TGU-ECON Discussion Paper Series
#2019-5

An Efficient Bayesian Estimation for the Space-Time Stationary Condition

Yoshihiro Ohtsuka
Faculty of Economics, Tohoku Gakuin University

September 2019

An efficient Bayesian estimation for the space-time stationary condition*

Yoshihiro Ohtsuka†

Tohoku Gakuin University, Faculty of Economics

Abstract

We propose an efficient posterior sampling algorithm for a spatial dynamic panel data model in terms of Bayesian inference. The stationary condition of the model mutually depends on three parameters: simultaneous and lagged spatial dependencies and serial correlation. In practice, the interdependence between these parameters yields low convergence to their target marginal distribution and fails to converge in the time available. To accelerate sampling efficiency, we develop a Bayesian estimation algorithm for these parameters using a Tailor approximation and blocked Metropolis-Hastings (TaB-MH) algorithm. The experimental study and illustration show that the TaB-MH algorithm is superior to the random walk MH.

Keywords: Taylor approximation and blocked Metropolis-Hastings algorithm; Markov chain Monte Carlo method; Inefficiency factor; Spatial dynamic panel data model;

JEL classification: C4, C33, R15

*We would like to thank the participants at the 3rd International Conference on Econometrics and Statistics (June 2019) for their helpful comments on this work. This work was supported by a Grant-in-Aid for Scientific Research by the Japan Society for the Promotion of Science (No.17K03770).

†1-3-1 Tsuchitoi, Aoba-ku, Sendai, Miyagi 980-8511, Japan; Tel./Fax: +81-22- 721-3275; E-mail: ohtsuka@mail.tohoku-gakuin.ac.jp

1 Introduction

Bayesian spatial econometrics attracts attention as empirical studies in spatial economics increasingly aim to examine geographical relationships and temporal structures. Since the contribution of a seminal work (Anselin, 1988) and comprehensive textbook (LeSage and Pace, 2009), the Bayesian approach, that is, Markov chain Monte Carlo (MCMC) methods and theory, rapidly extend to various empirical issues on spatial econometrics[1]. Gibbs sampling is strategy of sequentially drawing from the full conditional posterior distributions and is the main framework in MCMC methods. This sampling allows for a sophisticated model of spatial dependence and heterogeneity (Anselin, 2010), and complex hierarchical specification (Lacombe and McIntyre, 2016). Improvements in computational processing enabled such extensions of spatial econometric modeling.

Recent developments in empirical analysis show an increasing interest in high-dimensional data on land prices, micro economic data on states or municipalities, and so on. Analyses of high-dimensional data requires high numerical calculations. We focus on the accelerating MCMC algorithm for estimating spatial parameters, which one of main streams of in the literature on improving numerical computations and estimation methods[2]. However, Bayesian computation has critical issues such as simulation convergence and posterior propriety.

Gibbs sampling works well when the number of iterations is large. However, a researcher must decide how many times to select the iteration number. As mentioned above, increasing the iterations in handling a large data has a high computational cost and takes a lot of time in practice. Thus, for an empirical researcher, the desirable properties for sampling methods are efficiency and mixing well, which yield fast convergence or reduces the number of the Monte Carlo simulations. These points are often overlooked in empirical analysis. For spatial models, there are some studies on the development of sampling methods (Ohtsuka and Kakamu, 2009; Bivand, *et al*., 2014; Ohtsuka and Kakamu, 2015; Wolf *et al*., 2018) and discussions of the properties of sampling methods. Some studies advocate the importance of showing how the techniques are efficient or inefficient.

In this work, we deal with a spatial dynamic panel data (SDPD) model. Prior works use spatial panel data models to analyze spatial dependency and spillover effects in several fields, such as regional economics, urban economics, and real estate. Among the existing models, the SDPD model has the advantage of being able to evaluate how observations are related over time and space because the model allows spatiotemporal interdependency involving the dependent variable. A specification of this model has been development (see

[1]Arbia (2014) and Elhorst (2014) published well known, sophisticated textbooks, and summarized from the viewpoints of frequentist approaches.

