APPLICATION OF A NEW GIBBS HIT-AND-RUN SAMPLER TO A CONSTRAINED LINEAR MULTIPLE REGRESSION PROBLEM

by

Ming-Hui Chen and John Deely
Purdue University Canterbury University

Technical Report #92-21

Department of Statistics
Purdue University

May 1992

◇

Application of a New Gibbs Hit-and-Run Sampler to a Constrained Linear Multiple Regression Problem *

Ming-Hui Chen
Department of Statistics
Purdue University

John Deely †
Department of Mathematics
Canterbury University

Abstract

A new method for obtaining solutions to integration problems associated with Bayesian analysis is presented. The application motivating this study arises in the context of predicting one year ahead the new crop of apples and provides naturally a constrained linear multiple regression model. The method used, called the Gibbs Hit-and-Run (GH&R) sampler, is shown to be particularly suited to the constrainted problem and to provide effective and accessible methods for solving the Bayesian inference problems.

Keywords: Bayesian posterior distribution, Gibbs sampler, Hit-and-Run sampler, Importance weighted marginal density estimation, Marginal posterior density, Markov chain, Ordinary least square estimation, Simulation.

*This research was supported in part by NSF grant DMS-8717799 and DMS-8923071 at Purdue University.
†Work done while Visiting Professor in Department of Statistics at Purdue University.

1 Introduction

Recently a considerable amount of research has been devoted to the numerical solution of at worst difficult and at best tedious integration problems arising from Bayesian analysis to multiple parameter model. Various methods with corresponding optimality criteria have been introduced including the seemingly most popular methods, the Gibbs sampler as named by Geman and Geman [1984]. Numerous articles have appeared describing the theory and applications of this method (see, for example, Gelfand and Smith [1990], Tierney [1991]) along with articles describing competing methods such as importance sampling (see, for example, Geweke [1989]) and the Hit-and-Run sampler (see Belisle, Romeijn and Smith [1990], Schmeiser and Chen [1991]).

Because the specific problem being modeled has great influence on which method is "best" and because there is no universally accepted definition of "best", it is difficult to obtain quite general results. Instead specific problems are modeled and the performance of a particular method is documented. Some recent work has been devoted to comparing various methods applied to the same problem (see Chen and Schmeiser [1992]).

Coming out of these papers, there is some evidence to suggest that when the parameter space is constrained, the Gibbs sampler does not perform well in the sense that it requires extremely large numbers of samples to obtain a reasonably small sampling error due to the correlations of the Gibbs observations and more computing time to sample from the conditional distributions due to theirs complexities. Constrained parameter space models arise frequently and naturally in many applied problems and hence it is important to obtain accessible methods to solve the corresponding integration problems associated with Bayesian inferences. Some work in this area has been started; Gelfand, Smith and Lee [1990] and Geyer and Thompson, both of whom use the Gibbs sampler.

It is the purpose of this paper to show how a new Gibbs Hit-and-Run (GH&R) sampler can be used effectively to provide accessible and easily implemented solutions required by the associated Bayesian analysis of a constrained linear multiple regression problem arising naturally in an applied context. We describe the practical problem motivating this study in Section 2 and the Bayesian model used for analysis in Section 3. In Section 4 we propose a new GH&R sampler which generically combines the Gibbs sampler and the Hit-and-Run (H&R) sampler and then prove that the same asymptotic convergence results as given in Schmeiser and Chen [1991] still obtain here. We further demonstrate that the GH&R sampler is well suited for generating a dependent Markov chain of observations from the full joint posterior distribution under the constrained linear multiple regression model. In Section 5 the results of the computations are reported, robustness and predictions are discussed, the marginal posterior densities are estimated by the importance weighted marginal density estimation method and comparisons with ordinary least square estimation method are made.

2 Data and Model

The New Zealand Apple and Pear Marketing Board is a statutory body which amongst other responsibilities negotiates and arranges contracts for all exporting of the New Zealand apples throughout the world. In effect this means that all of the more than 1500 apple growers in the New Zealand are joined together as one grower when dealing with the international export market. The justification or lack thereof for such a board is not our concern here. Rather it is to derive a model for forecasting one year ahead the total crop of apples of any particular variety that will be available

for both export and local market consumption. The Board has a rich data base consisting of the historical record of more than thirty years for the total submissions from each individual grower for each variety, but more importantly, for the last five years, the total number of trees each grower has at each age for each variety. Ordinary time series modeling of the total submissions record does not provide a reasonable error bound. Hence a realistic data base that can be readily used for forecasting purposes consists of the tree numbers over the last four years and the total submissions for a grower during that time. By tree numbers we mean that the number of trees at each age is known and tabulated for a particular variety of apple and for an individual grower.

The forecasting model that has been evolved so far uses this data base to assign each grower for a particular year into a quality category. This quality category is determined on the basis of a computable QC index which is defined to be the average amount of fruit per average year of age; that is, it is the ratio of the average fruit per tree to the average age of the orchard. Further investigation indicated that seven categories of quality would be adequate. No grower is assigned to one category for all time; rather, in each year a grower can be in any category such an assignment being dependent upon many factors, such as weather, soil condition and farming practices. Instead of trying to unravel all the complexities involved with these factors, we simply compute a quality index for that year and thus assign a grower to a category on the basis of the individual year's data.

For illustrative purposes here we will report only a subset of data arising from one geographical region of growers within one category and for one variety. This region provided over 1500 grower years of data for the last four years, but we will use here only the 207 which fell in one particular category. Thus for each grower in that category the model developed suggests quite naturally that the average number of apples (in cartons) is given by a linear multiple regression model where the dependent variable is a ten dimensional vector consisting of the numbers of trees at each age and the regression coefficients are the prediction averages for each age, that is

$$y = E(Y|\underline{x}) + \epsilon, \tag{2.1}$$

and

$$E(Y|\underline{x}) = \sum_{j=1}^{10} \beta_j x_j, \tag{2.2}$$

where $\epsilon \sim N(0, \sigma^2)$, x_j = number of trees at age j and β_j = average number of cartons produced by trees at age j for $j = 1, 2, \cdots, 10$. Furthermore it is easy to see that this model must be constrained by

$$0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_{10}, \tag{2.3}$$

since growers do not allow poor trees to persist on average. So the problem then is to find the appropriate regression equation for this data and subsequently to use it to predict for the coming year what fruit is to be produced by any grower who is deemed to be in this category which in this paper we call category one. It is also of interest to obtain bounds for the beta coefficients.

The problem of assigning growers for the coming year to a category is a separate issue and will not be discussed here. The main problem that is of concern here is to find the best set of coefficients in (2.2) satisfying the constraints of (2.3) and to obtain the best error bounds that can be derived from the information provided.

It may be the case that one grower provides data for more than one year if that grower happened to be in the same category for more than just one year. Whereas there is considerable dependence

between the category of a grower from year to year, the prediction of a grower's submissions conditioned on a category is assumed to be independent from year to year and from other submissions from growers in the same category for the same year.

We will denote the data set arising from these 207 growers by $(y_i, \underline{x}_i)$, where $\underline{x}_i = (x_{1,i}, x_{2,i}, \cdots, x_{10,i})'$, $i = 1, 2, \cdots, n$, and $n = 207$ in this case.

In the next section, we will use the Bayesian method to analyze Model (2.1) and (2.2) with constraints (2.3) based on the above data.

