of  $N$  and  $M$  is discussed in Appendix). In the case of  $(\gamma_*, \Sigma_*) = (\gamma_{MLE}, \Sigma_{MLE})$  and  $d = 1.5$ , the acceptance rate in average is 0.5089, which gives us the largest one. Therefore, in this section,  $P_*(\gamma) = N(\gamma_{MLE}, 1.5\Sigma_{MLE})$  is chosen for the sampling density. That is, hereafter, we compare BMLE with M2SE and MLE.

Table 2: Basic Statistics from the Posterior Distributions

	$\beta_1$	$\beta_2$	$\beta_3$	$\gamma_1$	$\gamma_2$	$N(0, 1)$
IR	1.2693	1.3143	1.2498	1.4710	1.5119	1.349
50%	-0.0519	0.0138	0.0132	-0.0386	0.0228	0
Skewness	0.3212	-0.0466	-0.1529	0.0139	0.0659	0
Kurtosis	3.9670	3.4492	3.6201	2.4516	2.5438	3

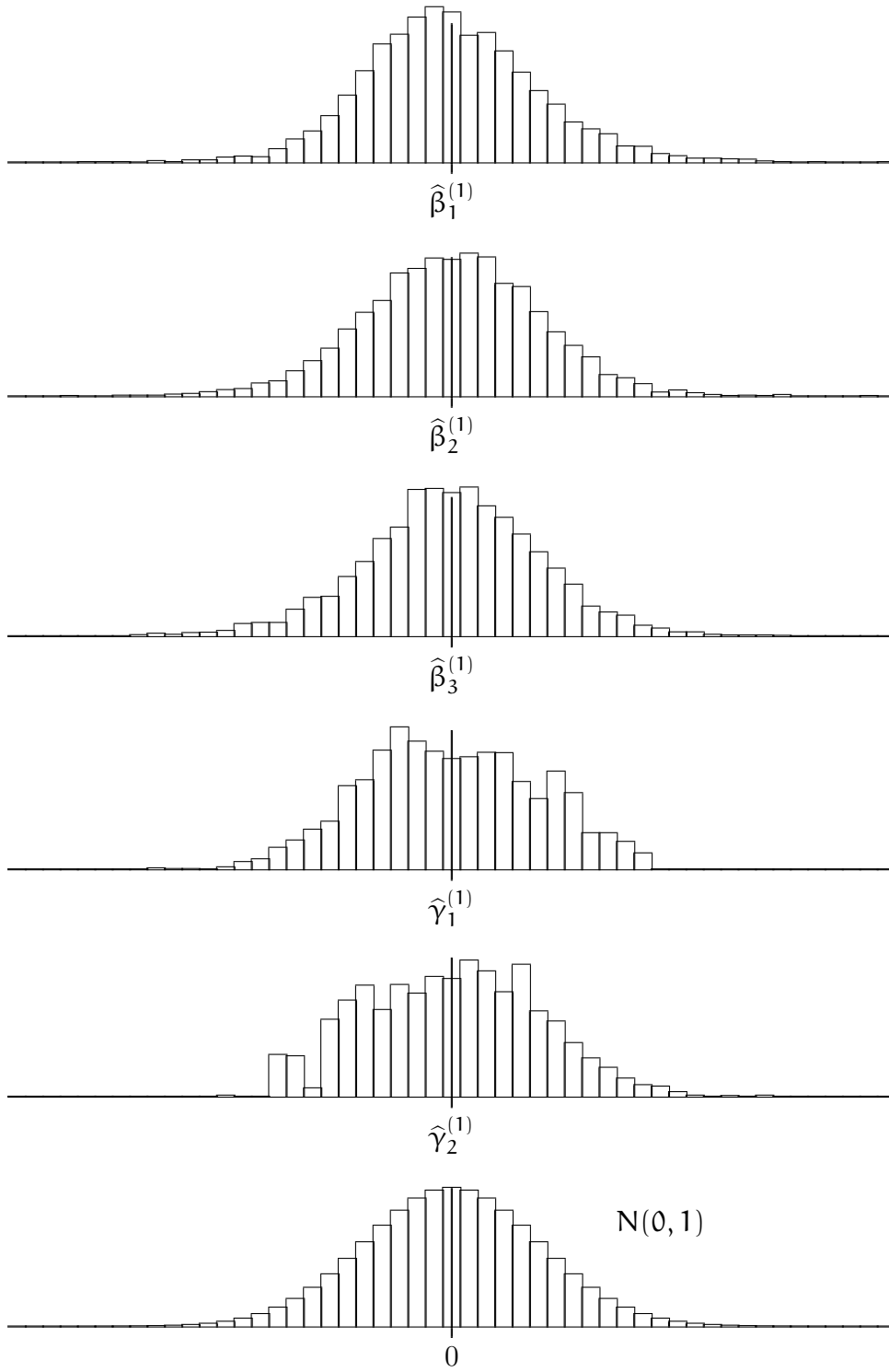
As for computational time, the case of  $n = 20$ ,  $M = 1000$ ,  $N = 6000$  and  $G = 10000$  takes about 75 minutes for each of  $d = 0.1, 0.2, \dots, 3.0$  and each of BM2SE and BMLE, where Pentium III 733MHz CPU personal computer and Watcom FORTRAN 77/32 Compiler (Version 11.0) are utilized.