[2]Typical improvements in numerical computations adopt approaches to the determinant of a weight matrix, such as Chebyshev approximation (Pace and LeSage, 2004) and characteristic polynomial approach (Smirnov and Anselin, 2001).

Elhorts, 2004, for an excellent review). In this work, we extend the SDPD model with the random effects used in Parent and LeSage (2012) by incorporating heteroskedasticity as in Mills and Parent (2014).

Although the SDPD model is highly flexible, the inclusion of the time series structure adds complexity to the stationary condition. Parent and LeSage (2011) show that the stationary condition of the model mutually depends on three parameters related to simultaneous spatial correlation, lagged serial correlation, and lagged spatial correlation parameters. In this work, we refer to the stationary condition of the SDPD model as the space-time stationary condition. Thus, the model is assumed to be the explicit relationship among parameters under the space-time stationary condition. When we estimate the model, the ranges of these parameters depend on the space-time condition.

With the sampling algorithm for the SDPD model, most studies conduct analyses using a random walk Metropolis-Hastings (RW-MH) algorithm and single sampling related to spatial and serial correlation parameters. The RW-MH algorithm is an inefficient estimation method because the algorithm depends on the previous sample value to draw the posterior sample from the proposed distribution. Moreover, prior studies assume that these parameters have an implicit relationship under the space-time stationary condition. However, sampling individually correlated parameters causes a slow convergence (Liu 1994). Thus, the existing method fails to converge in the time available in practice. While some studies on the sampling algorithm for spatial models focus on only simultaneous spatial parameters, few studies examine the case of multiple spatial and dynamic parameters. Therefore, it is necessary to accelerate the MCMC algorithm for the SDPD model and to demonstrate the sampling efficiency or inefficiency to empirical researchers.

In this work, we propose a Taylor approximation and blocked Metropolis-Hastings (TaB-MH) algorithm to accelerate convergence. This approach is widely used in econometric analysis, for instance, the stochastic volatility model (Watanabe and Omori, 2004) and the dynamic stochastic general equilibrium model (Chib and Ramamurthy, 2010) are popular. Our approach is based on Chib and Ramamurthy (2010) and joint sampling parameters related to space and time correlations and considering the proposal distribution, which mimics the target distribution using the Taylor expansion.

We illustrate the properties of both TaB-MH and RW-MH algorithm using a simulated data set given a small sample and strong correlations in spatial and temporal dependence. From the results, we find that the TaB-MH is the more efficient method in terms of inefficiency factors. Moreover, an empirical illustration on a regional spillover in Japan also works better and with fewer simulations than the existing approach.

The remainder of this paper is organized as follows. Section 2 introduces the SDPD model and MCMC methods. Section 3 shows the Monte Carlo experiments using simulated data and demonstrating applied work using regional productivity in Japan. Finally, we provide concluding remarks.

2 Econometric model and Bayesian inference

2.1 Spatial dynamic panel data model

We first introduce the spatial dynamic panel data model[3] with heteroskedasticity among regions in Mills and Parent (2014). Let y_t define the $n \times 1$ vector of observations at time t, and X_t denotes the $n \times k$ matrix of exogenous variables for $t = 1, \ldots, T$. The dynamic spatial lag model is defined as follows:

$$y_t = \rho W y_t + \phi y_{t-1} + \theta W y_{t-1} + \alpha \iota_n + X_t \beta + \eta_t, \tag{1}$$

$$\eta_t = \mu + \sigma \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \Lambda), \quad \Lambda = diag(\lambda_1, \ldots, \lambda_n), \tag{2}$$

where ρ captures the degree of simultaneous spatial interaction, the scaler ϕ and θ are reflected parameters of the strength of the first order serial and spatial dependence. W denotes the weight matrix including the geographic relationship among regions, and the components of the weight matrix are defined with the row-standardized contiguity dummy[4]. ι_n is the $n \times 1$ vector of ones associated with the constant term α.

