Efficient Gibbs sampler for Bayesian analysis of a sample selection model

Yasuhiro Omori

Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan

omori@e.u-tokyo.ac.jp

Abstract

We consider Bayesian estimation of a sample selection model and propose a highly efficient Gibbs sampler using the additional scale transformation step to speed up the convergence to the posterior distribution. Numerical examples are given to show the efficiency of our proposed sampler.

Key words: Bayesian analysis, Gibbs sampler, Sample selection model, Tobit model.

1 Introduction

A sample selection model or generalized Tobit (Type II Tobit) model has been very popular in the econometric analysis of the labour supply and wage function. It has been well-known as a generalization of the standard Tobit (Type I Tobit) model in econometrics since it was first introduced by Tobin (1958) to analyze the relationship between household income and household expenditures on a durable good where there are some households with zero expenditures (see e.g. Amemiya (1984) for a survey). Bayesian estimation method of a standard Tobit model using Monte Carlo method was proposed by Chib (1992). Chib (1992) developed Gibbs sampling procedure using the idea of data augmentation, which is widely used in the literature, and compared the efficacy of several Monte Carlo methods.

This article describes a highly efficient Gibbs sampler for a sample selection model and show that the convergence to the posterior distribution can be greatly accelerated by adding one more sampling step to the original Gibbs sampler. Numerical examples suggest that the original sampler suffers from inefficiencies and produces highly autocorrelated samples. The additional Gibbs move recovers efficiencies and reduces the sample autocorrelations dramatically. The rest of the paper is organised as follows. In Section 2, we introduce a sample selection model and describe a simple Gibbs sampler. Section 3 proposes an additional sampling step to accelerate the convergence to the posterior distribution. Numerical examples are given in Section 4.

2 Bayesian analysis of a sample selection model

In a sample selection model, the sample rule is determined by a latent random variable z_i^*, and we observe the response variable y_i when $z_i^* \geq 0$. The latent variable z_i^* is allowed to be correlated with the response variable y_i^*. When the correlation coefficient, ρ, between (z_i^*, y_i^*) is not equal to zero, a sample selection model is considered a Type I Tobit (a censored regression) model with a stochastic threshold model.

The sample selection model or a generalized Tobit (Type II Tobit) model for the i-th individual is given by

$$y_i = \begin{cases} y_i^*, & \text{if } z_i^* \geq 0, \\ \text{n.a.}, & \text{otherwise,} \end{cases} \qquad i = 1, 2, \ldots, n, \tag{1}$$

$$z_i^* = \boldsymbol{w}_i'\boldsymbol{\theta} + \xi_i, \tag{2}$$

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \eta_i, \tag{3}$$

$$(\xi_i, \eta_i)' \sim i.i.d.\ \mathcal{N}(\mathbf{0}, \Sigma),$$

where y_i is a dependent variable, y_i^*, z_i^* are latent dependent variables, $(\boldsymbol{w}_i, \boldsymbol{x}_i)$ are independent variable vectors, and $(\boldsymbol{\theta}, \boldsymbol{\beta})$ are corresponding coefficient vectors. The disturbance vector $(\xi_i, \eta_i)'$ follows a bivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ. The $(1, 1)$ element of Σ is set equal to 1 for the identification and we use the following parameterisation as in McCulloch *et al.* (2000) to implement Gibbs sampler:

$$\Sigma = \begin{pmatrix} 1 & \gamma \\ \gamma & \phi + \gamma^2 \end{pmatrix}.$$

This implies that the variance of the dependent variable, σ^2, is equal to $\phi + \gamma^2$ and that the correlation coefficient, ρ, between disturbances (ξ_i, η_i) in (2)–(3) is given by $\gamma/\sqrt{\phi + \gamma^2}$. To conduct Bayesian analysis, we assume that

$$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}_0, \Theta_0), \quad \boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, B_0), \quad \gamma \sim \mathcal{N}(\gamma_0, G_0), \quad \phi \sim \mathcal{IG}\left(\frac{n_0}{2}, \frac{S_0}{2}\right), \tag{4}$$

for prior distributions where $\mathcal{IG}$ denotes an inverse gamma distribution. Let $\boldsymbol{y}_c$ and $\boldsymbol{y}_o$ denote vectors of censored (latent) dependent variables and observed dependent variables, respectively. Then we obtain the joint posterior probability density of $(\boldsymbol{y}_c^*, \boldsymbol{z}^*, \boldsymbol{\theta}, \boldsymbol{\beta}, \gamma, \phi)$ given by

$$\begin{aligned}
&\pi(\boldsymbol{y}_c^*, \boldsymbol{z}^*, \boldsymbol{\theta}, \boldsymbol{\beta}, \gamma, \phi|\boldsymbol{y}_o) \\
&\quad \propto \quad \phi^{-\left(\frac{n_1}{2}+1\right)} \\
&\qquad \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(1+\phi^{-1}\gamma^2)(z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})^2 - 2\phi^{-1}\gamma(z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}) + \phi^{-1}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2\right\} \\
&\qquad \times \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'B_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\Theta_0^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \frac{(\gamma-\gamma_0)^2}{2G_0} - \frac{S_0}{2\phi}\right\}
\end{aligned}$$

