

Comparison of the Sampling Efficiency in Spatial Autoregressive Model

Yoshihiro Ohtsuka¹, Kazuhiko Kakamu²

¹Department of Economics, University of Nagasaki, Nagasaki, Japan

²Faculty of Law, Politics and Economics, Chiba University, Chiba, Japan

Email: ohtsuka@sun.ac.jp, kakamu@le.chiba-u.ac.jp

Received 29 December 2014; accepted 18 January 2015; published 22 January 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

A random walk Metropolis-Hastings algorithm has been widely used in sampling the parameter of spatial interaction in spatial autoregressive model from a Bayesian point of view. In addition, as an alternative approach, the griddy Gibbs sampler is proposed by [1] and utilized by [2]. This paper proposes an acceptance-rejection Metropolis-Hastings algorithm as a third approach, and compares these three algorithms through Monte Carlo experiments. The experimental results show that the griddy Gibbs sampler is the most efficient algorithm among the algorithms whether the number of observations is small or not in terms of the computation time and the inefficiency factors. Moreover, it seems to work well when the size of grid is 100.

Keywords

Acceptance-Rejection Metropolis-Hastings Algorithm, Griddy Gibbs Sampler, Markov Chain Monte Carlo (MCMC), Random Walk Metropolis-Hastings Algorithm, Spatial Autoregressive Model

1. Introduction

Spatial models have been widely used in various research fields such as physical, environmental, biological science and so on. Recently, a lot of researches are also emerging in econometrics (e.g., [3] [4] and so on), and [5] gave an excellent survey from the viewpoint of econometrics. When we focus on the estimation methods, properties of several estimation methods are studied. For example, the efficient maximum likelihood (ML) method was proposed by [6], and [7] first formally proved that the quasi maximum likelihood estimator had the usual asymptotic properties, including \sqrt{n} -consistency, asymptotic normality, and asymptotic efficiency. A class of moment estimators was examined by [8] and [9]. The Bayesian approach was first considered by [10] and [11] proposed a Markov chain Monte Carlo (hereafter MCMC) method to estimate the parameters of the

model. We have to mention that in economic analysis typically the sample size is small, for instance, areal data such as state-level data is widely used. The maximum likelihood methods depend on their asymptotic properties while the Bayesian method does not, because the latter evaluates the posterior distributions of the parameters conditioned on the data. Therefore, it is reasonable to examine the properties of Bayesian estimators (see [12]).

Although there are a lot of works using spatial models in a Bayesian framework, previous literature has rarely examined sampling methods for the parameter of spatial correlation. [13] proposed a random walk Metropolis-Hastings (hereafter RMH) algorithm. This method is widely used (e.g., [11] [12] [14] and so on). On the other hand, [2] applied a griddy Gibbs sampler (hereafter GGS) proposed by [1] and showed the GGS got an advantage over the RMH method from a simulated data and estimated the regional electricity demand in Japan. However, [2] has examined only one case. In this paper, we compare the properties of the GGS in the case that the number of observation is small (or large) through the Monte Carlo experiments. Desirable properties for sampling methods in the Bayesian inference are efficiency and well mixing, which yield fast convergence. In addition to these properties, computational requirements and model flexibility are important for applied econometrics. Therefore, the purpose of this paper is to investigate the properties of some sampling algorithms given several parameters of a model.

In this paper, we examine the efficiency of the existing Markov chain Monte Carlo methods for the spatial autoregressive (hereafter SAR) model which is the simplest and most commonly used model in the spatial models. Moreover, we propose an acceptance-rejection Metropolis-Hastings (hereafter ARMH) algorithm as an alternative MH algorithm, which is proposed by [15] because it is well known that the RMH is inefficient. This algorithm is widely used for the acceleration of MCMC convergence, for example, in the time series models (see [16]-[18] and so on). The advantage of this method is that the computational requirement is very small since it is irrelevant to the shape of the full conditional density. Therefore, we apply the algorithm to the SAR model.

We illustrate the properties of these algorithms using simulated data set given the three number of observations and the seven values of spatial correlation. From the results, we find that the GGS is the most efficient method whether the number of observations is small or not in terms of both the computation time and the inefficiency factors. Furthermore, we show that it is efficient when the number of grid in the GGS sampler is one hundred. These results give a benchmark of sampling the spatial correlation parameter of the models.

The rest of this paper is organized as follows. Section 2 summarizes the SAR model. Section 3 discusses the computational strategies of the MCMC methods, and reviews three sampling methods for spatial correlation parameter. Section 4 gives the Monte Carlo experiments using simulated data set and discusses the results. Finally, we summarize the results and provide concluding remarks.

2. Spatial Autoregressive (SAR) Model

Spatial autoregressive model explains the spatial spillover using a weight matrix (see [19]). There are numerous approaches to construct the weight matrix, which plays an important role in the model. For example, those are a first order contiguity matrix, inverse distance one and so on. Among the approaches, [20] recommended the first order contiguity dummies, because they showed that the first order contiguity weight matrix identifies the true model more frequently than the other matrices through the Monte Carlo simulations. Thus, we also utilize the first order contiguity dummies as the weight matrix.