3 Bayesian Analysis of the Constrained Linear Multiple Regression Model

For the constrained linear multiple regression model, which is described in Equations (2.1) and (2.2), we choose independent noninformative priors for $\underline{\beta}$ and σ^2. Let

$$S = \{(\beta_1, \beta_2, \cdots, \beta_{10})' : 0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_{10}, \underline{\beta} \in R^{10}\}. \tag{3.1}$$

Then the noninformative prior for $\underline{\beta}$ is proportional to

$$\pi_1(\underline{\beta}) = 1 \cdot I_S(\underline{\beta}), \tag{3.2}$$

where $I_S(\underline{\beta}) = 1$ if $\underline{\beta} \in S$ and 0 otherwise. The noninformative prior for σ^2 is proportional to

$$\pi_2(\sigma^2) = \frac{1}{\sigma^2}, \sigma^2 > 0 \tag{3.3}$$

So, the prior for $\underline{\beta}$ and σ^2 is proportional to

$$\pi(\underline{\beta}, \sigma^2) = \pi_1(\underline{\beta})\pi_2(\sigma^2) = \frac{1}{\sigma^2} I_S(\underline{\beta}). \tag{3.4}$$

According to Equations (2.1) & (2.2) and the data set described in Section 2, the likelihood function, ignoring the constant, is

$$L(\underline{\beta}, \sigma^2, data) = \frac{1}{(\sigma^2)^{\frac{207}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{207} \left(y_i - \sum_{j=1}^{10} \beta_j x_{j,i}\right)^2\right\}, \quad \underline{\beta} \in S, \sigma^2 > 0. \tag{3.5}$$

From Equation (3.4) and (3.5), the posterior density function is given as follows:

$$\begin{aligned} &\pi(\underline{\beta}, \sigma^2 | data) = c(\underline{x}) L(\underline{\beta}, \sigma^2, \ data)\pi(\underline{\beta}, \sigma^2) \\ &= c(\underline{x}) \frac{1}{(\sigma^2)^{104.5}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{207} \left(y_i - \sum_{j=1}^{10} \beta_j x_{j,i}\right)^2\right\} I_S, \quad \sigma^2 > 0, \end{aligned} \tag{3.6}$$

where $c(\underline{x})$ is the normalization constant. The respective posterior means of $\underline{\beta}$ and σ^2 will be denoted by

$$\hat{\underline{\beta}} = \text{posterior mean of } \underline{\beta} = E^{\pi(\underline{\beta}, \sigma^2 | data)}\left(\underline{\beta}\right), \tag{3.7}$$

and

$$\hat{\sigma}^2 = \text{posterior mean of } \sigma^2 = E^{\pi(\underline{\beta},\sigma^2|data)}\left(\sigma^2\right) \tag{3.8}$$

and the posterior covariances of $\underline{\beta}$ by

$$\begin{aligned} &\hat{V}_{ij} = \text{ posterior covariance of } \beta_i,\ \beta_j \\ = \ &E^{\pi(\underline{\beta},\sigma^2|data)}(\beta_i\beta_j) - E^{\pi(\underline{\beta},\sigma^2|data)}(\beta_i)\,E^{\pi(\underline{\beta},\sigma^2|data)}(\beta_j), \quad \text{for } i,j = 1,\cdots,10. \end{aligned} \tag{3.9}$$

Let $\underline{\beta}_{-j} = (\beta_1,\cdots,\beta_{j-1},\beta_{j+1},\cdots,\beta_{10})'$ and

$$\begin{aligned} S_{-j}(\beta_j) = \{(\beta_1,\cdots,\beta_{j-1},\beta_{j+1},\cdots,\beta_{10})' \ : \ &0 \le \beta_1 \le \cdots \le \beta_{j-1} \le \beta_j \le \beta_{j+1} \le \cdots \le \beta_{10}, \\ &\underline{\beta} \in R^{10}, \text{for prefixed } \beta_j\}. \end{aligned} \tag{3.10}$$

Then the marginal posterior density for β_j is

$$\pi_j(\beta_j|data) = \int_0^\infty \int_{S_{-j}(\beta_j)} \frac{c(\underline{x})}{(\sigma^2)^{104.5}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{207}\left(y_i - \sum_{l=1}^{10}\beta_l x_{l,i}\right)^2\right\} d\underline{\beta}_{-j}d\sigma^2. \tag{3.11}$$

In order to predict fruit to be produced by any grower deemed to be in category one for the coming year, we need to derive the predictive distribution for an individual grower. Let $(y_\nu, \underline{x}_\nu)$ denote respectively the total number of cartons of fruit produced by the tree numbers vector in the coming year. Then

$$f_\nu(y|\underline{\beta},\sigma^2,\underline{x}_\nu) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y-\underline{\beta}'\underline{x}_\nu)^2}{2\sigma^2}\right\}. \tag{3.12}$$

Then the predictive density of Y_ν is

$$\begin{aligned} f_\nu(y) \ &= \ \int_S\int_0^\infty f_\nu(y|\underline{\beta},\sigma^2,\underline{x}_\nu)\pi(\underline{\beta},\sigma^2|\ data\ )d\sigma^2 d\underline{\beta} \\ &= \ \int_S\int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(y-\underline{\beta}'\underline{x}_\nu)^2}{2\sigma^2}\} \\ &\qquad \cdot\frac{c(\underline{x})}{(\sigma^2)^{104.5}}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{207}(y_i - \sum_{j=1}^{10}\beta_j x_{j,i})^2\}d\sigma^2 d\underline{\beta}. \end{aligned} \tag{3.13}$$

Therefore, the predictive mean and variance of Y_ν are

$$\mu_{Y_\nu} = \text{predictive mean of } Y_\nu \ = E^{f_\nu}(Y_\nu) = E^{\pi(\underline{\beta},\sigma^2|data)}\left(\underline{\beta}'\underline{x}_\nu\right) \tag{3.14}$$

and

$$\begin{aligned} &V_{Y_\nu} = \text{predictive variance of } Y_\nu \ = E^{f_\nu}(Y_\nu - \mu_{Y_\nu})^2 \\ &= E^{\pi(\underline{\beta},\sigma^2|data)}\left(\sigma^2\right) + \left[E^{\pi(\underline{\beta},\sigma^2|data)}\left\{(\underline{\beta}'\underline{x}_\nu)^2\right\} - \left\{E^{\pi(\underline{\beta},\sigma^2|data)}\left(\underline{\beta}'\underline{x}_\nu\right)\right\}^2\right]. \end{aligned} \tag{3.15}$$

Equation (3.15) says that the predictive variance of Y_ν is the sum of the posterior mean of σ^2 and the posterior variance of $\underline{\beta}'\underline{x}_\nu$. This illustrates the intuitive appealing fact that there are two sources of error for predicting y_ν, one is the random error, and another is the method error.

In order to explore the Bayesian properties of this constrained linear multiple regression model, we have to compute the quantities of interest, e.g., posterior means, marginal posterior densities, and predictive distribution. However, it is nearly impossible to get the analytic results for those quantities because of the constraints. So we will develop a new Markov chain sampling scheme to evaluate them numerically. We also use a new importance weighted marginal density estimation method (see, Chen [1992]) to obtain estimators of the marginal posterior densities.

4 Sampling Approaches

To determine the properties of relevant Bayesian distributions, three Markov chain sampling approaches have been developed recently. These are the Gibbs sampler (Geman and Geman [1984], Gelfand and Smith [1990]), the Hit-and-Run (H&R) sampler (Belisle, Romeijn and Smith [1990], Schmeiser and Chen [1991], Chen and Schmeiser [1992]), and the Metropolis sampler (Hastings [1970], Tierney [1991], Müller [1991]). As discussed in Chen and Schmeiher [1992], the Gibbs sampler works well when the components of a random vector are nearly independent, and are easily sampled from the conditional distribution. They also point out that if the components of a random vector are highly correlated, the H&R sampler performs better than the Gibbs sampler. These facts indicate that there may exist a Markov chain sampler, which generically combines the Gibbs sampler and the H&R sampler.