where $\boldsymbol{z}^* = (z_1^*, z_2^*, \ldots, z_n^*)'$, and $n_1 = n_0 + n$. As shown in the Appendix A, the conditional posterior distributions of ϕ, γ, $\boldsymbol{\psi} = (\boldsymbol{\theta}', \boldsymbol{\beta}')'$ are

$$\begin{aligned}
\boldsymbol{\psi}|\gamma, \phi, \boldsymbol{z}^*, \boldsymbol{y}_c^*, \boldsymbol{y}_o &\sim \mathcal{N}(\boldsymbol{\psi}_1, \Psi_1), \\
\gamma|\boldsymbol{\psi}, \phi, \boldsymbol{z}^*, \boldsymbol{y}_c^*, \boldsymbol{y}_o &\sim \mathcal{N}(\gamma_1, G_1), \\
\phi|\boldsymbol{\psi}, \gamma, \boldsymbol{z}^*, \boldsymbol{y}_c^*, \boldsymbol{y}_o &\sim \mathcal{IG}\left(\frac{n_1}{2}, \frac{S_1}{2}\right),
\end{aligned}$$

where $\boldsymbol{z}^* = (z_1^*, z_2^*, \ldots, z_n^*)'$, $n_1 = n_0 + n$, $G_1^{-1} = G_0^{-1} + \phi^{-1}\sum_{i=1}^{n}(z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})^2$,

$\gamma_1 = G_1 \left\{ G_0^{-1}\gamma_0 + \phi^{-1}\sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})(y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta}) \right\}$, $S_1 = S_0 + \gamma^2 \sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})^2 - 2\gamma \sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})(y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta}) + \sum_{i=1}^n (y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})^2$, and

$$
\begin{aligned}
\Psi_1 &= \left(\Psi_0^{-1} + \sum_{i=1}^n \tilde{X}_i' \Sigma^{-1} \tilde{X}_i \right)^{-1}, \quad \boldsymbol{\psi}_1 = \Psi_1 \left(\Psi_0^{-1}\boldsymbol{\psi}_0 + \sum_{i=1}^n \tilde{X}_i' \Sigma^{-1} \tilde{\boldsymbol{y}}_i^* \right) \\
\tilde{\boldsymbol{y}}_i^* &= \begin{pmatrix} z_i^* \\ y_i^* \end{pmatrix}, \quad \tilde{X}_i = \begin{pmatrix} \boldsymbol{w}_i' & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{x}_i' \end{pmatrix}, \quad \boldsymbol{\psi}_0 = \begin{pmatrix} \theta_0 \\ \boldsymbol{\beta}_0 \end{pmatrix}, \quad \Psi_0 = \begin{pmatrix} \Theta_0 & O \\ O & B_0 \end{pmatrix}.
\end{aligned}
$$

Using these conditional posterior distributions, we implement the Gibbs sampler as follows:

1. Initialise ϕ, γ and $\boldsymbol{\psi}$.
2. Sample $(\boldsymbol{y}_c^*, \boldsymbol{z}^*)|\phi, \gamma, \boldsymbol{\psi}, \boldsymbol{y}_o$.
   (a) For censored observations, we generate $y_i^*|\boldsymbol{\psi}, \phi, \gamma \sim \mathcal{N}(\boldsymbol{x}_i'\boldsymbol{\beta}, \phi + \gamma^2)$ and $z_i^*|y_i^*, \boldsymbol{\psi}, \phi, \gamma \sim \mathcal{TN}_{(-\infty,0)}(\mu_z, \sigma_z^2)$ where $\mu_z = \boldsymbol{w}_i'\boldsymbol{\theta} + \gamma(y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})/(\phi + \gamma^2)$, $\sigma_z^2 = 1 - \gamma^2/(\phi + \gamma^2)$ and $\mathcal{TN}_{(a,b)}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 truncated on the interval (a, b).
   (b) For uncensored observations, we generate $z_i^*|y_i, \boldsymbol{\psi}, \phi, \gamma \sim \mathcal{TN}_{[0,\infty)}(\mu_z, \sigma_z^2)$.
3. Sample $\phi|\boldsymbol{\psi}, \gamma, \boldsymbol{z}^*, \boldsymbol{y}_c^*, \boldsymbol{y}_o \sim \mathcal{IG}(n_1/2, S_1/2)$,
4. Sample $\gamma|\phi, \boldsymbol{\psi}, \boldsymbol{z}^*, \boldsymbol{y}_c^*, \boldsymbol{y}_o \sim \mathcal{N}(\gamma_1, G_1)$.
5. Sample $\boldsymbol{\psi}|\phi, \gamma, \boldsymbol{z}^*, \boldsymbol{y}_c^*, \boldsymbol{y}_o \sim \mathcal{N}(\boldsymbol{\psi}_1, \Psi_1)$.
6. Go to 2.