## 4.2 Results and Discussion

### 4.2.1 Posterior Distributions in the 1st Simulation Run

Taking the 1st simulation run in the simulation procedure (i) – (v), i.e., the case of  $g = 1$ , the posterior distributions of  $\beta$  and  $\gamma$  are shown in Figure 2 to see if they are symmetric or not, where  $(M, N) = (1000, 6000)$  is taken, i.e., each distribution is based on the  $N - M$  random draws.  $\hat{\beta}_1^{(1)}$ ,  $\hat{\beta}_2^{(1)}$ ,  $\hat{\beta}_3^{(1)}$ ,  $\hat{\gamma}_1^{(1)}$  and  $\hat{\gamma}_2^{(1)}$  denote the estimated posterior means in the 1st simulation run. For comparison, all the distributions are normalized, where mean is zero and variance is one. The last distribution in Figure 2 is a standard normal density  $N(0, 1)$ . As for the vertical line in the center, the height of each posterior density is also equal to that of the standard normal density. The posterior density of  $\beta_1$  is slightly skewed to the right, while the densities of  $\beta_2$  and  $\beta_3$  are almost symmetric. The distributions of  $\gamma_1$  and  $\gamma_2$  are quit flat around the posterior mean, which implies that the interquartile range (IR) of the heteroscedasticity parameter  $\gamma$  is larger than that of the regression coefficient  $\beta$ . Based on the normalized random draws used in Figure 2, some basic statistics are shown in Table 2. Each density shown in Figure 2 is characterized by IR of  $\gamma_1$  and  $\gamma_2$ , skewness of  $\beta_1$ , and kurtosis of  $\gamma_1$  and  $\gamma_2$  in Table 2. For comparison, IR, 50%, skewness and kurtosis of the standard normal distribution are in Table 2.

Figure 2: Posterior Distributions in the 1st Simulation Run



#### 4.2.2 Simulation Study Based on the 10000 Simulation Runs

We have shown the posterior densities of  $\beta$  and  $\gamma$  using one sample path, which corresponds to the case of  $g = 1$  in the procedure (i) – (v) of Section 4.1. In this section, through Monte Carlo simulation, the Bayesian estimator BMLE is compared with the traditional estimators M2SE and MLE.