Let C be an $n \times n$ matrix of contiguity dummies, with $c_{ij} = 1$ if areas i and j are adjacent and $c_{ij} = 0$ otherwise (with $c_{ii} = 0$). We standardized the weight matrix as follows

$$w_{ij} = \frac{c_{ij}}{\sum_{j=1}^n c_{ij}}$$

and we define $W = \{w_{ij}\}$, where w_{ij} denotes the spatial weight on the j -th unit with respect to the i -th unit.

Note that we have $\sum_{j=1}^n w_{ij} = 1$ for all i .

Next, let y_i and \mathbf{x}_i be a dependent variable and a $1 \times k$ vector of covariates on the i th unit for $i = 1, \dots, n$, respectively. Then, the SAR model conditioned on the parameters ρ , β , σ^2 is written as follows:

$$y_i = \rho \sum_{j=1}^n w_{ij} y_j + \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \quad |\rho| < 1 \quad (1)$$

where ρ and σ^2 indicates the spatial correlation, and the variance of the disturbance term, respectively. As is shown in [21], we know that $\lambda_{\min}^{-1} = -1$ and $\lambda_{\max}^{-1} = 1$, where λ_{\min} and λ_{\max} denote the minimum and maximum eigenvalue of \mathbf{W} , since we standardize the weight matrix like \mathbf{W} . Thus, we restrict ρ to $\rho \in (-1, 1)$.

Then the likelihood function of the model (1) is given as follows:

$$L(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \rho, \mathbf{X}, \mathbf{W}) = (2\pi\sigma^2)^{-\frac{n}{2}} |\mathbf{I}_n - \rho\mathbf{W}| \exp\left(-\frac{\mathbf{e}'\mathbf{e}}{2\sigma^2}\right) \quad (2)$$

where $\mathbf{y} = (y_1, \dots, y_n)'$, $\mathbf{X} = (\mathbf{x}'_1, \dots, \mathbf{x}'_n)'$, $\mathbf{e} = (e_1, \dots, e_n)'$, $e_i = y_i - \sum_{j=1}^n \rho w_{ij} y_j - \mathbf{x}_i \boldsymbol{\beta}$, and \mathbf{I}_n is an $n \times n$ unit matrix.

3. Posterior Analysis and Simulation

3.1. Joint Posterior Distribution

Since we adopt the Bayesian approach, we complete the model by specifying the prior distribution over the parameters. We use the following independent prior distribution:

$$\pi(\boldsymbol{\beta}, \sigma^2, \rho) = \pi(\boldsymbol{\beta}) \pi(\sigma^2) \pi(\rho)$$

Given a prior density $\pi(\boldsymbol{\beta}, \sigma^2, \rho)$ and the likelihood function given in (2), the joint posterior distribution can be expressed as

$$\pi(\boldsymbol{\beta}, \sigma^2, \rho | \mathbf{y}, \mathbf{X}, \mathbf{W}) \propto \pi(\boldsymbol{\beta}, \sigma^2, \rho) L(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \rho, \mathbf{X}, \mathbf{W}) \quad (3)$$

Finally, we assume the following prior distributions:

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad \sigma^2 \sim IG(\nu_0/2, \lambda_0/2), \quad \rho \sim U(-1, 1)$$

where $IG(a, b)$ denotes an inverse gamma distribution with scale and shape parameters a and b .

Since the joint posterior distribution is given by (3), we can now adopt the MCMC method. The Markov chain sampling scheme can be constructed from the full conditional distributions of ρ , $\boldsymbol{\beta}$ and σ^2 .

3.2. Sampling ρ

From (3), the full conditional distribution of ρ is written as

$$p(\rho | \boldsymbol{\beta}, \sigma^2, \rho, \mathbf{y}, \mathbf{X}, \mathbf{W}) \propto |\mathbf{I}_n - \rho\mathbf{W}| \exp\left(-\frac{\mathbf{e}'\mathbf{e}}{2\sigma^2}\right) \quad (4)$$

As it is difficult to sample from the standard distribution, we examine three approaches for sampling ρ . First, we introduce the GGS, which is applied by [2]. Second, we overview the RMH algorithm, which is extended by [13]. Finally, we propose an ARMH algorithm. These sampling methods are summarized in the following.