3 Acceleration of the Gibbs sampler

As we shall see in the illustrative examples, samples from the simple Gibbs sampler in Section 2 are highly autocorrelated and large number of iterations would be required to conduct the appropriate statistical inferences for the parameters. To speed up the convergence, we consider the additional step to the Gibbs sampler which transforms some parameters without changing the stationary distribution of the Markov chain.

Consider the scale group $\Gamma = \{g > 0 : g(\boldsymbol{\varphi}) = (g\sqrt{\phi}, g\gamma, g\boldsymbol{\theta}, g\boldsymbol{z}^*)\}$ where $\boldsymbol{\varphi} = (\sqrt{\phi}, \gamma, \boldsymbol{\theta}, \boldsymbol{z}^*)$, the unimodular left-Harr measure is $L(dg) = g^{-1}dg$ and the corresponding Jacobian is $J_g = g^{2+J+n}$ (where J is a dimension of the vector $\boldsymbol{\theta}$). The conditional probability density of g which preserves a stationary distribution of the chain can be obtained as follows using Theorem 1 of Liu and Sabatti (2000).

$$
\begin{aligned}
&\pi(g|\boldsymbol{\varphi}, \boldsymbol{\beta}, \boldsymbol{y}_c^*, \boldsymbol{y}_o) \\
&\quad \propto \ \pi(g\sqrt{\phi}, g\gamma, g\boldsymbol{\theta}, g\boldsymbol{z}^*|\boldsymbol{\beta}, \boldsymbol{y}_c^*, \boldsymbol{y}_o) \times |J_g| \times L(dg) \\
&\quad \propto \ g^{\nu_1 - 1} \times \exp\left\{-\frac{1}{2}(a^2 g^{-2} + b^2 g^2)\right\} \times \exp\left\{g(\boldsymbol{\theta}'\Theta_0^{-1}\boldsymbol{\theta}_0 + \gamma\gamma_0 G_0^{-1})\right\}, \qquad (5)
\end{aligned}
$$

where

$$
\begin{aligned}
\nu_1 &= J - n_0 + 1, && (6)\\
a^2 &= \phi^{-1}\left\{S_0 + \sum_{i=1}^{n}(y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})^2\right\}, && (7)\\
b^2 &= (1 + \phi^{-1}\gamma^2)\sum_{i=1}^{n}(z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})^2 + \boldsymbol{\theta}'\Theta_0^{-1}\boldsymbol{\theta} + \gamma^2 G_0^{-1}. && (8)
\end{aligned}
$$

When $\gamma = 0$ and $\boldsymbol{\theta}_0 = \mathbf{0}$, the g^2 follows generalized inverse Gaussian distribution $\mathcal{GIG}(\nu_1/2, a, b)$ $(a, b \geq 0)$ where $\mathcal{GIG}(\nu, a, b)$ denotes the probability density function given by

$$
f(x|\nu, a, b) = \frac{(b/a)^{\nu}}{2K_{\nu}(ab)} x^{\nu-1} \exp\left\{-\frac{1}{2}(a^2 x^{-1} + b^2 x)\right\}, \qquad x > 0, \quad a, b \geq 0, \quad -\infty < \nu < \infty,
$$

and K_ν is a modified Bessel function of the third kind (see e.g. Barndorff-Nielsen and Shephard (2001)). To generate a random sample from $\mathcal{GIG}(\nu, a, b)$, see e.g. Dagpunar (1989), Doornik (2002) and Hörmann *et al.* (2004).

When $\gamma \neq 0$ or $\boldsymbol{\theta}_0 \neq \mathbf{0}$, the conditional posterior distribution of g is not a well-known probability distribution and we need to conduct the Metropolis-Hastings (MH) algorithm to sample g. Given the current point g, we generate a candidate $g'^2 \sim \mathcal{GIG}(\nu_1/2, a, b)$ and accept it with probability $\min\left[\exp\left\{1, (g'-g)(\boldsymbol{\theta}'\Theta_0^{-1}\boldsymbol{\theta}_0 + \gamma\gamma_0 G_0^{-1})\right\}\right]$. To obtain a random sample from this conditional distribution, we usually need to repeat the MH algorithm many times until the distribution of the sample converges to its stationary distribution. However, in sampling g from (5), we can show that we only need to implement the MH algorithm once using the initial value $g = 1$.

Theorem 3.1. *Suppose that $\gamma \neq 0$ or $\boldsymbol{\theta}_0 \neq \mathbf{0}$ in (5). To sample from the conditional distribution of g, it suffices to generate a candidate $g'^2 \sim \mathcal{GIG}(\nu_1/2, a, b)$ where ν_1, a, b are given in (6)–(8) and accept it with probability*

$$\min\left[\exp\left\{1, (g'-1)(\boldsymbol{\theta}'\Theta_0^{-1}\boldsymbol{\theta}_0 + \gamma\gamma_0 G_0^{-1})\right\}\right].$$

If it is rejected, set $g = 1$.

Proof: (See Appendix B).

Thus, to accelerate the convergence of the Gibbs sampler described in Section 2, we replace Step 6 by

6'. (a) Generate $g^2 \sim \mathcal{GIG}(\nu_1/2, a, b)$.