The arithmetic mean (AVE) and the root mean square error (RMSE) have been usually used in Monte Carlo study. Moments of the parameters are needed in the calculation of the mean and the mean square error. However, we cannot assure that these moments actually exist. Therefore, in addition to AVE and RMSE, we also present values for quartiles, i.e., the 1st quartile (25%), median (50%), the 3rd quartile (75%) and the interquartile range (IR). Thus, for each estimator, AVE, RMSE, 25%, 50%, 75% and IR are computed from  $G$  simulation runs. The results are given in Table 3, where BMLE is compared with M2SE and MLE. The case of  $n = 20$ ,  $M = 1000$  and  $N = 6000$  is examined in Table 3. Moreover, a discussion on  $M$  and  $N$  is given in Appendix, where we examine whether  $M = 1000$  and  $N = 6000$  are sufficient.

First, we compare the two traditional estimators, i.e., M2SE and MLE. Judge et al. (1980, pp.141–142) indicated that 2SE of  $\gamma_1$  is inconsistent although 2SE of the other parameters is consistent but asymptotically inefficient. For M2SE, the estimate of  $\gamma_1$  is modified to be consistent. But M2SE is still asymptotically inefficient while MLE is consistent and asymptotically efficient. Therefore, for  $\gamma$ , MLE should have better performance than M2SE in the sense of efficiency. In Table 3, for all the parameters, RMSE and IR of MLE are smaller than those of M2SE. For both M2SE and MLE, AVEs of  $\beta$  are close to the true parameter values. Therefore, it might be concluded that M2SE and MLE are unbiased for  $\beta$  even in the case of small sample. However, the estimates of  $\gamma$  are different from the true values for both M2SE and MLE. That is, AVE and 50% of  $\gamma_1$  are  $-0.951$  and  $-0.905$  for M2SE, and  $-2.708$  and  $-2.680$  for MLE, which are far from the true value  $-2.0$ . Similarly, AVE and 50% of  $\gamma_2$  are  $0.197$  and  $0.199$  for M2SE, which are different from the true value  $0.25$ . But  $0.270$  and  $0.272$  for MLE are slightly larger than  $0.25$ . Thus, the traditional estimators work well for the regression coefficients  $\beta$  but not for the heteroscedasticity parameters  $\gamma$ .

From AVE of  $\beta$ , next, the Bayesian estimator (i.e., BMLE) is compared with the traditional ones (i.e., M2SE and MLE). For all the parameters, we can find from Table 3 that BMLE shows better performance in RMSE and IR than the traditional estimators, because RMSE and IR of BMLE are smaller than those of M2SE and MLE. Furthermore, from AVEs of BMLE, we can

Table 3: The AVE, RMSE and Quartiles:  $n = 20$

		$\beta_1$ 10	$\beta_2$ 1	$\beta_3$ 1	$\gamma_1$ -2	$\gamma_2$ 0.25
M2SE	AVE	10.013	1.002	0.998	-0.951	0.197
	RMSE	7.580	0.417	0.329	3.022	0.144
	25%	5.276	0.734	0.773	-2.723	0.114
	50%	10.057	0.998	1.000	-0.905	0.199
	75%	14.937	1.275	1.215	0.865	0.284
	IR	9.661	0.542	0.442	3.588	0.170
MLE	AVE	10.058	1.000	0.997	-2.708	0.270
	RMSE	7.073	0.383	0.327	2.917	0.135
	25%	5.436	0.744	0.778	-4.438	0.190
	50%	10.027	0.994	0.998	-2.680	0.272
	75%	14.783	1.257	1.215	-0.962	0.352
	IR	9.347	0.512	0.437	3.476	0.162
BMLE	AVE	10.046	1.001	0.997	-1.975	0.248
	RMSE	6.896	0.379	0.324	2.467	0.116
	25%	5.533	0.746	0.778	-3.525	0.177
	50%	10.034	0.997	1.000	-1.953	0.250
	75%	14.646	1.254	1.213	-0.405	0.321
	IR	9.113	0.509	0.436	3.120	0.144

see that the heteroscedasticity parameters as well as the regression coefficients are unbiased in the small sample. Thus, Table 3 also shows the evidence that for both  $\beta$  and  $\gamma$ , AVE and 50% of BMLE are very close to the true parameter values. The values of RMSE and IR also indicate that the estimates are concentrated around AVE and 50%.