3.2.1. Griddy Gibbs Sampler

The GGS was proposed by [1]. This sampling algorithm approximates a cumulative distribution function of the full conditional distribution by each kernel function over a grid of points and uses a numerical integration method, and is sampling method from the full conditional distribution by using the inverse transform method. Let the grid be as follows

$$-1 = a_1 < a_2 < \dots < a_m < a_{m+1} = 1$$

and ρ^i ($i \in \{1, \dots, m\}$), which is centered in the interval $[a_i, a_{i+1}]$. Then, the full conditional distribution in the interval $[a_i, a_{i+1}]$ is approximated as follows

$$\omega_i = \frac{p(\rho^i | \beta, \sigma^2, \mathbf{y}, \mathbf{X}, \mathbf{W})}{\sum_{h=1}^m p(\rho^h | \beta, \sigma^2, \mathbf{y}, \mathbf{X}, \mathbf{W})}$$

Thus, we select the grid a_i^* with probabilities,

$$h(\rho^i) = \frac{\omega_i (a_{i+1} - a_i)}{\sum_{j=1}^m \omega_j (a_{j+1} - a_j)}$$

Finally, we sample ρ from the uniform (a_i^*, a_{i+1}^*) . [22] stated that the choice of the grid of points has to be made carefully and constitute the main difficulty in applying GGS. In this paper, we select the equal interval among $a_{m+1} - a_m$ as in [1]. Then, our numerical experiments examines to choice the size of grid for estimating the spatial correlation.

3.2.2. Random Walk Metropolis-Hastings Algorithm

The RMH method is a simple algorithm because it needs the previous value and a random walk process such as $\phi^{\text{new}} \sim N(\phi^{\text{old}}, \tau^2)$, where ϕ^{old} is the parameter of the previous sampling, and τ denotes the tuning parameter, respectively. Therefore, the following Metropolis step is used: Sample ρ^{new} from

$$\rho^{\text{new}} \sim N(\rho^{\text{old}}, s^2)$$

where s is the tuning parameter. In the numerical example below, we select the tuning parameter such that the acceptance rate lies between 0.4 and 0.6 (see [13]). Next, we evaluate the acceptance probability

$$\alpha(\rho^{\text{old}}, \rho^{\text{new}}) = \min \left(\frac{p(\rho^{\text{new}} | \beta, \sigma^2, \rho, \mathbf{y}, \mathbf{X}, \mathbf{W})}{p(\rho^{\text{old}} | \beta, \sigma^2, \rho, \mathbf{y}, \mathbf{X}, \mathbf{W})}, 1 \right)$$

And finally set $\rho = \rho^{\text{new}}$ with probability $\alpha(\rho^{\text{old}}, \rho^{\text{new}})$, otherwise $\rho = \rho^{\text{old}}$. The proposal value of ρ is not truncated to the interval $(-1, 1)$ because the constraint is part of the target density. Thus, if the proposed value of ρ is not within the interval, the conditional posterior is zero, and the proposal value is rejected with probability one (see [23]). It is well known that the method is not efficient because the convergence is slow for using the previous sampled parameter.

3.2.3. Acceptance-Rejection Metropolis-Hastings Algorithm

An acceptance-rejection Metropolis-Hastings (ARMH) algorithm method was proposed by [15]. This algorithm samples the parameter using the AR and MH steps. Suppose that there is a candidate function $g(\rho)$ such that it is possible to sample directly from $g(\rho)$ by some known method. Then, the AR step proceeds as follows. We sampling the parameter from the candidate function $g(\rho)$, and accepts the candidate draw with probability $p(\rho)/cg(\rho)$. This step is iterated until the candidate draw is accepted.

Next, suppose the candidate ρ^{new} is produced from above AR step. The MH part proceeds as follows. We calculate the acceptance probability, q as following:

$$\begin{aligned} &\text{If } p(\rho^{\text{old}}) < cg(\rho^{\text{old}}), && \text{then } q = 1; \\ &\text{If } p(\rho^{\text{old}}) \geq cg(\rho^{\text{old}}) \text{ and } p(\rho^{\text{new}}) < cg(\rho^{\text{new}}), && \text{then } q = \frac{cg(\rho^{\text{old}})}{p(\rho^{\text{old}})}; \\ &\text{If } p(\rho^{\text{old}}) \geq cg(\rho^{\text{old}}) \text{ and } p(\rho^{\text{new}}) \geq cg(\rho^{\text{new}}), && \text{then } q = \min \left[\frac{p(\rho^{\text{new}})g(\rho^{\text{old}})}{p(\rho^{\text{old}})g(\rho^{\text{new}})}, 1 \right]. \end{aligned}$$

In this step, the candidate draw is accepted with probability q and rejected with probability $1 - q$. If the draw is rejected, the previously sampled value is sampled again. If q is small, the probability of sampling the same value consecutively is high, causing high autocorrelation across sample values (see [24]). Hence, we

should also make q as close to one as possible.

The advantage of this method is that it is free to functional form which differs from the GGS and RMH. In this paper, in order to construct the candidate function, we utilize the result of [7], which showed the consistency and asymptotic normality of quasi-ML estimators of model parameters, to the candidate density. Then, we construct the candidate density $g(\rho)$ as an approximation to the the conditional posterior density by omitting the determinant $|\mathbf{I} - \rho\mathbf{W}|$ as follows:

$$g(\rho) \sim N(\hat{\mu}_\rho, \hat{\sigma}_\rho^2) \quad (5)$$

where $\hat{\mu}_\rho = (\mathbf{y}'\mathbf{W}\mathbf{W}\mathbf{y})^{-1} \{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W}\mathbf{y}\}$ and $\hat{\sigma}_\rho^2 = \sigma^2 (\mathbf{y}'\mathbf{W}\mathbf{W}\mathbf{y})^{-1}$. Thus we sample ρ^{new} from the distribution, and apply the ARMH algorithm.