With respect to the vector of the constant term μ in the equation (2), the $n \times 1$ vector μ is assumed to be the random effect. Thus, μ is assumed to be the following normal distribution such as $\mu \sim \mathcal{N}(0, \tau^2)$ where τ^2 is the unknown parameter of variance. ϵ_t is assumed to be independent and identically distributed and follows the normal distribution with mean zeros and variance I_n, which is $n \times n$ unit matrix. λ_i denotes the auxiliary parameter for the heteroskedasticity of the error distribution across regions. As Mills and Parent (2014) note, the extension of the model including the regional variance scaling parameter λ_i causes slow convergence because of the identification problem among the variance σ^2 and Λ. Thus, we assume that the error term ϵ_t follows the standardized Student's t-distribution. The variation scale of heteroskedasticity is assumed to be independent and identically distributed and following the chi-squared distribution,

$$\frac{\nu - 2}{\lambda_i} \sim \chi^2(\nu),$$

where ν represents an unknown parameter of the degree of freedom. Then, the error term follows the standardized Student's t-distribution; $\nu > 2$ is assumed to satisfy a finite variance. A pre-sample y_0 is assumed to be an exogenous variable because the main purpose of this study is to examine the improvements in sampling efficiency for the space-time parameters (ρ, ϕ, and θ).

Next, we discuss the stationary condition of the SDPD model. Allowing spatial interaction at both si-

[3]This work discusses the spatial autoregressive type of dynamic panel data models. Refer to the following references for other types of models such as the spatial error model and spatial Durbin model in Parent and LeSage (2011) and Mills and Parent (2014), respectively.

[4]As Stakhovych and Bijmolt (2009) recommend the contiguity dummy in their numerical experiments, we adopt the same approach.

multaneous, serial, and temporal dynamics, the stationary condition is more complex than the stationary condition of the conventional dynamic panel data model. Parent and LeSage (2011, 2012) show the stationary conditions as follows:

$$S(\psi) = \begin{cases} \phi + (\rho + \theta)\varpi_{\max} < 1 & (\rho + \theta \geq 0) \\ \phi + (\rho + \theta)\varpi_{\min} < 1 & (\rho + \theta < 0) \\ \phi - (\rho + \theta)\varpi_{\max} > -1 & (\rho - \theta \geq 0) \\ \phi - (\rho + \theta)\varpi_{\min} > -1 & (\rho - \theta < 0) \end{cases} \tag{3}$$

where $\psi = (\rho, \phi, \theta)'$ and $\varpi_{\min}$ ($\varpi_{\max}$) is the minimum (maximum) eigenvalue of W. It is obvious from the above expression that the stationary condition of the SDPD model mutually depends on three parameters such as simultaneous and lagged spatial dependences and serial clustering. Hereinafter, this condition is called the space-time stationary condition. Thus, the space-time stationary constraints must draw parameters ρ, ϕ, and θ consistent with the condition (3) to estimate the model.

To simplify the notations, let $\vartheta = (\psi, \alpha, \beta, \sigma^2, \mu, \tau^2, \lambda, \nu)$, $\lambda = (\lambda_1, \ldots, \lambda_n)$, $y = (y_1', \ldots, y_T')'$, $y_- = (y_0', \ldots, y_{T-1}')'$, and $X = (X_1', \ldots, X_T')'$. Therefore, the log likelihood function is the following:

$$\ln \mathscr{L}(y|\vartheta, y_0, X, W) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega| + T\sum_{j=1}^{n}\ln(1 - \rho\varpi_j) - \frac{1}{2}e'\Omega^{-1}e, \tag{4}$$

where $e = \{\iota_T \otimes (I_n - \rho W)\}y - \{\iota_T \otimes (\phi I_n + \theta W)\}y_- - \iota_{nT}\alpha - X\beta$, $\Omega = \tau^2(J_T \otimes I_n) + \sigma^2(I_T \otimes \Lambda)$ and $J_T = \iota_T\iota_T'$.