For the regression coefficient  $\beta$ , all of the three estimators are very close to the true parameter values. However, for the heteroscedasticity parameter  $\gamma$ , BMLE shows a good performance but M3SE and MLE are poor.

The larger values of RMSE for the traditional counterparts may be due to “outliers” encountered with the Monte Carlo experiments. This problem is also indicated in Zellner (1971, pp.281). Compared with the traditional counterpart, the Bayesian approach is not characterized by extreme values for posterior modal values.

In the next section, to see if there is a real improvement in RMSE and IR or whether the improvements are only due to the sampling error of the simulation, we show the 95% confidence intervals of MSE and IR.

#### 4.2.3 Confidence Intervals of MSE and IR

We present the 95% confidence intervals of MSE, because the confidence interval of RMSE is not easy. The construction of the confidence intervals is briefly discussed as follows.

Confidence Interval of MSE: Let  $\delta_j$  be the  $j$ -th element of  $\delta = (\beta, \gamma)$  and  $\hat{\delta}_j$  be the estimate of  $\delta_j$ . Suppose that  $MSE_j$  represents MSE of  $\hat{\delta}_j$ . Denote the estimate of  $MSE_j$  by  $\overline{MSE}_j$ , which is computed by:

$$\overline{MSE}_j = \frac{1}{G} \sum_{i=1}^G (\hat{\delta}_j^{(i)} - \delta_j)^2,$$

where  $G$  is the number of simulation runs. Accordingly,  $\hat{\delta}_j^{(i)}$  denotes the estimate of  $\delta_j$  in the  $i$ -th simulation run.

Let  $\hat{\sigma}_j^2$  be the estimate of  $\text{Var}(\overline{MSE}_j)$ , which is obtained by:

$$\hat{\sigma}_j^2 = \frac{1}{G} \sum_{i=1}^G ((\hat{\delta}_j^{(i)} - \delta_j)^2 - \overline{MSE}_j)^2.$$

By the central limit theorem, we have:

$$\frac{\sqrt{G}(\overline{MSE}_j - MSE_j)}{\hat{\sigma}_j} \longrightarrow N(0, 1), \quad \text{as } G \longrightarrow \infty.$$

Table 4: 95% Confidence Intervals of MSE and IR:  $n = 20$ 

		$\beta_1$	$\beta_2$	$\beta_3$	$\gamma_1$	$\gamma_2$
M2SE	MSE <sub>L</sub>	55.594	0.168	0.105	8.866	0.0202
	MSE	57.450	0.174	0.108	9.135	0.0209
	MSE <sub>U</sub>	59.306	0.179	0.111	9.404	0.0215
	IR <sub>L</sub>	9.339	0.524	0.428	3.468	0.165
	IR	9.661	0.542	0.442	3.558	0.170
	IR <sub>U</sub>	9.984	0.559	0.457	3.708	0.176
MLE	MSE <sub>L</sub>	48.516	0.142	0.104	8.220	0.0176
	MSE	50.032	0.147	0.107	8.510	0.0183
	MSE <sub>U</sub>	51.547	0.151	0.110	8.799	0.0189
	IR <sub>L</sub>	9.041	0.496	0.423	3.357	0.156
	IR	9.347	0.512	0.437	3.476	0.162
	IR <sub>U</sub>	9.653	0.529	0.452	3.594	0.167
BMLE	MSE <sub>L</sub>	46.173	0.139	0.102	5.887	0.0130
	MSE	47.558	0.143	0.105	6.085	0.0135
	MSE <sub>U</sub>	48.942	0.147	0.108	6.282	0.0139
	IR <sub>L</sub>	8.815	0.492	0.422	3.015	0.140
	IR	9.113	0.509	0.436	3.120	0.144
	IR <sub>U</sub>	9.411	0.525	0.450	3.225	0.149