3.3. Sampling Other Parameters

The full conditional distributions of $\boldsymbol{\beta}$ and σ^2 are

$$\boldsymbol{\beta} \sim N(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}), \quad \text{and} \quad \sigma^2 \sim IG(\hat{\nu}/2, \hat{\lambda}/2)$$

where $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}} \{ \sigma^{-2} \mathbf{X}'(\mathbf{y} - \rho\mathbf{W}\mathbf{y}) + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\beta}_0 \}$, $\hat{\boldsymbol{\Sigma}} = (\sigma^{-2} \mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1}$, $\hat{\nu} = n + \nu_0$, and $\hat{\lambda} = \mathbf{e}'\mathbf{e} + \lambda_0$. These parameters are easily sampled from the Gibbs sampler (see [25]).

4. Comparison of MCMC Methods

4.1. Measures of Efficiency for Comparison

In this section, we examine the properties of three MCMC methods by simulated data sets. Desirable properties for sampling methods in MCMC are efficiency and well mixing, which yield fast convergence. [17] compared from the view point of acceptance rate in the AR and MH step. [26] [27] evaluated the efficiency of sampling methods, comparing the inefficiency factor and time of MCMC simulation. Following previous literatures, we also compare inefficiency factor and computational time.

The inefficiency factor is defined as $1 + 2\sum_{s=1}^{\infty} r_s$ where r_s is the sample autocorrelation at lag s calculated from the sampled values. It is used to measure how well the chain mixes and is the ratio of the numerical variance of the sample posterior mean to the variance of the sample mean from the hypothetical uncorrelated draws (see [28]).

4.2. Data Generating Process and Estimation Procedures

We now explain the simulated data for an experiment. First, we give the weight matrix as an exogenous variable. We construct the spatial weight matrix \mathbf{W} as follows: 1) generate c_{ij} for $i > j$ from Bernoulli distribution with a probability of success 0.3, 2) set $c_{ij} = c_{ji}$ for $i \neq j$ and $c_{ij} = 0$ for $i = j$, and 3) compute $w_{ij} = c_{ij} / \sum_{j=1}^n c_{ij}$ for all i, j . Next, for the independent variables $\mathbf{x}_i = (1, x_{1i}, x_{2i}, x_{3i})$, we take the standard normal variates and set the \mathbf{X} , which are $n \times 4$ covariate matrices.

Given \mathbf{W} , \mathbf{X} , $\rho = -0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9$, and $n = 50, 100, 200$, the true data generating process is as follows:

$$y_i = \rho \sum_{j=1}^n w_{ij} y_j + \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i \quad (6)$$

where the ϵ_i is normally and independently distributed with $E(u_i) = 0$ and $E(u_i^2) = \sigma^2$. The parameter is set to be $\boldsymbol{\beta}' = (\beta_0, \beta_1, \beta_2, \beta_3) = (1, 1, 1, 1)$ and $\sigma^2 = 0.1$, respectively. The parameters of ρ for simulated data reflect the values obtained in [12]. All the results in this paper were calculated using the Ox version 5.1 (see [29]).

The prior distributions are as follows:

$$\beta \sim N(\mathbf{0}, 100 \times \mathbf{I}_4), \quad \sigma^2 \sim IG(1.0/2, 0.01/2), \quad \text{and} \quad \rho \sim U(-1, 1)$$

We perform the MCMC procedure by generating 35,000 draws in a single sample path and discard the first 20,000 draws as the initial burn-in. For the GGS, we consider the number of grid, $m = 50, 100, 300$ for estimating the parameters.

4.3. Results of Comparison

Table 1 reports inefficiency factors by using three methods. Although there are some differences, the performances of the GGS are almost equivalent to those of the ARMH. In addition, these algorithms are more efficient than RMH. For example, from the table in $n = 50$, the inefficiency factors calculated by the ARMH are smaller than the other methods. However, if spatial correlation is positive strong such as $\rho = 0.9$, the value by the GGS ($m = 100$) has the smallest inefficiency factor. Next, we focus on the results in $n = 100$. In this case, the GGS ($m = 50, 100$) perform the best for $\rho = 0.6, 0.9$, respectively. In the case of $n = 200$, the values of the GGS ($m = 100$) and the ARMH are similar in each parameter. We can also find such similarity in sample paths and autocorrelation functions. **Figure 1** shows the results of MCMC simulation in each method in the cases of $\rho = 0.3$, $n = 50$ and $m = 100$. The figure shows that the marginal posterior densities (middle of the figure)

Table 1. Inefficiency factor of models.