2.2 Bayesian analysis

2.2.1 Markov chain Monte Carlo method

Our estimation procedure relies on Bayesian inference, such as the MCMC method. Empirical studies using the SDPD model are mostly based on a quasi-maximum likelihood (QML) estimation via the concentrated log-likelihood approach (Yu *et al.*, 2008; Su and Yang, 2015)[5]. The QML method depends on asymptotic properties while the MCMC method does not. The MCMC evaluates the posterior distributions of the parameters conditioned on the data. Moreover, for regional economic data, the number of time series is small. Thus, it makes sense to estimate the SDPD model using the MCMC method.

Since we adopt Bayesian inference, we evaluate the posterior distribution given the prior distribution

[5] As an alternative approach, Lee and Yu (2014) propose the generalized method of moments.

and the likelihood. The joint posterior distribution is given by

$$p(\vartheta|y, y_0, X, W) \propto p(\vartheta)\mathscr{L}(y|\vartheta, y_0, X, W), \tag{5}$$

where $p(\vartheta)$ denotes the joint prior distribution of the parameters. As the dimension of the parameter set is high to directly evaluate the joint posterior distribution, we simulate the parameter from the full conditional distribution. Thus, we must specify the joint prior distribution. The joint prior distribution can be decomposed as follows:

$$p(\vartheta) = p(\alpha)p(\beta)p(\sigma^2)p(\mu|\tau^2)p(\tau^2)p(\psi)p(\lambda|\nu)p(\nu).$$

We assume that the prior distributions for the parameters (α, β, τ^2, σ^2, and ψ, ν) are all independent. Since the joint posterior distribution is given by (5), we can now adopt the MCMC method. This approach must use multiple iterations to evaluate the marginal posterior distribution in the joint posterior distribution. It is analytically difficult to evaluate the marginal posterior distribution if the joint posterior distribution is complicated. Then, we draw the parameters from the full conditional distributions that use Markov sampling and Monte Carlo integration to approximate the full conditional distribution. For the SDPD model, we implement the MCMC algorithm, which involves the following steps

0. Initial ϑ
1. Generate β from $p(\beta|\vartheta_{-\beta}, y, y_0, X, W)$
2. Generate σ^2 from $p(\sigma^2|\vartheta_{-\sigma^2}, y, y_0, X, W)$
3. Generate μ from $p(\mu|\vartheta_{-\mu}, y, y_0, X, W)$
4. Generate τ^2 from $p(\tau^2|\vartheta_{-\tau^2}, y, y_0, X, W)$
5. Generate α from $p(\alpha|\vartheta_{-\alpha}, y, y_0, X, W)$
6. Generate ψ from $p(\psi|\vartheta_{-\psi}, y, y_0, X, W)$
7. Generate λ from $p(\lambda|\vartheta_{-\lambda}, y, y_0, X, W)$
8. Generate ν from $p(\nu|\vartheta_{-\nu}, y, y_0, X, W)$
9. Repeat 1. to 8.

where $\vartheta_{-\beta}$ denotes the parameters of ϑ excluding β. This strategy of sequentially drawing from the full conditional posterior distributions is called Gibbs sampling. When the number of iterations is large, the drawing from the full conditionals will yield a posterior sample sequence that can be averaged to produce parameter estimates in the same manner as Monte Carlo integration (see Gilks *et al.*, 1996).

The Gibbs sampling approach is a powerful tool that is broadly used in econometric models. However, in the case of the SDPD model involving an explicit relationship among parameters such as the space-time stationary condition, the critical question is how to draw the parameters ψ in step 6. To estimate these parameters, the RW-MH algorithm, which generates a single parameter at a time, is widely applied. This method produces a highly correlated and also between parameters sample sequence and yields low convergence to the target marginal distribution. Thus, it is necessary to repeat the sampling for a large number of iterations to obtain independent posterior samples. In this work, we propose a more efficient estimation method than the existing approach and explain the sampling algorithm for the SDPD model in the following subsections.