   (b) Accept g with probability

$$\min\left[1, \exp\left\{(g-1)(\boldsymbol{\theta}'\Theta_0^{-1}\boldsymbol{\theta}_0 + \gamma\gamma_0 G_0^{-1})\right\}\right].$$

   If rejected, set $g = 1$.

   (c) Let $g\sqrt{\phi} \to \sqrt{\phi}$, $g\gamma \to \gamma$, $g\boldsymbol{\theta} \to \boldsymbol{\theta}$ and $g\boldsymbol{z}^* \to \boldsymbol{z}^*$.

7'. Go to 2.

Note that the Metropolis-Hastings algorithm reduces to Gibbs sampler when $\gamma_0 = 0$ and $\boldsymbol{\theta}_0 = \mathbf{0}$. When the absolute value of ρ is close to one (*i.e.*, ϕ is very small), the speed of the random sample generation from the generalized inverse Gaussian distribution may become very slow, and it is recommended to skip this additional transformation step (say when $\sqrt{ab} > 150$).

4 Numerical example

We illustrate our proposed procedure using the simulated data from the sample selection model. We set

$$\boldsymbol{\theta} = (1, 5, 10)', \quad \boldsymbol{\beta} = (2, 1, 1)', \quad \sigma^2 = 1.0, \quad \rho = 0.9,$$

and all covariates are generated using a standard normal distribution independently. The total number of generated observations was 1000, and 46.5 percent of them were censored. The sampling results are based on less informative proper prior distributions given by

$$\begin{aligned} \boldsymbol{\theta} &\sim \mathcal{N}(0, 10I_3), \quad \boldsymbol{\beta} \sim \mathcal{N}(0, 10I_3), \\ \gamma &\sim \mathcal{N}(0, 10), \quad \phi \sim \mathcal{IG}\,(0.001, 0.001)\,. \end{aligned}$$

The initial 5,000 variates are discarded as so-called burn-in period and the subsequent $50,000$ values are recorded to conduct an inference. Figure 1 & 2 show the sample paths and the sample autocorrelations functions for the original Gibbs sampler described in Section 2. It is clear that the sample paths show very slow convergence to the posterior distribution for $\theta_i' s$, parameters of selection equation (2), and their autocorrelations do no decay even at 7,500 lags.

The summary statistics are given in Table 1. The inefficiency factors in Table 1 are calculated to measure how well the chain mixes. The inefficiency factor is defined as $1 + 2\sum_{s=1}^{\infty} \rho_s$ where ρ_s is the sample autocorrelation at lag s calculated from the sampled values (see e.g. Chib (2001)). It is the ratio of the numerical variance of the sample posterior mean to the variance of the sample mean from the hypothetical uncorrelated draws.

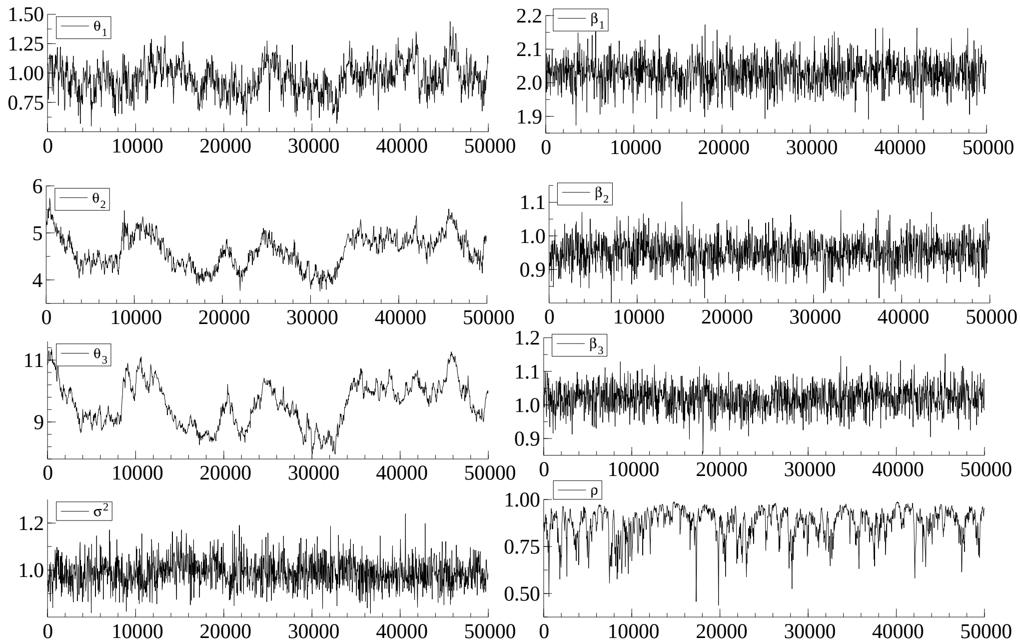

Figure 1: Sample paths from original Gibbs sampler.