In this section, we propose such a new Markov chain sampler named the Gibbs Hit-and-Run (GH&R) sampler for sampling $(\sigma^2, \underline{\beta})$ from the posterior distribution $\pi(\underline{\beta}, \sigma^2|data)$ without knowing the normalization constant $c(\underline{x})$. Intuitively the GH&R sampler operates as follows. Firstly given any starting point $(\sigma_0^2, \underline{\beta}_0)$ such that $\sigma_0^2 > 0$ and $\underline{\beta}_0 \in S$, we generate σ_1^2 from the conditional distribution $\pi(\sigma^2|\underline{\beta}_0, data)$, which is an Inverse Gamma

$$\mathcal{IG}\left(104.5, \frac{2}{\sum_{i=1}^{207}(y_i - \sum_{j=1}^{10}\beta_{0,j}x_{j,i})^2}\right) \tag{4.1}$$

(e.g., see Berger [1985], p. 561), where $\underline{\beta}_0 = (\beta_{1,0}, \cdots, \beta_{10,0})'$. Secondly we use the H&R sampler to generate $\underline{\beta}_1$ from $\pi(\underline{\beta}|\sigma_1^2, data)$, which is a truncated multi-variate normal distribution and finish one whole cycle with a new point $(\sigma_1^2, \underline{\beta}_1)$. The technical details of the GH&R sampler for sampling from the posterior distribution $\pi(\underline{\beta}, \sigma^2|data)$ are given in *Algorithm 1*.

Algorithm 1

step 0. Choose a starting point $\sigma_0^2 > 0$, $\underline{\beta}_0 \in S$, and set i=0.

step 1. Generate σ_{i+1}^2 from $\mathcal{IG}\left(104.5, \frac{2}{\sum_{l=1}^{207}(y_l - \sum_{j=1}^{10}\beta_{j,i}x_{j,l})^2}\right)$.

step 2. Generate a uniformly distributed unit-length direction $\underline{d}_i \stackrel{def}{=} (d_{1,i}, d_{2,i}, \cdots, d_{10,i})'$.

step 3. Find the set $S_i = (R_1^i, R_2^i)$, where

$$R_1^i \stackrel{def}{=} \inf_{\lambda}\left\{\lambda : \underline{\beta}_i + \lambda\underline{d}_i \in S\right\}, \text{ and } R_2^i \stackrel{def}{=} \sup_{\lambda}\left\{\lambda : \underline{\beta}_i + \lambda\underline{d}_i \in S\right\}.$$

step 4. Generate a signed distance λ_i from density

$$\pi_i(\lambda) = \frac{\pi(\underline{\beta}_i + \lambda \underline{d}_i | \sigma^2_{i+1}, data)}{\int_{R_1^i}^{R_2^i} \pi(\underline{\beta}_i + u\underline{d}_i | \sigma^2_{i+1}, data) du}, \quad \lambda \in (R_1^i, R_2^i). \tag{4.2}$$

step 5. Set $\underline{\beta}_{i+1} = \underline{\beta}_i + \lambda_i \underline{d}_i$ and set i=i+1. Go to step 1.

A random unit-length direction $\underline{d}_i$ can be generated in Step 2 by independently generating $z_l \sim N(0,1)$ and setting $d_i^l = z_l \left(\sum_{j=1}^{10} z_j^2\right)^{-\frac{1}{2}}, l = 1, 2, ..., 10$ (e.g., see Devroye [1986, Section 4.2]).

Notice that since R_1^i and R_2^i can not both be infinity, $\pi_i(\lambda)$ in Equation (4.2) is a probability density function of the truncated normal $TN(\mu_i, (\sigma_i^N)^2, R_1^i, R_2^i)$ with

$$\text{mean} = \mu_i = \frac{\sum_{l=1}^{207} \left[\left(y_l - \sum_{j=1}^{10} \beta_{j,i} x_{j,l} \right) \sum_{j=1}^{10} d_{j,i} x_{j,l} \right]}{\sum_{l=1}^{207} \left(\sum_{j=1}^{10} d_{j,i} x_{j,l} \right)^2}, \tag{4.3}$$

and

$$\text{variance} = (\sigma_i^N)^2 = \frac{\sigma^2_{i+1}}{\sum_{l=1}^{207} \left(\sum_{j=1}^{10} d_{j,i} x_{j,l} \right)^2}. \tag{4.4}$$

A random variable λ_i of the truncated univariate normal $TN(\mu_i, (\sigma_i^N)^2, R_1^i, R_2^i)$ can be generated by generating $U \sim U(0,1)$ and setting

$$\lambda_i \stackrel{def}{=} \mu_i + \sigma_i^N \Phi^{-1} \left(\Phi(\frac{R_1^i - \mu_i}{\sigma_i^N}) + U \left(\Phi(\frac{R_2^i - \mu_i}{\sigma_i^N}) - \Phi(\frac{R_1^i - \mu_i}{\sigma_i^N}) \right) \right), \tag{4.5}$$

where Φ is the $N(0,1)$ *cdf*, and Φ^{-1} is the inverse $N(0,1)$ *cdf* (e.g., see Devroye [1986, p.39]). Another way to generate λ_i is the mixed rejection algorithm for truncated univariate normal sampling (see Geweke [1991, Section 2]).

The advantage of *Algorithm 1* is that only one univariate truncated normal random variate, which it is expensive to generate, is needed for getting a new Markov chain state. For the primary Gibbs sampler, 10 univariate truncated normal random variates are needed for completing one Gibbs sampling cycle. Since β_j is the average number of cartons produced by trees at age j, $j = 1, 2, \cdots, 10$, it is likely that they would be correlated. Thus to obtain the same standard errors in the estimates when sampling $\underline{\beta}$ from its conditional posterior distribution given σ^2, the H&R sampler will require fewer iterations than the Gibbs sampler and it is in this sense that the H&R sampler performs better.

Now, we consider the asymptotic properties of the homogeneous Markov chain $\{(\sigma_i^2, \underline{\beta}_i), i \geq 0\}$ generated by *Algorithm 1*. Since *Algorithm 1* combines the Gibbs sampler and the H&R sampler, we can not obtain a time reversible Markov chain which would easily assure the asymptotic convergence results. However, it is shown in Schmeiser and Chen [1991] that if the probability measure induced by $\pi(\underline{\beta}, \sigma^2 | data)$ has the certain invariant property (see Lemma 4.2 below), the asymptotic convergence results then obtain. Hence we now prove that $\pi(\underline{\beta}, \sigma^2 | data)$ does process this property.

Firstly, we derive the one-step transition probability density since the Markov chain $\{(\sigma_i^2, \underline{\beta}_i), i \geq 0\}$ always moves to a new state from the current state. This transition probability density at $\sigma_{i+1}^2 = \sigma^2$ and $\underline{\beta}_{i+1} = \underline{\beta}$ given $\sigma_i^2 = \sigma^{*2}$ and $\underline{\beta}_i = \underline{\beta}^*$ is

$$\begin{aligned} p(\sigma^2, \underline{\beta}|\sigma^{*2}, \underline{\beta}^*) &= \frac{\pi(\underline{\beta}^*, \sigma^2|data)}{\pi(\underline{\beta}^*|data)} \\ &\quad \cdot \frac{2}{C_{10}||\underline{\beta} - \underline{\beta}^*||^9} \cdot \frac{\pi(\underline{\beta}|\sigma^2, data)}{\int_{R_1}^{R_2} \pi(\underline{\beta}^* + u\frac{\underline{\beta}-\underline{\beta}^*}{||\underline{\beta}-\underline{\beta}^*||}|\sigma^2, data)du}, \end{aligned} \tag{4. 6}$$

for all $\sigma^2 \neq \sigma^{*2} > 0, \underline{\beta} \neq \underline{\beta}^* \in S$, where $C_{10} = \frac{2\pi^5}{\Gamma(5)}$ is the surface area of the 10-dimensional unit hypersphere, $||\underline{\beta} - \underline{\beta}^*|| = \sqrt{\sum_{j=1}^{10}(\beta_j - \beta_j^*)^2}$, $\pi(\underline{\beta}^*|data)$ is the marginal posterior density of $\underline{\beta}$ at $\underline{\beta} = \underline{\beta}^*$, and

$$R_1 \stackrel{def}{=} \inf_{\lambda}\left\{\lambda : \underline{\beta}^* + \lambda\frac{\underline{\beta} - \underline{\beta}^*}{||\underline{\beta} - \underline{\beta}^*||} \in S\right\} \text{ and } R_2 \stackrel{def}{=} \sup_{\lambda}\left\{\lambda : \underline{\beta}^* + \lambda\frac{\underline{\beta} - \underline{\beta}^*}{||\underline{\beta} - \underline{\beta}^*||} \in S\right\}. \tag{4.7}$$