Observation: $n = 50$						
Parameter	RMH	GGS			ARMH	
ρ		$m = 50$	$m = 100$	$m = 300$		
-0.9	7.2	3.2	3.4	3.4	2.8	
-0.6	27.6	4.4	4.4	4.7	3.7	
-0.3	15.4	23.7	6.6	6.9	4.3	
0	41.6	9.0	10.1	11.5	6.6	
0.3	79.8	24.6	19.6	20.9	13.1	
0.6	117.0	46.3	45.2	52.3	44.6	
0.9	806.1	312.9	223.1	327.2	324.9	
Observation: $n = 100$						
Parameter	RMH	GGS			ARMH	
ρ		$m = 50$	$m = 100$	$m = 300$		
-0.9	10.7	4.8	5.0	5.3	4.7	
-0.6	17.0	6.7	7.3	7.5	4.6	
-0.3	34.7	9.0	10.2	10.9	5.0	
0	72.4	15.4	16.2	17.8	9.5	
0.3	85.1	24.5	25.5	32.6	19.9	
0.6	202.3	36.1	56.6	65.8	51.3	
0.9	609.1	379.3	338.0	342.1	338.9	
Observation: $n = 200$						
Parameter	RMH	GGS			ARMH	
ρ		$m = 50$	$m = 100$	$m = 300$		
-0.9	22.2	7.1	8.3	5.7	7.8	
-0.6	31.0	11.5	12.4	13.5	9.0	
-0.3	64.8	17.6	17.6	19.1	13.8	
0	75.7	23.5	26.4	33.6	23.7	
0.3	163.6	57.4	67.3	65.6	50.5	
0.6	697.3	164.2	117.5	163.3	159.3	
0.9	860.4	695.1	628.7	694.0	780.6	

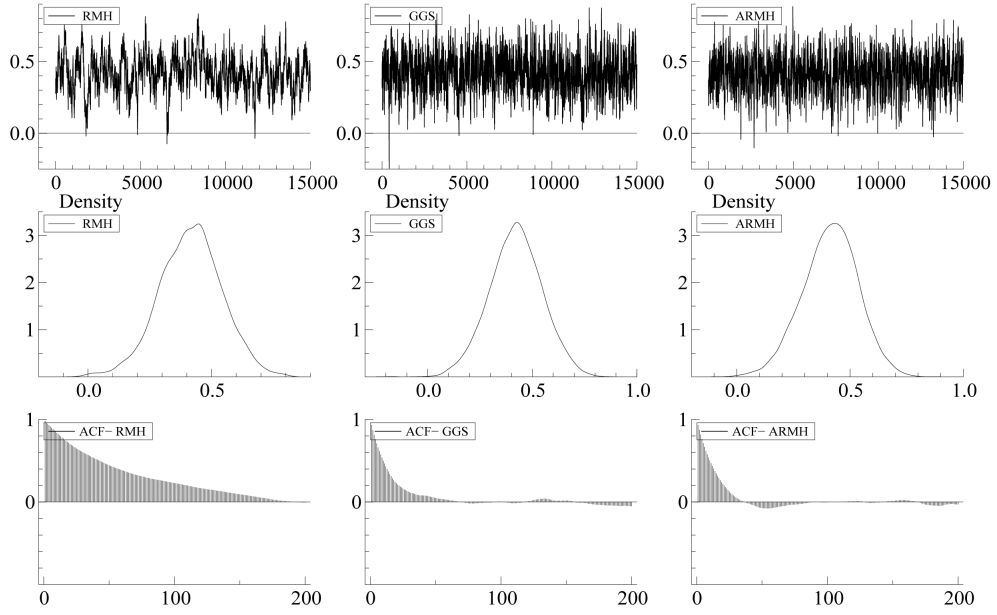


Figure 1. Sample paths, sample autocorrelation and posterior density of $\rho = 0.3$, $n = 50$.

have similar shapes but that the sample paths (top of the figure) and autocorrelation functions (bottom of the figure) are different. From the sample paths, we can find that the ARMH and GGS mix better than the RMH. As same as the sample paths, autocorrelation functions shows the same tendency. The figure of autocorrelation indicates that both GGS and ARMH perform similarly in the autocorrelation disappear. On the contrary, the result for the RMH indicates that serious autocorrelation for parameter at large lag length.

Table 2 shows CPU time on a Pentium Core2 Duo 2.4GHz including discarded and rejected draws. For the GGS, the computation time depends on the number of grid because the increase of grid number causes the cost of computation time. In all cases, the GGS ($m = 50$) overwhelms the others. If we focus on the case of $n = 50$, the computational time of the GGS ($m = 100$) are as same as those of the RMH and ARMH methods. Furthermore, if $n = 200$, the GGS needs much shorter time than the RMH and ARMH methods. Summarizing the results of inefficiency factors and computational time, if the number of observation is not only small (like $n = 50$) but also large, then it is suitable to use the GGS. In addition, the choice of grid number affects to the computational time. In this numerical experiments, the results of selecting $m = 100$ seem to work well in terms of inefficiency factors and computational time.