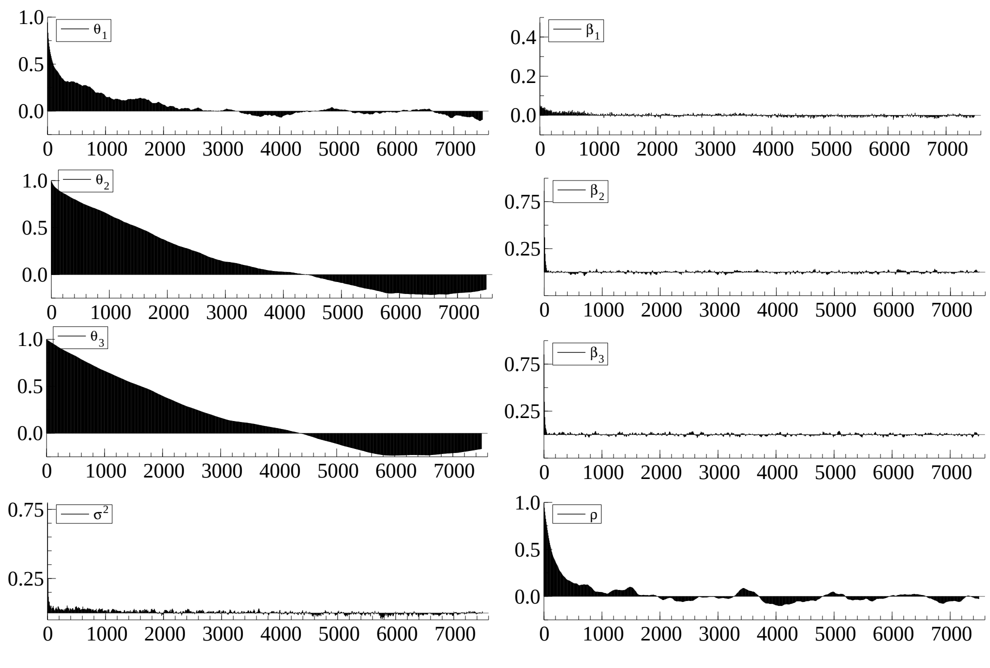

Figure 2: Sample autocorrelations from original Gibbs Sampler.

	True	Mean	Stdev	95% Interval	Inefficiency
θ_1	1.0	0.957	0.133	(0.695, 1.218)	744.3
θ_2	5.0	4.620	0.354	(3.991, 5.297)	2172.5
θ_3	10.0	9.587	0.717	(8.326, 10.968)	2310.1
β_1	2.0	2.029	0.043	(1.944, 2.114)	32.8
β_2	1.0	0.952	0.037	(0.868, 1.032)	15.6
β_3	1.0	1.019	0.039	(0.943, 1.095)	33.1
σ^2	1.0	0.988	0.061	(0.876, 1.115)	84.6
ρ	0.9	0.873	0.081	(0.668, 0.973)	471.3

Table 1: Posterior means, standard deviations, 95% credible intervals and inefficiency factors obtained from original Gibbs sampler ($\rho = 0.9$).

The inefficiency factors for θ_i's are quite large in the range of $700 \sim 2300$, indicating the poor mixing property of the simple Gibbs sampler. It implies that we need to sample from the Gibbs sampler about 2300 as many times as the hypothetical uncorrelated sampler to obtain the same variance of the posterior sample mean. We note that the corresponding factors for σ^2 and ρ are also large, while those for $\beta_i's$, parameters of the regression equation (3), are relatively small.

For the accelerated Gibbs sampler described in Section 3, Figure 3 & 4 show the sample paths and sample autocorrelation functions. The sample paths seem to mix well and autocorrelations die out very quickly. We skipped only 6 percent of the acceleration steps due to slow random generations, and succeeded in improving the mixing property of obtained samples. The summary statistics are shown in Table 2. The inefficiency factors for $\theta_i's$ are drastically decreased to $300 \sim 450$.