A requirement of Schmeiser and Chen [1991] is that the transition probability density $p(\sigma^2, \underline{\beta}|\sigma^{*2}, \underline{\beta}^*) > 0$. It is apparent from the above that $p(\sigma^2, \underline{\beta}|\sigma^{*2}, \underline{\beta}^*) > 0$ for $\sigma^2 \neq \sigma^{*2} > 0, \underline{\beta} \neq \underline{\beta}^* \in S$ and we can define $p(\sigma^{*2}, \underline{\beta}^*|\sigma^{*2}, \underline{\beta}^*) > 0$ since the Lebesgue measure of one single point $(\sigma^{*2}, \underline{\beta}^*)$ is zero.

Secondly we will prove that if we start $(\sigma^2, \underline{\beta})$ from the joint density $\pi(\underline{\beta}, \sigma^2|data)$, the joint distribution of the observation at the next iteration generated by *Algorithm 1* is still $\pi(\cdot, \cdot|data)$, that is,

Lemma 4.1

$$\int_S \int_0^\infty \pi(\underline{\beta}^*, \sigma^{*2}|data)p(\sigma^2, \underline{\beta}|\sigma^{*2}, \underline{\beta}^*)d\sigma^{*2}d\underline{\beta}^* = \pi(\underline{\beta}, \sigma^2|data). \tag{4.8}$$

Proof: From Equation (4.6),

$$\begin{aligned} &\int_S \int_0^\infty \pi(\underline{\beta}^*, \sigma^{*2}|data)p(\sigma^2, \underline{\beta}|\sigma^{*2}, \underline{\beta}^*)d\sigma^{*2}d\underline{\beta}^* \\ = &\int_S \int_0^\infty \pi(\underline{\beta}^*, \sigma^{*2}|data)\frac{\pi(\underline{\beta}^*, \sigma^2|data)}{\pi(\underline{\beta}^*|data)}d\sigma^{*2} \\ &\frac{2}{C_{10}||\underline{\beta} - \underline{\beta}^*||^9} \cdot \frac{\pi(\underline{\beta}|\sigma^2, data)}{\int_{R_1}^{R_2} \pi(\underline{\beta}^* + u\frac{\underline{\beta}-\underline{\beta}^*}{||\underline{\beta}-\underline{\beta}^*||}|\sigma^2, data)du}d\underline{\beta}^* \\ = &\int_S \pi(\underline{\beta}^*, \sigma^2|data) \cdot \frac{2\pi(\underline{\beta}|\sigma^2, data)}{C_{10}||\underline{\beta} - \underline{\beta}^*||^9 \int_{R_1}^{R_2} \pi(\underline{\beta}^* + u\frac{\underline{\beta}-\underline{\beta}^*}{||\underline{\beta}-\underline{\beta}^*||}|\sigma^2, data)du}d\underline{\beta}^* \\ = &\pi(\sigma^2|data)\int_S \pi(\underline{\beta}^*|\sigma^2, data) \cdot \frac{2\pi(\underline{\beta}|\sigma^2, data)}{C_{10}||\underline{\beta} - \underline{\beta}^*||^9 \int_{R_1}^{R_2} \pi(\underline{\beta}^* + u\frac{\underline{\beta}-\underline{\beta}^*}{||\underline{\beta}-\underline{\beta}^*||}|\sigma^2, data)du}d\underline{\beta}^*. \end{aligned} \tag{4. 9}$$

Since the H&R sampler is used to generate a new $\underline{\beta}$ from the conditional distribution $\pi(\cdot|\sigma^2, data)$, thus Lemma 2.2 in Chen and Schmeiser [1992] yields

$$\int_S \pi(\underline{\beta}^*|\sigma^2, data) \cdot \frac{2\pi(\underline{\beta}|\sigma^2, data)}{C_{10}||\underline{\beta} - \underline{\beta}^*||^9 \int_{R_1}^{R_2} \pi(\underline{\beta}^* + u\frac{\underline{\beta}-\underline{\beta}^*}{||\underline{\beta}-\underline{\beta}^*||}|\sigma^2, data)du}d\underline{\beta}^* = \pi(\underline{\beta}|\sigma^2, data). \tag{4.10}$$

Therefore,

$$\begin{aligned} & \int_S \int_0^\infty \pi(\underline{\beta}^*, \sigma^{*2}|data) p(\sigma^2, \underline{\beta}|\sigma^{*2}, \underline{\beta}^*) d\sigma^{*2} d\underline{\beta}^* \\ = \; & \pi(\sigma^2|data)\pi(\underline{\beta}|\sigma^2, data) = \pi(\underline{\beta}, \sigma^2|data). \end{aligned} \tag{4. 11}$$

■

Let $R^+ \stackrel{def}{=} \{\sigma^2 : \sigma^2 > 0\}$, let $\mathcal{B}_+^{11}$ denote the Borel sets of $R^+ \times S$. For every $A \in \mathcal{B}_+^{11}$, let the probability measure Π, defined by $\pi(\underline{\beta}, \sigma^2|data)$, be

$$\Pi(A) = \int_A \pi(\underline{\beta}, \sigma^2|data) d\sigma^2 d\underline{\beta}. \tag{4.12}$$

Then $\Pi(R^+ \times S) = \int_{R^+ \times S} \pi(\underline{\beta}, \sigma^2|data) d\sigma^2 d\underline{\beta} = 1$. Finally we have

Lemma 4.2 *The Probability measure* Π *is invariant for the Markov chain* $\{(\sigma_i^2, \underline{\beta}_i), i \geq 0\}$, *i.e.,*

$$\Pi(A) = \int_{R^+ \times S} \int_A p(\sigma^2, \underline{\beta}|\sigma^{*2}, \underline{\beta}^*) \pi(\underline{\beta}^*, \sigma^{*2}|data) d\sigma^2 d\underline{\beta} d\sigma^{*2} d\underline{\beta}^*, \tag{4.13}$$

for every $A \in \mathcal{B}_+^{11}$.

Proof: By Lemma 4.1 and Fubini's Theorem, we have

$$\begin{aligned} \text{The RHS of Equation (4.13)} &= \int_A \left\{ \int_S \int_0^\infty p(\sigma^2, \underline{\beta}|\sigma^{*2}, \underline{\beta}^*) \pi(\underline{\beta}^*, \sigma^{*2}|data) d\sigma^{*2} d\underline{\beta}^* \right\} d\sigma^2 d\underline{\beta} \\ &= \int_A \pi(\underline{\beta}, \sigma^2|data) d\sigma^2 d\underline{\beta} = \Pi(A). \end{aligned} \tag{4. 14}$$

■

From the above we can see that $\pi(\underline{\beta}, \sigma^2|data)$ does satisfy the requirements given in Schmeiser and Chen [1991]. Hence we have the following main result.