Table 3 shows the results with acceptance probabilities in both AR and MH parts in the ARMH. From the table, the acceptance probabilities in those part are exceed 89%. This result shows that our candidate function seems to work well, and the probabilities of sampling the same value consecutively are low. However, our ARMH algorithm does not improve the values of inefficiency factor. Thus, we think that the SAR model has the problem of identification.

Figure 2 and **Table 4** depict the sample path and the correlation among the parameters in the case of $n = 100$, $\rho = 0.9$, $m = 100$ using the GGS. From β_2 to β_4 and σ^2 in the figure, the MCMC draws seem to be well mixing. In addition, correlations among these parameters are very small. On the other hand, strong correlation between β_0 and ρ can be found from the figure. Moreover, the correlation between β_0 and ρ is -0.995 from the table. Therefore, we assume that the spatial correlation and constant term is weakly identified.

5. Concluding Remarks

This paper reviewed the MCMC estimation procedures for sampling the spatial correlation of SAR model, and proposed the ARMH algorithm as more efficient than the RMH in order to show the property of the GGS proposed by [2]. To illustrate the differences between the estimates of three MCMC methods, we compared these algorithms by simulated data set. From the Monte Carlo experiments, we found that the GGS was the most efficient algorithm with respect to the mixing, efficiency and computational requirement of the MCMC. Moreover,

Table 2. Time of convergence.

Observation: $n = 50$					
Parameter	RMH	GGS			ARMH
ρ		$m = 50$	$m = 100$	$m = 300$	
-0.9	22.17	11.12	22.68	1:05.96	24.24
-0.6	23.22	11.57	23.06	1:05.99	23.95
-0.3	23.31	11.71	23.21	1:07.18	23.99
0	23.27	11.83	23.27	1:07.87	24.01
0.3	23.20	12.26	23.10	1:09.49	23.99
0.6	24.16	12.07	22.70	1:08.36	24
0.9	23.17	12.06	23.64	1:08.66	24.02
Observation: $n = 100$					
Parameter	RMH	GGS			ARMH
ρ		$m = 50$	$m = 100$	$m = 300$	
-0.9	1:31.73	18.67	35.49	1:37.61	1:44.90
-0.6	1:41.36	17.25	35.75	1:37.83	1:42.99
-0.3	1:43.81	19.93	36.64	1:39.25	1:43.22
0	1:40.10	18.30	40.15	1:38.47	1:43.53
0.3	1:40.90	19.90	40.04	1:41.40	1:43.55
0.6	1:41.73	18.91	37.35	1:43.97	1:43.13
0.9	1:43.36	17.93	37.89	1:40.33	1:42.27
Observation: $n = 200$					
Parameter	RMH	GGS			ARMH
ρ		$m = 50$	$m = 100$	$m = 300$	
-0.9	8:40.79	26.88	56.62	2:43.63	9:05.58
-0.6	8:43.81	26.66	56.76	2:45.73	9:07.63
-0.3	8:59.71	26.74	58.84	2:44.22	9:08.03
0	8:57.87	26.92	57.48	2:46.64	8:56.41
0.3	9:03.95	27.02	58.49	2:45.99	8:51.45
0.6	9:12.82	28.24	58.13	2:48.13	9:01.35
0.9	9:22.86	27.10	57.84	2:48.15	8:59.61

Note: Time denotes CPU time on a Pentium Core2 Duo, including discarded and rejected draws.

Table 3. Acceptance probability of the ARMH methods.

Parameter	$n = 50$		$n = 100$		$n = 200$	
	AR step	MH step	AR step	MH step	AR step	MH step
-0.9	0.9866	0.9116	0.9578	0.8975	0.9881	0.9505
-0.6	0.9999	0.9500	0.9999	0.9438	1.0000	0.9724
-0.3	1.0000	0.9848	1.0000	0.9805	1.0000	0.9906
0	1.0000	0.9849	1.0000	0.9787	1.0000	0.9949
0.3	1.0000	0.9670	1.0000	0.9544	1.0000	0.9861
0.6	0.9991	0.9553	0.9958	0.9375	1.0000	0.9802
0.9	0.9997	0.9716	0.9977	0.9649	0.9997	0.9821

Table 4. Correlation of parameters.

	β_1	β_2	β_3	β_4	σ^2
β_2	-0.050				
β_3	0.087	0.106			
β_4	-0.059	0.205	0.183		
σ^2	0.161	-0.007	0.002	-0.025	
ρ	-0.995	0.054	-0.075	0.078	-0.160

Note: True parameter is 0.9. The number of observation set to be 100.