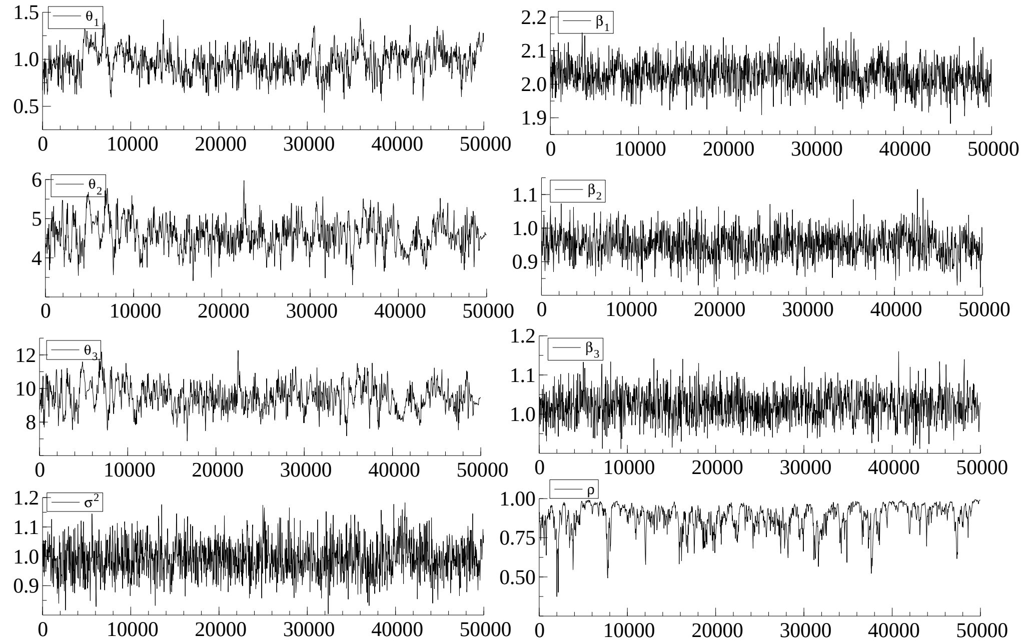

Figure 3: Sample paths from accelerated Gibbs sampler.

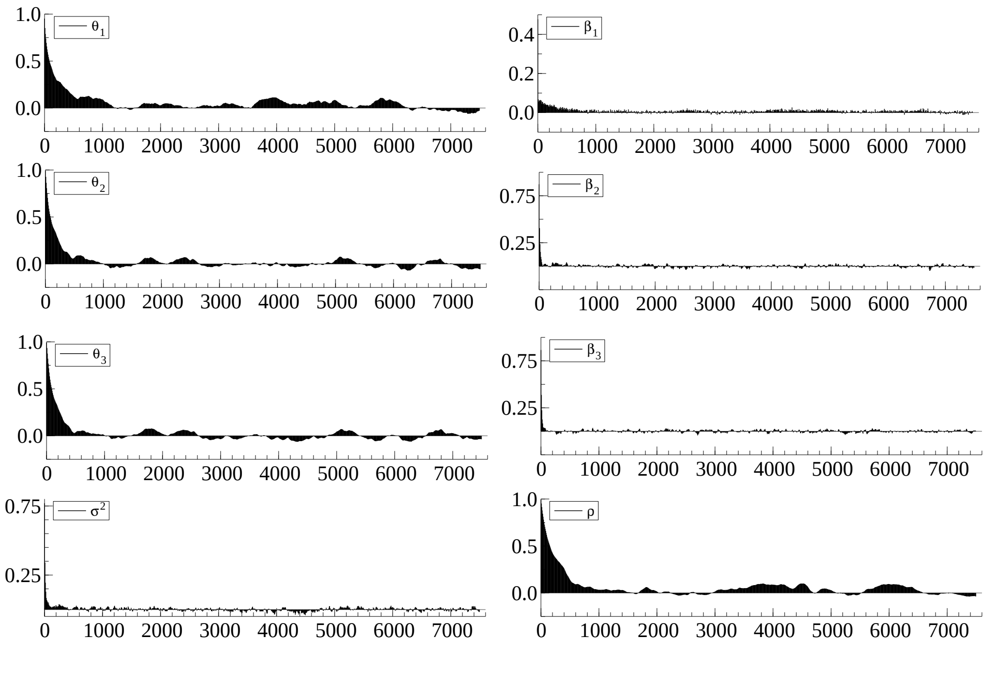

Figure 4: Sample autocorrelations from accelerated Gibbs sampler.

	True	Mean	Stdev	95% Interval	Inefficiency
θ_1	1.0	0.971	0.148	(0.679, 1.262)	454.5
θ_2	5.0	4.591	0.390	(3.851, 5.401)	324.0
θ_3	10.0	9.506	0.802	(8.024, 11.134)	318.0
β_1	2.0	2.027	0.043	(1.942, 2.112)	51.8
β_2	1.0	0.951	0.041	(0.870, 1.033)	38.2
β_3	1.0	1.022	0.039	(0.947, 1.098)	23.8
σ^2	1.0	0.989	0.061	(0.879, 1.115)	45.0
ρ	0.9	0.882	0.086	(0.661, 0.983)	501.5

Table 2: Posterior means, standard deviations, 95% credible intervals and inefficiency factors obtained from accelerated Gibbs sampler ($\rho = 0.9$).

To investigate the influence of the correlation coefficient, ρ, on the sampling efficiencies, we repeated the experiments using $\rho = 0.5$ and 0.98. First, for $\rho = 0.5$, Table 3 shows summary statistics for the accelerated Gibbs sampler. The inefficiency factors are still large for $\theta_i^\prime s$, but relatively smaller compared with those values in Table 1 & 2 with $\rho = 0.9$. The obtained samples seem to be less autocorrelated when the correlation coefficient $\rho = 0.5$. We did not need to skip any acceleration step due to slow random generations, and accomplished great improvements in decreasing the inefficiency factors.

For $\rho = 0.98$, Table 4 shows very high inefficiency factors for $\theta_i^\prime s$, and the obtained samples are highly autocorrelated. Since ρ is very close to one, we have very small values of ϕ. This resulted in skipping 71% of acceleration steps due to the slow random number generations from the $\mathcal{GIG}(\nu_1/2, a, b)$ distribution, and we were not as successful as in the case $\rho = 0.5$ or $\rho = 0.9$.