Proposition 4.1 *If* h *is integrable with respect to* $\pi(\cdot, \cdot|data)$, *i.e.,*

$$\int_{R^+ \times S} |h(\sigma^2, \underline{\beta})| \pi(\underline{\beta}, \sigma^2|data) d\sigma^2 d\underline{\beta} < \infty,$$

then for every fixed $0 \leq j_0 < \infty$

$$\lim_{n \to \infty} \frac{1}{n - j_0 + 1} \sum_{j=j_0}^{n} h(\sigma_j^2, \underline{\beta}_j) = E^{\pi(\underline{\beta}, \sigma^2|data)}(h) \quad a.s., \tag{4.15}$$

where $E^{\pi(\underline{\beta}, \sigma^2|data)}(h) = \int_{R^+ \times S} h(\sigma^2, \underline{\beta}) \pi(\underline{\beta}, \sigma^2|data) d\sigma^2 d\underline{\beta}$.

The conclusion of Proposition 4.1 assures us that the use of the Markov chain sample generated by *Algorithm 1* to obtain the approximations of the unavailable Bayesian estimates is valid.

Although *Algorithm 1* is stated specifically for the constrained linear multiple regression model, it can be stated in general terms. Denote densities by square brackets, so that joint, conditional and marginal forms for random variables $\underline{\Theta}_1$, $\underline{\Theta}_2$, appear as $[\underline{\Theta}_1, \underline{\Theta}_2]$, $[\underline{\Theta}_1|\underline{\Theta}_2]$ and $[\underline{\Theta}_2]$, respectively. Let a k-dimensional random variate $\underline{\Theta} = (\Theta_1, \Theta_2, \cdots, \Theta_k)'$ be distributed as $\pi(\underline{\theta})$ with the support S^*. Set $0 = j_0 \leq j_1 \leq j_2 \leq \cdots \leq j_{k^*} = k$, $\underline{\Theta}_{j_l} = (\Theta_{j_{l-1}+1}, \cdots, \Theta_{j_l})', l = 1, 2, \cdots, k^*$. Then the basic scheme of the GH&R sampler is as follows.

Algorithm Gibbs Hit-and-Run Sampler

step 0. Choose an arbitrary starting point $\underline{\Theta}_{(0)} = (\Theta_{1,0}, \Theta_{2,0}, \cdots, \Theta_{k,0})' \in S^*$, and set i=0.

step 1. Generate $\underline{\Theta}_{j_1,i+1} = (\Theta_{1,i+1}, \Theta_{2,i+1}, \cdots, \Theta_{j_1,i+1})'$.

- Generate $\Theta_{1,i+1} \sim [\Theta_1 | \Theta_{2,i}, \cdots, \Theta_{k,i}]$;
- Generate $\Theta_{2,i+1} \sim [\Theta_2 | \Theta_{1,i+1}, \Theta_{3,i}, \cdots, \Theta_{k,i}]$;

  ... ...
- Generate $\Theta_{j_1,i+1} \sim [\Theta_{j_1} | \Theta_{1,i+1}, \Theta_{2,i+1}, \cdots, \Theta_{j_1-1,i+1}, \Theta_{j_1+1,i}, \cdots, \Theta_{k,i}]$.

step 2. Use the H&R sampler to generate each of

$$\underline{\Theta}_{j_2,i+1} \sim [\underline{\Theta}_{j_2} | \underline{\Theta}_{j_1,i+1}, \underline{\Theta}_{j_3,i}, \cdots, \underline{\Theta}_{j_{k^*},i}], \cdots, \underline{\Theta}_{j_{k^*},i+1} \sim [\underline{\Theta}_{j_{k^*}} | \underline{\Theta}_{j_1,i+1}, \cdots, \underline{\Theta}_{j_{k^*-1},i+1}],$$

e.g., to generate $\underline{\Theta}_{j_2,i+1} \sim [\underline{\Theta}_{j_2} | \underline{\Theta}_{j_1,i+1}, \underline{\Theta}_{j_3,i}, \cdots, \underline{\Theta}_{j_{k^*},i}]$ as follows:

- Generate a uniformly distributed unit-length direction $\underline{d}_{j_2,i} \stackrel{def}{=} (d^1_{j_2,i}, d^2_{j_2,i}, \cdots, d^{j_2-j_1}_{j_2,i})'$;
- Find the set $S_{j_2,i} \stackrel{def}{=} \{\lambda \in R | \underline{\Theta}_{j_2,i} + \lambda \underline{d}_{j_2,i} \in S_{j_2}\}$, where S_{j_2} is the support of the conditional distribution $[\Theta_{j_1} | \underline{\Theta}_{j_1,i+1}, \underline{\Theta}_{j_3,i}, \cdots, \underline{\Theta}_{j_{k^*},i}]$ with the density function

$$\pi(\underline{\theta}_{j_2} | \underline{\theta}_{j_1,i+1}, \underline{\theta}_{j_3,i}, \cdots, \underline{\theta}_{j_{k^*},i});$$

- Generate a signed distance $\lambda_{j_2,i}$ from density

$$\pi_{j_2,i}(\lambda) = \frac{\pi(\underline{\theta}_{j_2,i} + \lambda \underline{d}_{j_2,i} | \underline{\theta}_{j_1,i+1}, \underline{\theta}_{j_3,i}, \cdots, \underline{\theta}_{j_{k^*},i})}{\int_{S_{j_2,i}} \pi(\underline{\theta}_{j_2,i} + u \underline{d}_{j_2,i} | \underline{\theta}_{j_1,i+1}, \underline{\theta}_{j_3,i}, \cdots, \underline{\theta}_{j_{k^*},i}) du}, \lambda \in S_{j_2,i}; \qquad (4.16)$$

- Set $\underline{\Theta}_{j_2,i+1} = \underline{\Theta}_{j_2,i} + \lambda_{j_2,i} \underline{d}_{j_2,i}$.

step 3. Set $i = i + 1$, and go to step 1.

It can thus be seen that the GH&R sampler is a generalization of the Gibbs sampler and the H&R sampler. If $k^* = 1$ and $j_1 = k$, the GH&R sampler is the Gibbs sampler; if $k^* = 2$, $j_1 = 0$, and $j_2 = k$, the GH&R sampler is the H&R sampler. The choice of $j_1, j_2, \cdots, j_{k^*}$ varies from problem to problem. In practice, this choice is often quite natural, e.g., in *Algorithm 1*, $k^* = 2$, $j_1 = 1$, and $j_2 = 10$.

5 Computation Results

In this section, we report the numerical solutions for the posterior means of $\underline{\beta}$ and σ^2 along with their associated numerical standard errors. Some comparisons between the Bayesian estimations and the ordinary least square estimations are also made. We also illustrate how the marginal posterior densities of the coefficients $\underline{\beta}$ can be obtained by using a new method proposed by Chen [1992] called the importance weighted marginal density estimation (IWMDE). In particular such densities are obtained for β_1, β_2, β_5, and β_{10}. In addition we derive estimators of the predictive mean and variance and estimate the predictive density. In conclusion robustness of the prior and the implementation of the GH&R sampler are also discussed.

5.1 Bayesian Estimation Versus Ordinary Least Square Estimation

The Bayesian posterior means $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ are given in Equations (3.7) and (3.8). Because of the constrained parameter space S, closed form solutions for posterior means and variances are not possible. Hence the true values of the desired estimators $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ are unavailable. To estimate them numerically, we used the Markov chain sample, $\{(\sigma_i^2, \underline{\beta}_i), 0 \leq i \leq n\}$, generated by *Algorithm 1* . The estimates of $\hat{\beta}_j$ and $\hat{\sigma}^2$ so obtained will be denoted by $\tilde{\beta}_j$ and $\tilde{\sigma}^2$ respectively and $s(\tilde{\beta}_j)$ and $s(\tilde{\sigma}^2)$ will denote the respective numerical standard errors. Since $\{(\sigma_i^2, \underline{\beta}_i), 0 \leq i \leq n\}$, which was obtained using the GH&R sampler, is a dependent sample, we used the batch statistics method suggested by Schmeiser, Avramidis and Hashem [1990], to obtain $s(\tilde{\beta}_j)$ and $s(\tilde{\sigma}^2)$.