Therefore, when  $G$  is large enough, the standard error of  $MSE_j$  is approximately given by  $\hat{\sigma}_j/\sqrt{G}$ . That is, the approximated 95% confidence interval is represented by:  $(MSE_j - 1.96\hat{\sigma}_j/\sqrt{G}, MSE_j + 1.96\hat{\sigma}_j/\sqrt{G})$ .

The results are in Table 4, where  $MSE_L$  and  $MSE_U$  denote the lower and upper confidence limits at significance level 5%.  $MSE$  is exactly equal to the square of  $RMSE$  in Table 3. In the case where M2SE is compared with BMLE, for all the parameters except for  $\beta_3$ , the lower limits of M2SE are larger than the upper limits of BMLE. For  $\gamma_1$  and  $\gamma_2$ , the MSEs of BMLE are significantly smaller than those of MLE. Taking an example of  $\gamma_2$ , the MSE of BMLE is 0.0135, which are evidently smaller than 0.0183 for MLE, and 0.0209 for M2SE. The 95% confidence interval is (0.0130, 0.0139) for BMLE, while the lower confident limit is 0.0176 for MLE, and 0.0202 for M2SE. These facts mean that for  $\gamma_1$  and  $\gamma_2$  the MSEs of BMLE are significantly smaller than those of the traditional counterparts.

Confidence Interval of IR: To obtain the confidence interval of IR, we utilize the following formula on the order statistics, which is found in a textbook, for example, Stuart and Ord (1994, p.360).

Let  $X_1, X_2, \dots, X_n$  be independent identically distributed random variables with common probability density  $f(\cdot)$  and cumulative distribution function  $F(\cdot)$ . Assume that  $F(x)$  is strictly monotone for  $0 < F(x) < 1$ . Let  $\mu_p$  be the unique solution in  $x$  of  $F(x) = p$  for some  $0 < p < 1$ , i.e.,  $\mu_p$  is the  $p$ -th quantile. Let  $Y_1, Y_2, \dots, Y_n$  be the order statistics corresponding to the sample  $X_1, X_2, \dots, X_n$ , which satisfies  $Y_1 \leq Y_2 \leq \dots \leq Y_n$ . The  $np$ -th order statistic is denoted by  $Y_{np}$ .

Under the above setup, the distribution of range  $Y_{pn} - Y_{p'n}$  for  $p > p'$  is approximately given by:

$$Y_{pn} - Y_{p'n} \longrightarrow N(\mu_p - \mu_{p'}, \sigma_r^2),$$

where

$$\sigma_r^2 = \frac{p(1-p)}{n(f(\mu_p))^2} + \frac{p'(1-p')}{n(f(\mu_{p'}))^2} - \frac{2(1-p)p'}{nf(\mu_p)f(\mu_{p'})}.$$

Accordingly, the 95% confidence interval of  $\mu_p - \mu_{p'}$  is approximately given by:

$$\left( (Y_{pn} - Y_{p'n}) - 1.96\hat{\sigma}_r, (Y_{pn} - Y_{p'n}) + 1.96\hat{\sigma}_r \right).$$

$\hat{\sigma}_r$  denotes the estimate of  $\sigma_r$ , where  $\mu_p$  and  $\mu_{p'}$  in  $\sigma_r$  is replaced by  $Y_{pn}$  and  $Y_{p'n}$ , respectively,  $f(\cdot)$  is also approximated as a normal density, where mean and variance are replaced by their estimates

In our case, we want to have the 95% confidence interval of IR, where  $p = 0.75$  and  $p' = 0.25$  should be taken. Moreover, note that  $n$  in the above statement corresponds to  $G$  in this paper. The 75% and 25% in Table 3 are utilized for  $Y_{pn}$  and  $Y_{p'n}$ . The mean and variance for  $f(\cdot)$  are estimated from the  $G$  estimates of  $\delta$ , i.e.,  $\hat{\delta}^{(g)}$  for  $g = 1, 2, \dots, G$ . That is, the estimated mean corresponds to AVE in Table 3. Thus, the 95% confidence intervals of IR are obtained in Table 4, where  $IR_L$  and  $IR_U$  denote the lower and upper limits of the 95% confidence interval.