	True	Mean	Stdev	95% Interval	Inefficiency	[Inefficiency]
θ_1	1.0	0.902	0.147	(0.627, 1.197)	69.8	[690.7]
θ_2	5.0	4.419	0.452	(3.569, 5.348)	89.2	[1859.7]
θ_3	10.0	9.243	0.934	(7.492, 11.183)	99.0	[1955.8]
β_1	2.0	1.991	0.044	(1.904, 2.078)	11.2	[8.5]
β_2	1.0	0.968	0.044	(0.881, 1.054)	5.4	[3.1]
β_3	1.0	1.031	0.040	(0.952, 1.111)	3.4	[5.3]
σ^2	1.0	1.001	0.062	(0.887, 1.129)	2.7	[5.8]
ρ	0.5	0.501	0.140	(0.201, 0.744)	146.5	[82.8]

Table 3: Accelerated Gibbs sampler when $\rho = 0.5$. Inefficiency factors using the original Gibbs sampler are given in brackets.

	True	Mean	Stdev	95% Interval	Inefficiency	[Inefficiency]
θ_1	1.0	0.997	0.102	(0.789, 1.194)	770.9	[1689.3]
θ_2	5.0	4.669	0.253	(4.239, 5.205)	1220.1	[3057.7]
θ_3	10.0	9.468	0.504	(8.661, 10.571)	1281.0	[3102.7]
β_1	2.0	2.038	0.042	(1.956, 2.121)	17.9	[42.6]
β_2	1.0	0.951	0.039	(0.871, 1.028)	102.3	[171.4]
β_3	1.0	1.005	0.039	(0.930, 1.081)	246.7	[394.2]
σ^2	1.0	0.992	0.059	(0.882, 1.112)	307.2	[640.5]
ρ	0.98	0.979	0.029	(0.903, 0.999)	758.6	[1209.5]

Table 4: Accelerated Gibbs sampler ($\rho = 0.98$). Inefficiency factors using the original Gibbs sampler are given in brackets.

Since the inefficiency factors for θ_1 are found to be smaller than those for θ_2 and θ_3, the above experiments are repeated using different set of parameters for θ. We found that the inefficiency of the original Gibbs sampler may become moderate when the values of θ_i are sufficiently small.

5 Conclusion

The efficient Markov chain Monte Carlo implementations are described for Bayesian analysis of a sample selection model. The proposed estimation method is illustrated using numerical examples and is found to be highly efficient.

Acknowledgement. This work is supported by the Grants-in-Aid for Scientific Research 18330039 from the Japanese Ministry of Education, Science, Sports, Culture and Technology. The author thanks Nobuhiro Yoneda and Hideo Kozumi for their helpful comments and discussions.

Appendix A

Conditional posterior probability density of ϕ.

$$\begin{aligned}
&\pi(\phi|\boldsymbol{y}_c^*, \boldsymbol{z}^*, \boldsymbol{\theta}, \boldsymbol{\beta}, \gamma, \boldsymbol{y}_o) \\
&\quad\propto \quad \phi^{-\left(\frac{n_1}{2}+1\right)} \exp\left[-\frac{1}{2\phi}\left\{\gamma^2 \sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})^2 - 2\gamma \sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})(y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta}) + \sum_{i=1}^n (y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + S_0\right\}\right] \\
&\quad\propto \quad \phi^{-\left(\frac{n_1}{2}+1\right)} \exp\left\{-\frac{S_1}{2\phi}\right\}
\end{aligned}$$

where $S_1 = S_0 + \gamma^2 \sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})^2 - 2\gamma \sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})(y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta}) + \sum_{i=1}^n (y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})^2$.

Conditional posterior probability density of γ.

$$\begin{aligned}
&\pi(\gamma|\boldsymbol{y}_c^*, \boldsymbol{z}^*, \boldsymbol{\theta}, \boldsymbol{\beta}, \phi, \boldsymbol{y}_o) \\
&\quad\propto \quad \exp\left[-\frac{1}{2\phi}\left\{\gamma^2 \sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})^2 - 2\gamma \sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})(y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})\right\} - \frac{(\gamma - \gamma_0)^2}{2G_0}\right] \\
&\quad\propto \quad \exp\left\{-\frac{(\gamma - \gamma_1)^2}{2G_1}\right\}
\end{aligned}$$

where $G_1^{-1} = G_0^{-1} + \phi^{-1} \sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})^2$ and $\gamma_1 = G_1 \left\{G_0^{-1}\gamma_0 + \phi^{-1} \sum_{i=1}^n (z_i^* - \boldsymbol{w}_i'\boldsymbol{\theta})(y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})\right\}$.