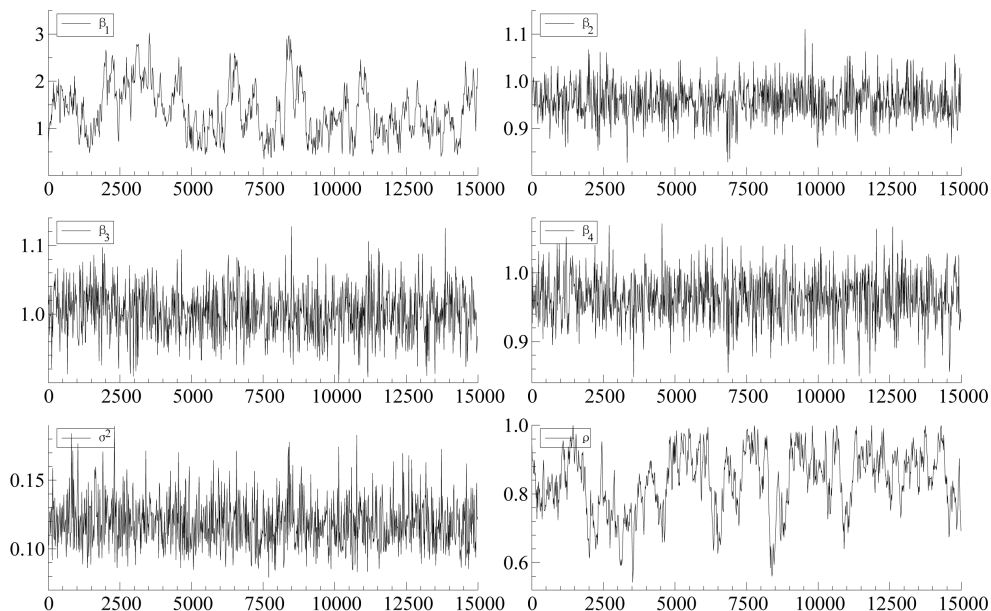


Figure 2. Sample paths of SAR model with GGS ($\rho = 0.9$, $n = 100$, $m = 100$).

the results of selecting $m = 100$ seem to work well in terms of inefficiency factors and computational time. Therefore, the GGS is beneficial algorithm for estimating the spatial parameter as same as the result of [22].

Finally, we will state our remaining issues. In this paper, we found that the GGS was the most efficient algorithm in sampling the intensity of spatial interaction. On the other hand, we showed the problem of the SAR model such that the spatial correlation and constant term was weakly identified. Thus, we have to construct the model which is identified, or appropriate algorithm to sample the intensity of spatial interaction. Furthermore, we found that the number of grids is appropriate when $m = 100$. In this paper, we could not derive the theoretical reason why $m = 100$ was appropriate number of grids, that was, we only showed the results of Monte Carlo experiments. However, it is important to know the properties of the existing sampling methods, though research on the convergence of the GGS algorithm has never been examined. We think that, in this respect, our experiment gives the benchmark in applied econometrics.

Acknowledgements

We gratefully acknowledge helpful discussions and suggestions with Toshiaki Watanabe on several points in the paper. Mototsugu Fukushima gave insightful comments and suggestions when we made a presentation at Japan Association for Applied Economics in June, 2009. This research was partially supported by KAKENHI # 25245035 for K. Kakamu.

References

[1] Ritter, C. and Tanner, M. (1992) Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler.