2.2.2 Random walk Metropolis-Hastings algorithm

The RW-MH algorithm has been widely used in empirical analysis. In this work, we illustrate a sampling scheme of ρ as in Parent and LeSage (2012). The authors assumed the following as the joint prior distribution of the parameter ψ:

$$p(\psi) = p(\rho)p(\theta|\rho)p(\phi|\rho,\theta).$$

Moreover, the authors assumed the individual prior distribution (ρ, ϕ, θ) as follows:

$$\rho \sim \mathcal{U}(-1,1), \quad \theta|\rho \sim \mathcal{U}(-1+|\rho|, 1+|\rho|), \quad \phi|\rho,\theta \sim \mathcal{U}(-1+|\rho-\theta|, 1+|\rho+\theta|)$$

The full conditional distribution of ρ is as follows:

$$p(\rho|\vartheta_{-\rho}, y, y_0, X, W) \propto |I_n - \rho W| \exp\left(-\frac{1}{2}e'\Omega^{-1}e\right). \tag{6}$$

It is difficult to directly draw the parameter. Thus, we employ the MH algorithm with the random walk chain. The proposal value is drawn from the following distribution,

$$\rho_{new} \sim \mathcal{N}(\rho_{old}, s_\rho^2),$$

where s_ρ^2 denotes 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 Holloway *et al.*, 2002). Next, we evaluate the acceptance probability

$$Pr(\rho_{old}, \rho_{new}) = \min\left[\frac{p(\rho_{new}|\vartheta_{-\rho}, y, y_0, X, W)}{p(\rho_{old}|\vartheta_{-\rho}, y, y_0, X, W)}, 1\right],$$

and finally set $\rho = \rho_{new}$ with probability $Pr(\rho_{old}, \rho_{new})$; otherwise, $\rho = \rho_{old}$. The proposal value of ρ is not truncated to satisfy the space-time stationary condition $S(\psi)$. Thus, if the proposed value of ρ is not within the condition, the conditional posterior is zero, and the proposal value is rejected with probability one. Pace and LeSage (2012) apply the RW-MH method to the sampling of other parameters.

2.2.3 Tailored approximation blocking MH algorithm

As noted above, the RW-MH, which generates a single parameter at a time, produces a highly correlated sample sequence. Moreover, the SDPD model has complicated interdependence in parameters arising from the space-time stationary condition. Liu (1994) suggests that to lump the high correlated parameters together improves the sample mixing. Thus, the solutions for estimating high correlated parameters are to jointly sample these parameters at once and approximate the full conditional distribution of the parameters ψ. Thus, we develop an estimation approach by incorporating a Tailored approximation and blocking MH algorithm (hereafter, we call this TaB-MH). The approximation and blocked sampling approach is applied in econometric models such as the stochastic volatility model (Watanabe and Omori, 2004) and the dynamic stochastic general equilibrium (DSGE) model (Chib and Ramamurthy, 2010). Our estimation algorithm is based on Chib and Ramamurthy (2010)[6].

At first, we assume the following as the prior distribution for the parameter set ψ

$$p(\psi) \sim \mathcal{N}(\psi_0, \Sigma_{\psi_0}) I[S(\psi)]$$

where $I[\cdot]$ is an indicator function that takes one if the condition in parentheses is satisfied and zero otherwise. Let $f(\psi)$ denote the full conditional distribution of ψ with

$$f(\psi|\vartheta_{-\psi}, y, y_0, X, W) \quad \propto \quad \mathscr{L}(y|\vartheta, y_0, X, W)p(\psi). \tag{7}$$