Conditional posterior probability density of $\boldsymbol{\psi} = (\boldsymbol{\theta}', \boldsymbol{\beta}')'$.

$$\begin{aligned}
&\pi(\boldsymbol{\psi}|\boldsymbol{y}_c^*, \boldsymbol{z}^*, \gamma, \phi, \boldsymbol{y}_o) \\
&\quad\propto \quad \exp\left\{-\frac{1}{2}\sum_{i=1}^n (\tilde{\boldsymbol{y}}_i^* - \tilde{X}_i\boldsymbol{\psi})'\Sigma^{-1}(\tilde{\boldsymbol{y}}_i^* - \tilde{X}_i\boldsymbol{\psi}) - \frac{1}{2}(\boldsymbol{\psi} - \boldsymbol{\psi}_0)'\Psi_0^{-1}(\boldsymbol{\psi} - \boldsymbol{\psi}_0)\right\} \\
&\quad\propto \quad \exp\left\{-\frac{1}{2}(\boldsymbol{\psi} - \boldsymbol{\psi}_1)'\Psi_1^{-1}(\boldsymbol{\psi} - \boldsymbol{\psi}_1)\right\}
\end{aligned}$$

where

$$\begin{aligned}
\Psi_1 &= \left(\Psi_0^{-1} + \sum_{i=1}^n \tilde{X}_i'\Sigma^{-1}\tilde{X}_i\right)^{-1}, \quad \boldsymbol{\psi}_1 = \Psi_1\left(\Psi_0^{-1}\boldsymbol{\psi}_0 + \sum_{i=1}^n \tilde{X}_i'\Sigma^{-1}\tilde{\boldsymbol{y}}_i^*\right) \\
\tilde{\boldsymbol{y}}_i^* &= \begin{pmatrix} z_i^* \\ y_i^* \end{pmatrix}, \quad \tilde{X}_i = \begin{pmatrix} \boldsymbol{w}_i' & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{x}_i' \end{pmatrix}, \quad \boldsymbol{\psi}_0 = \begin{pmatrix} \theta_0 \\ \boldsymbol{\beta}_0 \end{pmatrix}, \quad \Psi_0 = \begin{pmatrix} \Theta_0 & O \\ O & B_0 \end{pmatrix}.
\end{aligned}$$

Appendix B

Proof of Theorem 3.1. Given the current sample g of the conditional distribution, let $\alpha(g, g')$ denote the acceptance probability of the candidate g' where $g'^2 \sim \mathcal{GIG}(\nu_1/2, a, b)$,

$$\alpha(g, g') \;=\; \min\left[\exp\left\{(g' - g)(\boldsymbol{\theta}'\Theta_0^{-1}\boldsymbol{\theta}_0 + \gamma\gamma_0 G_0^{-1})\right\}, 1\right].$$

Then the Markov transition function of g is $T_x(g, g')L(dg')$ given parameters $x = (\sqrt{\phi}, \gamma, \boldsymbol{\theta}, \boldsymbol{z}^*)$ where

$$T_x(g, g') \;=\; \frac{(b/a)^{\frac{\nu_1}{2}}}{K_{\nu_1/2}(ab)} g'^{\nu_1} \exp\left\{-\frac{1}{2}\left(a^2 g'^{-2} + b^2 g'^2\right)\right\} \alpha(g, g').$$

Since

$$\begin{aligned} T_{g_0^{-1}(x)}(gg_0, g'g_0) \;&=\; \frac{\{(b/g_0)/(ag_0)\}^{\frac{\nu_1}{2}}}{K_{\nu_1/2}((ag_0)(b/g_0))}(g'g_0)^{\nu_1} \\ &\quad\times \exp\left\{-\frac{1}{2}\left((ag_0)^2 g'^{-2} g_0^{-2} + (b/g_0)^2 g'^2 g_0^2\right)\right\} \alpha_{g_0^{-1}(x)}(gg_0, g'g_0) \\ &=\; T_x(g, g'), \end{aligned}$$

the result follows by Theorem 2 in Liu and Sabatti (2000).

$\square$

References

Amemiya, T. (1984), "Tobit models, a survey," *Journal of Econometrics,* 24, 3-61.

Barndorff-Nielsen, O. E. and Shephard, N. (2001), "Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion) ," *Journal of the Royal Statistical Society, Series B,* 63, 167-241.

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

Chib, S. (2001), Markov chain Monte Carlo methods: computation and inference. in J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, **5**, 3569-3649. North Holland, Amsterdam.

Dagpunar, J. S. (1989), "An easily implemented generalized inverse Gaussian generator," *Communications in Statistics-Simulations,* 18, 703-710.

Doornik, J.A. (2002), *Object-Oriented Matrix Programming Using Ox, 3rd ed.* London: Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik.

Hörmann, W., Leydold, J. and Derflinger G. (2004), *Automatic Nonuniform Random Variate Generation.* Springer, Berlin.

Liu, J. S. and Sabatti, C. (2000), "Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation," *Biometrika,* 87, 353-369.

McCulloch, R., Polson, N. G. and Rossi, P. (2000), "A Bayesian analysis of the multinomial probit model with fully identified parameters," *Journal of Econometrics,* 99, 173-193.

Tobin, J. (1958), "Estimation of relationships for limited dependent variables, *Econometrica,* 26, 24-36.