- Journal of the American Statistical Association*, **87**, 861-868. <http://dx.doi.org/10.1080/01621459.1992.10475289>
- [2] Ohtsuka, Y. and Kakamu, K. (2009) Estimation of Electric Demand in Japan: A Bayesian Spatial Autoregressive AR(p) Approach. In: Schwartz, L.V., Ed., *Inflation: Causes and Effects*, Nova Science Publisher, New York, 156-178.
 - [3] Anselin, L. (2003) Spatial Externalities, Spatial Multipliers, and Spatial Econometrics. *International Regional Science Review*, **26**, 153-166. <http://dx.doi.org/10.1177/0160017602250972>
 - [4] Gelfand, A.E., Banerjee, S., Sirmans, C.F., Tu, Y. and Ong, S.E. (2007) Multilevel Modeling Using Spatial Processes: Application to the Singapore Housing Market. *Computational Statistics and Data Analysis*, **51**, 3567-3579. <http://dx.doi.org/10.1080/01621459.1990.10476213>
 - [5] Anselin, L. (2010) Thirty Years of Spatial Econometrics. *Papers in Regional Science*, **89**, 3-25. <http://dx.doi.org/10.1111/j.1435-5957.2010.00279.x>
 - [6] Ord, K. (1975) Estimation Methods for Models for Spatial Interaction. *Journal of the American Statistical Association*, **70**, 120-126. <http://dx.doi.org/10.1080/01621459.1975.10480272>
 - [7] Lee, L.F. (2004) Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica*, **72**, 1899-1925. <http://dx.doi.org/10.1111/j.1468-0262.2004.00558.x>
 - [8] Conley, T.G. (1999) GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics*, **92**, 1-45. [http://dx.doi.org/10.1016/S0304-4076\(98\)00084-0](http://dx.doi.org/10.1016/S0304-4076(98)00084-0)
 - [9] Kelejian, H.H. and Prucha, I.R. (1999) A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, **40**, 509-533. <http://dx.doi.org/10.1111/1468-2354.00027>
 - [10] Anselin, L. (1980) A Note on Small Sample Properties of Estimators in A First-order Spatial Autoregressive Model. *Environment and Planning A*, **14**, 1023-1030. <http://dx.doi.org/10.1068/a141023>
 - [11] LeSage, J.P. (1997) Regression Analysis of Spatial Data. *The Journal of Regional Analysis and Policy*, **27**, 83-94.
 - [12] Kakamu, K. and Wago, H. (2008) Small-Sample Properties of Panel Spatial Autoregressive Models: Comparison of the Bayesian and Maximum Likelihood Methods. *Spatial Economic Analysis*, **3**, 305-319. <http://dx.doi.org/10.1080/17421770802353725>
 - [13] Holloway, G., Shankar, B. and Rahman, S. (2002) Bayesian Spatial Probit Estimation: A Primer and an Application to HYV Rice Adoption. *Agricultural Economics*, **27**, 383-402. <http://dx.doi.org/10.1111/j.1574-0862.2002.tb00127.x>
 - [14] Ohtsuka, Y., Oga, T. and Kakamu, K. (2010) Forecasting Electricity Demand in Japan: A Bayesian Spatial Autoregressive ARMA Approach. *Computational Statistics & Data Analysis*, **54**, 2721-2735. <http://dx.doi.org/10.1016/j.csda.2009.06.002>
 - [15] Tierney, L. (1994) Markov Chains for Exploring Posterior Distributions (with Discussion). *Annals of Statistics*, **22**, 1701-1728. <http://dx.doi.org/10.1214/aos/1176325750>
 - [16] Chib, S. and Greenberg, E. (1994) Bayes Inference in Regression Models with ARMA(P, q) Errors. *Journal of Econometrics*, **64**, 183-206. [http://dx.doi.org/10.1016/0304-4076\(94\)90063-9](http://dx.doi.org/10.1016/0304-4076(94)90063-9)
 - [17] Watanabe, T. (2001) On Sampling the Degree-of-Freedom of Student's-t Disturbances. *Statistics & Probability Letters*, **52**, 177-181. [http://dx.doi.org/10.1016/S0167-7152\(00\)00221-2](http://dx.doi.org/10.1016/S0167-7152(00)00221-2)
 - [18] Mitsui, H. and Watanabe, T. (2003) Bayesian Analysis of GARCH Option Pricing Models. *Journal of the Japan Statistical Society (Japanese Issue)*, **33**, 307-324.
 - [19] LeSage, J.P. and Pace, R.K. (2008) Introduction to Spatial Econometrics (Statistics: A Series of Textbooks and Monographs). Chapman and Hall/CRC, London.
 - [20] Stakhovych, S. and Bijmolt, T.H.A. (2009) Specification of Spatial Models: A Simulation Study on Weights Matrices. *Papers in Regional Science*, **88**, 389-408. <http://dx.doi.org/10.1111/j.1435-5957.2008.00213.x>
 - [21] Sun, D., Tsutakawa, R.K. and Speckman, P.L. (1999) Posterior Distribution of Hierarchical Models Using CAR(1) Distributions. *Biometrika*, **86**, 341-350. <http://dx.doi.org/10.1093/biomet/86.2.341>
 - [22] Bauwens, L. and Lubrano, M. (1998) Bayesian Inference on GARCH Models Using the Gibbs Sampler. *The Econometrics Journal*, **1**, 23-46. <http://dx.doi.org/10.1111/1368-423X.11003>
 - [23] Chib, S. and Greenberg, E. (1998) Analysis of Multivariate Probit Models. *Biometrika*, **85**, 347-361. <http://dx.doi.org/10.1093/biomet/85.2.347>
 - [24] Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, **49**, 327-335.
 - [25] Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, **85**, 398-409. <http://dx.doi.org/10.1080/01621459.1990.10476213>
 - [26] Asai, M. (2005) Comparison of MCMC Methods for Estimating Stochastic Volatility Models. *Computational Eco-*

nomics, **25**, 281-301. <http://dx.doi.org/10.1007/s10614-005-2974-4>

- [27] Asai, M. (2006) Comparison of MCMC Methods for Estimating GARCH Models. *Journal of the Japan Statistical Society*, **36**, 199-212. <http://dx.doi.org/10.14490/jjss.36.199>
- [28] Chib, S. (2001) Markov Chain Monte Carlo Methods: Computation and Inference. In: Heckman, J.J. and Leamer, E., Eds., *Handbook of Econometrics*, Elsevier, Amsterdam, 3569-3649.
- [29] Doornik, J.A. (2006) Ox: An Object Oriented Matrix Programming Language. Timberlake Consultants Press, London.