In Table 5.1, we used 50 iterations to "warm up" the Markov chain, then used one single long run with 5 batches of size 10^5 to get $\tilde{\beta}_j$, $\tilde{\sigma}^2$, $s(\tilde{\beta}_j)$ and $s(\tilde{\sigma}^2)$. In Table 5.1, the ordinary least square estimators $\bar{\beta}_j$ and $\bar{\sigma}^2$ are reported along with the standard error $s(\bar{\beta}_j)$. $\tilde{\beta}_j$ (noninformative) and $\bar{\beta}_j$ are also displayed in Figure 1.

Table 5.1: Bayesian and Ordinary Least Square Estimations

	Bayesian Estimation (noninformative)		Ordinary Least Square Estimation		Bayesian Estimation (informative)	
parameter	$\tilde{\beta}_j$	$s(\tilde{\beta}_j)$	$\bar{\beta}_j$	$s(\bar{\beta}_j)$	$\tilde{\beta}_j$	$s(\tilde{\beta}_j)$
β_1	0.01315	0.00003	0.09478	0.03229	0.01315	0.00003
β_2	0.02487	0.00003	-0.0240	0.01929	0.02486	0.00004
β_3	0.17759	0.00006	0.19386	0.01211	0.17761	0.00006
β_4	0.30960	0.00049	0.29421	0.04208	0.30967	0.00045
β_5	0.55468	0.00113	0.53179	0.07831	0.55501	0.00203
β_6	0.78141	0.00073	0.83362	0.03922	0.78172	0.00080
β_7	0.81938	0.00097	0.74456	0.06559	0.82001	0.00103
β_8	0.98997	0.01006	1.23486	0.44459	0.99104	0.01102
β_9	1.13842	0.01261	0.52651	0.35547	1.13697	0.01327
β_{10}	1.84827	0.10287	1.00188	0.89903	1.78973	0.08232
	$\tilde{\sigma}^2$	$s(\tilde{\sigma}^2)$	$\bar{\sigma}^2$	$s(\bar{\sigma}^2)$	$\tilde{\sigma}^2$	$s(\tilde{\sigma}^2)$
σ^2	53909.1	79.3574	51170.45	·	53860.0	66.5492

From Table 5.1 it can be seen that the ordinary least square estimators of $\underline{\beta}$ are not in the constrained parameter space S while the Bayesian estimators $\tilde{\underline{\beta}}$ are. Further it is clear that for several coefficients, namely $\beta_3, \cdots, \beta_7$, the Bayesian estimators are very close to the ordinary least square estimators. Histograms of the residuals of the Bayesian and ordinary least square estimated regression were obtained but are not reported here. Both histograms were very similar. The respective mean square errors were very close as well being 226.2 and 233.6.

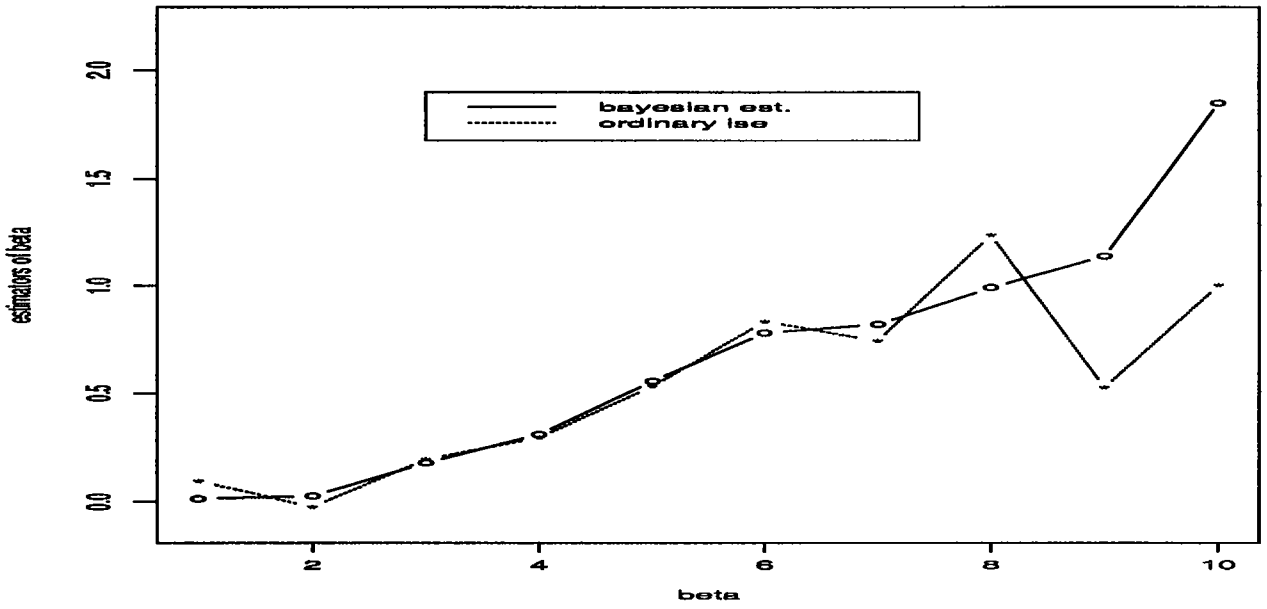

Figure 1: The Bayesian and ordinary least square estimations for $\underline{\beta}$.

5.2 Marginal Posterior Density Estimation

An interesting outcome of this study is to display the marginal posterior densities of β_j. Because of the constrained parameter space $S_j(\beta_j)$, closed forms for $\pi_j(\beta_j|data)$ given in Equation (3.11) are not available. Therefore the IWMDE method was applied to obtain the estimators of the marginal posterior densities for β_j. Based on the GH&R sample $\{(\sigma_i^2, \underline{\beta}_i), 0 \leq i \leq n\}$, the IWMDE of $\pi(\beta_j^*|data)$ is of form

$$\begin{aligned} & \hat{\pi}_j(\beta_j^*|data) \\ \stackrel{def}{=} & \frac{1}{n}\sum_{i=1}^{n} \frac{w(\beta_{j,i}|\beta_{l,i}, l \neq j)\pi(\beta_{1,i}, \cdots, \beta_{j-1,i}, \beta_j^*, \beta_{j+1,i}, \cdots, \beta_{10,i}, \sigma_i^2|data)}{\pi(\beta_{1,i}, \cdots, \beta_{j,i}, \beta_{j+1,i}, \cdots, \beta_{10,i}|data)}, \end{aligned} \tag{5. 1}$$

for any given point β_j^*, where $w(\beta_j|\beta_1, \cdots, \beta_{j-1}, \beta_{j+1}, \cdots, \beta_{10})$ is a conditional density playing the role of a weight function. A full description of this method along with technical details are given in Chen [1992]. In that paper it is shown that the choice of w greatly influences the rate of convergence and in particular that the rate is accelerated if w is chosen as close to the true conditional density as possible. Here for illustrative purposes we report the results for the marginal posterior densities for only β_1, β_2, β_5, and β_{10}.