For IR, the same results as MSE are obtained from Table 4. That is, for  $\gamma_1$  and  $\gamma_2$  the IRs of BMLE are significantly smaller than those of M2SE and

Table 5: The AVE, RMSE and Quartiles: the case  $\mathbf{n} = 15$  of BMLE

		$\beta_1$ 10	$\beta_2$ 1	$\beta_3$ 1	$\gamma_1$ -2	$\gamma_2$ 0.25
BMLE	AVE	10.002	0.995	0.999	-2.068	0.251
	RMSE	8.897	0.455	0.351	4.744	0.243
	25%	4.344	0.706	0.766	-4.684	0.105
	50%	10.171	0.995	1.001	-1.774	0.241
	75%	15.849	1.293	1.233	0.795	0.387
	IR	11.505	0.586	0.467	5.479	0.282

MLE, because  $\text{IR}_U$  of BMLE is smaller than  $\text{IR}_L$  of both M2SE and MLE in the case of  $\gamma_1$  and  $\gamma_2$ .

Thus, for the heteroscedasticity parameters, we can conclude that BMLE is significantly improved over M2SE and MLE in the criteria of MSE and IR.

#### 4.2.4 On the Sample Size $\mathbf{n}$

Finally, we examine how the sample size  $\mathbf{n}$  influences precision of the parameter estimates. Since we utilize the exogenous variable  $\mathbf{X}$  shown in Judge et al. (1980), we cannot examine the case where  $\mathbf{n}$  is greater than 20. In order to see the effect of the sample size  $\mathbf{n}$ , the case of  $\mathbf{n} = 15$  is compared with that of  $\mathbf{n} = 20$  in this section.

The case  $\mathbf{n} = 15$  of BMLE is shown in Table 5, which should be compared with BMLE in Table 3. As a result, AVE is very close to the corresponding true parameter value. Therefore, we can conclude from Tables 3 and 5 that the Bayesian estimator is unbiased even in the small sample such as  $\mathbf{n} = 15, 20$ . However, RMSE and IR become large as  $\mathbf{n}$  decreases. That is, for example, RMSEs of  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ ,  $\gamma_1$  and  $\gamma_2$  are given by 6.896, 0.376, 0.324, 2.467 and 0.116 in Table 3, and 8.897, 0.455, 0.351, 4.744 and 0.243 in Table 5. Thus, we can see that RMSE decreases as  $\mathbf{n}$  is large.

## 5 SUMMARY

In this paper, we have examined the multiplicative heteroscedasticity model discussed by Harvey (1976), where the two traditional estimators discussed in Harvey (1976) are compared with the Bayesian estimator. For the Bayesian

approach, we have evaluated the posterior mean by generating random draws from the posterior density, where the Markov chain Monte Carlo methods (i.e., the MH within Gibbs algorithm) are utilized. In the MH algorithm, the sampling density has to be specified. We examine the multivariate normal sampling density, which is the independence chain in the MH algorithm. For mean and variance in the sampling density, we consider using the mean and variance estimated by the two traditional estimators (i.e., M2SE and MLE). The Bayesian estimators with M2SE and MLE are called BM2SE and BMLE in this paper. Through our Monte Carlo study, the results are obtained as follows:

- (i) We compare BM2SE and BMLE with respect to the acceptance rates in the MH algorithm. In this case, BMLE shows higher acceptance rates than BM2SE, which is shown in Figure 1. For the sampling density, we utilize the independence chain through this paper. The high acceptance rate implies that the chain travels over the support of the target density. For the Bayesian estimator, therefore, BMLE is preferred to BM2SE.
- (ii) Taking one sample path, each posterior distribution is displayed in Figure 2. The regression coefficients are distributed to be almost symmetric ( $\beta_1$  is slightly skewed to the left), but the posterior densities of the heteroscedasticity parameters are flat around the center. For the posterior density, the functional form of the regression coefficient  $\beta$  is different from that of the heteroscedasticity parameter  $\gamma$ . These results are also obtained from Table 2.
- (iii) For the traditional estimators (i.e., M2SE and MLE), we have obtained the result that  $\gamma$  of MLE has smaller RMSE than that of M2SE, because for one reason the M2SE is asymptotically less efficient than the MLE. However, from Table 3, the estimates of  $\beta$  are unbiased but those of  $\gamma$  are different from the true parameter values.
- (iv) From Table 3, BMLE performs better than the two traditional estimators in the sense of RMSE and IR, because RMSE and IR of BMLE are smaller than those of the traditional ones for all the cases.
- (v) The 95% confidence intervals of MSE and IR are constructed in Tables 4. We make sure that MSEs and IRs of BMLE are significantly smaller than those of M2SE and MLE for the heteroscedasticity parameter  $\gamma$ . That is, we can see that the difference in RMSE and IR between BMLE and

Table 6: BMLE:  $n = 20$ 

		$\beta_1$ 10	$\beta_2$ 1	$\beta_3$ 1	$\gamma_1$ -2	$\gamma_2$ 0.25
$M = 1000$ $N = 6000$	AVE	10.046	1.001	0.997	-1.975	0.248
	RMSE	6.896	0.379	0.324	2.467	0.116
	50%	10.034	0.997	1.000	-1.953	0.250
	IR	9.113	0.509	0.436	3.120	0.144
$M = 1000$ $N = 11000$	AVE	10.049	1.001	0.997	-1.976	0.248
	RMSE	6.896	0.379	0.324	2.463	0.116
	50%	10.025	0.995	0.999	-1.955	0.250
	IR	9.140	0.510	0.435	3.130	0.145
$M = 5000$ $N = 10000$	AVE	10.046	1.001	0.997	-1.981	0.248
	RMSE	6.913	0.379	0.324	2.465	0.116
	50%	10.013	0.996	1.000	-1.963	0.251
	IR	9.101	0.510	0.435	3.114	0.145

the traditional estimators, shown in Table 3, is not due to the sampling errors of the simulation.

- (vi) As for BMLE, the case of  $n = 15$  is compared with  $n = 20$ . The case  $n = 20$  has smaller RMSE and IR than  $n = 15$ , while AVE and 50% are close to the true parameter values for  $\beta$  and  $\gamma$ . Therefore, it might be expected that the estimates of BMLE go to the true parameter values as  $n$  is large.

## APPENDIX: Are $M = 1000$ and $N = 6000$ Sufficient?

In Section 4, only the case of  $(M, N) = (1000, 6000)$  is examined. In this appendix, we check whether  $M = 1000$  and  $N = 6000$  are sufficient. For the burnin period  $M$ , there are some diagnostic tests, which are discussed in Geweke (1992) and Mengersen, Robert and Guhennenc-Jouyaux (1999). However, since their tests are applicable in the case of one sample path, we cannot utilize them. Because G simulation runs are implemented in this paper, we have G test statistics if we apply the tests. It is difficult to evaluate G testing results at the same time. Therefore, we consider using the alternative approach to see if  $M = 1000$  and  $N = 6000$  are sufficient.

For  $M$  and  $N$ , we consider the following two issues.