We use the MH algorithm based on a normal approximation of the posterior density because we must choose the proposal density that mimics the full conditional distribution. Applying the second order Taylor expansion to the logarithm of the density (7) around the mode $\psi = \psi^*$, we have an approximate normal

[6]The authors proposed randomized blocking for parameters because the dimension of parameters in the DSGE model is large. In the case of the SDPD model, we propose an approach without randomizing parameters as we should block the three parameters.

density as follows:

$$\begin{aligned}\ln f(\psi^*|\vartheta_{-\psi}, y, y_0, X, W) &\approx \ln f(\psi^*|\vartheta_{-\psi}, y, y_0, X, W) + (\psi - \psi^*)'a + \frac{1}{2}(\psi - \psi^*)'B(\psi - \psi^*) \\ &= \ln g(\psi^*|\vartheta_{-\psi}, y, y_0, X, W), \end{aligned} \tag{8}$$

where

$$a = \left.\frac{\partial \log f(\psi|\vartheta_{-\psi}, y, y_0, X, W)}{\partial \psi}\right|_{\psi=\psi^*}, \quad B = \left.\frac{\partial^2 \log f(\psi|\vartheta_{-\psi}, y, y_0, X, W)}{\partial \psi \partial \psi'}\right|_{\psi=\psi^*},$$

denoting the numerical gradient vector and Hessian matrix[7] of $f(\psi^*|\vartheta_{-\psi}, y, y_0, X, W)$. The mode ψ^* can be obtained by selecting in a way that maximizes the conditional posterior density $f(\psi|\vartheta_{-\psi}, y, y_0, X, W)$.

Next, we generate a candidate from the normal distribution,

$$\psi_{new} \sim \mathcal{N}(\hat{\psi}, \Sigma_\psi),$$

where $\hat{\psi} = \psi^* + \Sigma_\psi^{-1} a$ and $\Sigma_\psi = (-B)^{-1}$. Finally, we evaluate the acceptance probability

$$Pr(\psi_{old}, \psi_{new}) = \min\left[\frac{f(\psi_{new}|\vartheta_{-\psi}, y, y_0, X, W)g(\psi_{old}|\vartheta_{-\psi}, y, y_0, X, W)}{g(\psi_{new}|\vartheta_{-\psi}, y, y_0, X, W)f(\psi_{old}|\vartheta_{-\psi}, y, y_0, X, W)},\ 1\right],$$

and set $\psi = \psi_{new}$ with probability $Pr(\psi_{old}, \psi_{new})$; otherwise, $\psi = \psi_{old}$. When the $S(\psi)$ from the proposed values of ψ does not satisfy the space-time stationary condition, the conditional posterior is zero, and the proposal values are rejected with probability one.

3 Illustrative examples

3.1 Numerical experiment

We illustrate the properties of sampling approaches using simulated data. The desirable properties for sampling methods in MCMC are efficiency and well mixing, which yield fast convergence. Following previous literature (Chib and Ramamurthy, 2010; Ohtsuka and Kakamu, 2015), we also compare the efficiency of sampling methods using an inefficiency factor. The inefficiency factor is defined as $1 + 2\sum_{s=1}^{L} r_s$ where r_s is the sample autocorrelation at lag s calculated from the sampled values. The inefficiency factor 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 (Chib, 2001).

[7] Chib and Ramamurthy (2010) note that it is possible that the negative Hessian is not guaranteed. In this case, we should use the modified Cholesky decomposition in Nocedal and Wright (2000).

We now explain the data generation process. The data are simulated using the SDPD model presented in Section 2.1. First, we construct the weight matrix W as follows:

i. Generate d_{ij} for $(i > j)$ from the Bernoulli distribution with a 0.4 probability of success.

ii. Set $d_{ij} = d_{ji}$ for $i \neq j$ and $d_{ij} = 0$ for $i = j$.

iii. Compute $w_{ij} = d_{ij} / \sum_{j=1}^{n} d_{ij}$ for all i and j.