In Figure 2 we used 50 GH&R iterations to "warm up" the Markov chain $\{(\sigma_i^2, \underline{\beta}_i), i \geq 0\}$, then used 50,000 GH&R iterations to get the marginal posterior densities for β_1, β_2, β_5, and β_{10}. We evaluated the values of the IWMDEs at 101 griddy points for β_1, β_5, β_{10} and 201 griddy points for β_2. We chose w for β_1 as $w(\beta_1|\beta_j, j \geq 2) = 1/\beta_2$, for $0 < \beta_1 < \beta_2$. Partial justification for this choice is the fact that $\hat{\beta}_1$ is roughly half of $\hat{\beta}_2$.

The support of the true conditional density of β_2 given $\beta_j, j \neq 2$ is

$$\{\beta_2 : \beta_1 < \beta_2 < \beta_3, \text{ for any given } 0 < \beta_1 < \beta_3\}. \tag{5.2}$$

The posterior mean $\hat{\beta}_3 = 0.1776$ is relatively far away from $\hat{\beta}_1$ and $\hat{\beta}_2$, and $\hat{\beta}_2$ is relatively close to $\hat{\beta}_1$. Therefore, the true conditional distribution might be skewed to β_1. We have to put more conditional mass near β_1. For this case w was chosen as follows:

$$w(\beta_2|\beta_1, \beta_3, \cdots, \beta_{10}) = \frac{10(\beta_3 - \beta_2)^9}{(\beta_3 - \beta_1)^{10}}, \text{ for } \beta_1 < \beta_2 < \beta_3. \tag{5.3}$$

For β_5, the support of the true conditional density is

$$\{\beta_5 : \beta_4 < \beta_5 < \beta_6, \text{ for any given } 0 < \beta_4 < \beta_6\}. \tag{5.4}$$

The posterior means of β_4, β_5 and β_6 are

$$\hat{\beta}_4 = 0.3096, \hat{\beta}_5 = 0.5547, \text{ and } \hat{\beta}_6 = 0.7814.$$

So, $\hat{\beta}_5$ is roughly in the middle of $\hat{\beta}_4$ and $\hat{\beta}_6$. Thus, w was chosen as

$$w(\beta_5|\beta_1, \cdots, \beta_4, \beta_6, \cdots, \beta_{10}) = \frac{1}{\beta_6 - \beta_4}, \text{ for } \beta_4 < \beta_5 < \beta_6. \tag{5.5}$$

For β_{10}, the support of the true conditional density is

$$\{\beta_{10} : \beta_9 < \beta_{10}, \text{ for any given } 0 < \beta_9\}. \tag{5.6}$$

The posterior means of β_9 and β_{10} are $\hat{\beta}_9 = 1.1384$ and $\hat{\beta}_{10} = 1.8483$. In this case the set given in Equation (5.8) is unbounded. We chose w as

$$w(\beta_{10}|\beta_1, \cdots, \beta_9) = \frac{1}{1.5034147} e^{-|\beta_{10} - \beta_9 - 0.7|}, \text{ for } \beta_9 < \beta_{10} < \infty. \tag{5.7}$$

Figure 2 shows that all the above choices for w gave the good convergence results of the IWMDEs.

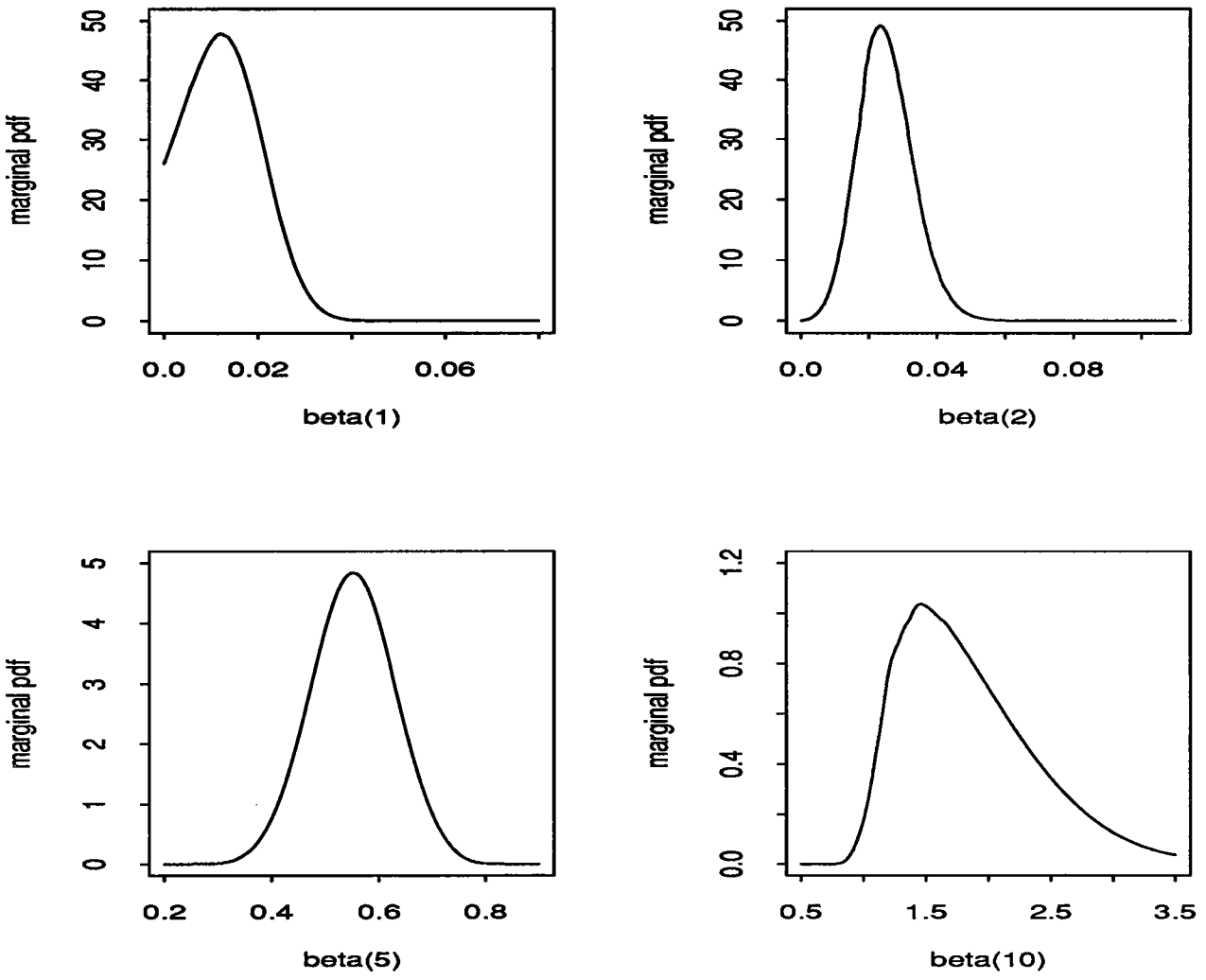

Figure 2: The IWM estimated marginal posterior density curves for β_1, β_2, β_5, and β_{10}.

5.3 Bayesian Prediction

The predictive mean and variance of y_ν at $\underline{x}_\nu$ and the predictive density of Y_ν are given in Equations (3.14), (3.15) and (3.13) respectively. As an example, we used

$$\underline{x}_\nu = (3654, 5373, 13204, 23859, 962, 580, 1787, 5443, 598, 371)' \tag{5.8}$$

as a point for prediction.