- (i) Given fixed  $M = 1000$ , compare  $N = 6000$  and  $N = 10000$ .
- (ii) Given fixed  $N - M = 5000$ , compare  $M = 1000$  and  $M = 5000$ .

(i) examines whether  $N = 6000$  is sufficient, while (ii) checks whether  $M = 1000$  is large enough. If the case of  $(M, N) = (1000, 6000)$  is close to that of  $(M, N) = (1000, 11000)$ , we can conclude that  $N = 6000$  is sufficiently large. Similarly, if the case of  $(M, N) = (1000, 6000)$  is not too different from that of  $(M, N) = (5000, 10000)$ , it might be concluded that  $M = 1000$  is also sufficient.

The results are in Table 6, where AVE, RMSE, 50% and IR are shown for each of the regression coefficients and the heteroscedasticity parameters. Note that the case of  $(M, N) = (1000, 6000)$  in Table 6 is equivalent to that of  $(M, N) = (1000, 6000)$  in Table 3. For comparison, the case of  $(M, N) = (1000, 6000)$  is put in Table 6, again. From Table 6, the three cases are very close to each other. Therefore, we can conclude that both  $M = 1000$  and  $N = 6000$  are large enough in the simulation study shown in Section 4.

## ACKNOWLEDGMENTS

The authors thank two anonymous referees and the editor for their helpful suggestions.

## References

- [1] Amemiya, T. (1995), *Advanced Econometrics*, Cambridge:Harvard University Press.
- [2] Boscardin, W.J. and A. Gelman (1996), "Bayesian Computation for parametric Models of Heteroscedasticity in the Linear Model," in *Advances in Econometrics*, Vol.11 (Part A), edited by Hill, R.C., Connecticut:JAI Press Inc.
- [3] Chib, J. and E. Greenberg (1995), "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, Vol.49, pp.327 – 335.
- [4] Gelman, A., G.O. Roberts and W.R. Gilks (1996), "Efficient Metropolis Jumping Rules," in *Bayesian Statistics*, Vol.5, edited by Bernardo, J.M.,

- J.O. Berger, A.P. Dawid and A.F. Smith, New York:Oxford University Press.
- [5] Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics*, Vol.4, pp.169 – 193 (with discussion), edited by Bernardo, J.M., J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford University Press.
  - [6] Greene, W.H. (1997), *Econometric Analysis*, 3rd ed. Prentice-Hall International, Inc.
  - [7] Harvey, A.C. (1976), "Estimating Regression Models with Multiplicative Heteroscedasticity," *Econometrica*, Vol.44, No.3, pp.461 – 465.
  - [8] Jorgenson, D.W. (1961), "Multiple Regression Analysis of a Poisson Process," *Journal of the American Statistical Association*, Vol.56, No.1, pp.235 – 245.
  - [9] Judge, G., C. Hill, W. Griffiths and T. Lee (1980), *The Theory and Practice of Econometrics*, New York; John Wiley and Sons, Inc.
  - [10] Koop, G., J. Osiewalski and M.F. Steel (1994), "Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function," *Journal of Business and Economic Statistics*, Vol.12, No.3, pp.339 – 346.
  - [11] Mengersen, K.L., C.P. Robert and C. Guhennec-Jouyaux (1999), "MCMC Convergence Diagnostics: A Review," in *Bayesian Statistics*, Vol.6, pp.514 – 440 (with discussion), edited by Bernardo, J.M., J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford University Press.
  - [12] Ohtani, K. (1982), "Small Sample Properties of the Two-step and Three-step Estimators in a Heteroscedastic Linear Regression Model and the Bayesian Alternative," *Economics Letters*, Vol.10, pp.293 – 298.
  - [13] Stuart, A and A.K. Ord (1994), *KenDall's Advanced Theory of Statistics*, Vol.1 (Sixth Edition), Edward Arnold.
  - [14] Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, Vol.22, No.4, pp.1701 – 1762.
  - [15] Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York; John Wiley and Sons, Inc.