Let $\tilde{\mu}_{Y_\nu}$ and $\tilde{V}_{Y_\nu}$ denote the estimated predictive mean μ_{Y_ν} and variance V_{Y_ν} by simulation, and $s(\tilde{\mu}_{Y_\nu})$ and $s(\tilde{V}_{Y_\nu})$ denote the corresponding numerical standard errors. We used 50 iterations to

"warm up" the Markov chain, generated the GH&R Markov chain with $n = 10^4$ iterations and $m = 20$ *i.i.d.* macro replications and obtained

$$\tilde{\mu}_{Y_\nu} = 19296.9, s(\tilde{\mu}_{Y_\nu}) = 138.488 \quad (5.\ 9)$$

$$\tilde{V}_{Y_\nu} = 1597690, s(\tilde{V}_{Y_\nu}) = 77260.9. \quad (5.\ 10)$$

We used 10^5 GH&R iterations to evaluate the values of $f_\nu(y)$ in Equation (3.13) at 391 griddy points in Figure 3.

From Figure 3, the predictive density $f_\nu(y)$ is quite symmetric and unimodal. Therefore, we can use 3-σ limits confidence interval to predict y_ν at $\underline{x}_\nu$. This confidence interval is (15504.9, 23088.9).

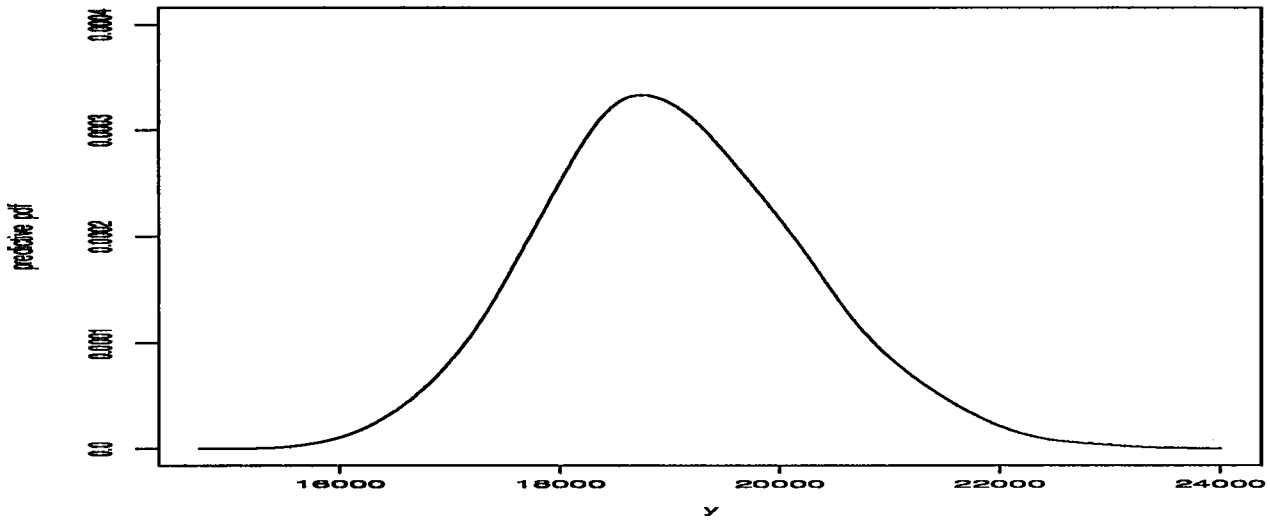

Figure 3: The predictive density of $f_\nu(y)$.

5.4 Robustness of the Prior

In Section 3, the flat noninformative prior on $\underline{\beta}$ was chosen, i.e.,

$$\pi_1(\underline{\beta}) = 1 \cdot I_S(\underline{\beta}),$$

where $S = \{(\beta_1, \cdots, \beta_{10})' : 0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_{10}, \underline{\beta} \in R^{10}\}$. In practice, upper bounds ($\underline{c}$) on the average number of cartons produced by trees at age j ($j = 1, 2, \cdots, 10$) may be known, at least approximately. For category one the following values were suggested:

$$c(1) = 1, c(2) = 1.22, c(3) = 1.44, c(4) = 1.67, c(5) = 1.89,$$
$$c(6) = 2.11, c(7) = 2.33, c(8) = 2.56, c(9) = 2.78, c(10) = 3, \quad (5.\ 11)$$

and $\beta_j \leq c(j)$, for $j = 1, 2, \cdots, 10$. A reasonable prior that accommodates this information is

$$\pi_1^*(\underline{\beta}) = 1 \cdot I_{S^*}(\underline{\beta}), \quad (5.12)$$

where

$$S^* = \{\underline{\beta} : \beta_j \leq c(j), j = 1, 2, \cdots, 10, \underline{\beta} \in S\}, \tag{5.13}$$

which we call the flat informative prior. The simulated estimates obtained with this prior are reported in Table 5.1. It can be seen that there is very little difference in the estimates from the two different priors.

5.5 Comments on the Implementation of the GH&R Sampler

Algorithm 1, a special case of the GH&R sampler, was programmed in single precision Fortran-77 using the IMSL library. Computation times for the Bayesian estimators reported in Table 5.1 were in the order of 1.6 hours on Sun Sparc-station 1 and were based on 500,050 GH&R iterations. It is clear from examining the standard errors thus obtained that considerably fewer iterations are necessary to obtain modest standard errors.

It should be noted that it was not possible to generate all λ_i directly from Equation (4.5). When the absolute values of both arguments $(R_2^i - \mu_i)/\sigma_i^N$ and $(R_2^i - \mu_i)/\sigma_i^N$ of Φ are suitably large, the computer returns 0 or 1 for the argument of Φ^{-1} in Equation (4.5). Therefore the true signed distance λ_i in this case could not be directly generated. For those extreme cases, we used the uniform rejection/acceptance sampling instead of Equation (4.5), which yields good results.

References

[1] Belisle, Claude J.P., Romeijn, H. Edwin and Smith, Robert L. (1990), "Hit-and-Run Algorithms for Generating Multivariate Distributions," Technical Report 90-18, The University of Michigan, Department of Industrial and Operations Engineering.

[2] Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Second Edition, Springer-Verlag, New York.

[3] Chen, Ming-Hui (1992), "Importance Weighted Marginal Bayesian Posterior Density Estimation," Working Paper, Purdue University, Department of Statistics.

[4] Chen, Ming-Hui and Schmeiser, B.W. (1992), "Performance of the Gibbs, Hit-and-Run, and Metropolis Samplers," Technical Report 92-18, Purdue University, Department of Statistics.

[5] Devroye, Luc (1986), *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.

[6] Gelfand, A. E. and Smith, A.F.M. (1990), "Sampling Based Approaches to Calculating Marginal Densities," *Journal of American Statistical Association, 85*, pp. 398-409.

[7] Gelfand, A. E., Smith, A.F.M. and Lee, T.M. (1990), "Bayesian Analysis of Constrained Parameter and Truncated Data Problems," Technical Report 90-01, University of Connecticut, Department of Statistics.

[8] Geman, S. and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6,* pp. 721-741.

[9] Geyer, C. J. and Thompson, E. A. (1992), "Constrained Monte Carlo Maximum Likelihood for Dependent Data," *J. R. Statist. Soc. B, 54, 3* (to appear).

[10] Geweke, J. (1991), "Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints," *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, pp. 571-578.

[11] Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrika, 57*, pp. 97-109.

[12] Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika, 57*, pp. 97 -109.

[13] Müller, P. (1991), "A Generic Approach to Posterior Integration and Gibbs Sampling," Technical Report 91-09, Purdue University, Department of Statistics.

[14] Schmeiser, B. W., Avramidis, T. N., and Hashem, S. (1990), "Overlapping Batch Statistics," In *Proceedings of the 1990 Winter Simulation Conference*, pp. 395-398.

[15] Schmeiser, Bruce W. and Chen, Ming-Hui (1991),"On Random-Direction Monte Carlo Sampling for Evaluating Multidimensional Integrals," Technical Report 91-39, Purdue University, Department of Statistics.

[16] Tierney, Luke (1991), "Markov Chains for Exploring Posterior Distributions," Technical Report No. 560, University of Minnesota, School of Statistics.