Next, for the independent variables $x_{it} = (x_{1it}, x_{2it}, x_{3it})$, we take the standard normal variates and set the X_t, which are $n \times 3$ covariate matrices. With regard to the parameters, we implement the same settings in Parent and LeSage (2012). The setting of the space-time parameters is $(\rho, \phi, \theta) = (0.9, 0.9, -0.85)$, which corresponds to a case of strong interaction in both spatial and serial parameters. Moreover, we set $T = 5$ to explore small sample properties. The other parameters are set to be $\beta' = (\beta_1, \beta_2, \beta_3) = (2, 2, 2)$, $\alpha = 2$; μ generates from $\mathcal{N}(0, \tau^2 I_n)$ with $\tau^2 = 0.05$ and $\nu = 6$. A pre-sample y_0 draws from the initial distribution $\mathcal{N}(\alpha\iota_n, \tau^2 I_n + \sigma^2 \Lambda)$.

The hyper-parameters of the prior distribution are the following

$$\beta_0 = \mathit{0}_3,\ \Sigma_{\beta_0} = 10I_3,\ \alpha_0 = 0.0,\ \sigma^2_{\alpha_0} = 10,\ \delta_{\tau_0} = 0.05,\ s_{\tau_0} = 2.0,\ \delta_{\sigma_0} = 0.05,\ s_{\sigma_0} = 2.0,$$
$$\upsilon_0 = \mathit{0}_3,\ \Sigma_{\upsilon_0} = 10I_3,\ \delta_{\nu_0} = 6.0,\ s_{\nu_0} = 1.0.$$

In the numerical experiment using the RW-MH, we set the tuning parameters $s^2_\rho = s^2_\phi = s^2_\theta = 1.0$. For each simulation, we use $10,000$ iterations for the posterior sample after a burn-in period of $5,000$ iterations. All the results in this paper were calculated using the Ox version 6.2 (see Doornik, 2006).

Table 1 summarizes the estimated results of the TaB-MH and RW-MH algorithm where Mean, SD, 95%CI, CD, and IF represent the posterior mean, standard deviation, 95% credible interval, Geweke's (1992) convergence diagnostics[8], and inefficiency factor, respectively. From Table 1, we first observed that all parameters include true values in the 95% credible intervals. However, some values of the CD in the RW-MH algorithm are less than 0.01 against the values in the TaB-MH. These results imply that the RW-MH method does not work well in the parameters of the SDPD model under this experiment. Regarding the efficiency of parameter sampling corresponding to the space and time correlations, the inefficiency factors of the RW-MH are 10 times more than those of the TaB-MH algorithm. This implies that the RW-MH method takes a greater number of simulations and more computation time than our algorithm.

[8] The convergence diagnostics represent the p-value based on the test statistic on the difference between two sample means (i.e., dividing all the generated random draws into three parts; we compute two sample means from the first 20% and last 50% of the random draws) where the test statistic is asymptotically distributed as standard normal random variables. We confirm that the random draws generated by the MCMC do not converge to the random draws generated from the target distribution when the CD is less than 0.01 (see Geweke, 1992, for a detailed discussion of the CD).

Figure 1 shows the sample draws from the posterior distribution and their autocorrelation functions (ACFs) to illustrate the differences in the mixing of the two ways. The top panel corresponding to the draws using the RW-MH algorithm and the bottom panel corresponds to the draws using the TaB-MH algorithm. From the top panel of the figure, it seems that the posterior samples using the RW-MH include persistent autocorrelations that decay slowly. In comparison, the TaB-MH algorithm performs well with autocorrelations that decay to zero rapidly. Thus, we also see that our algorithm has faster convergence than the RW-MH algorithm.

Finally, we examine the correlations among drawing parameters and discuss the reason why the convergence of the posterior samples is slow. Figure 2 shows the scatter plots of drawing from the full conditional distribution of the parameters associated with the space-time stationary condition using the TaB-MH approach. From the figure, we see that the negative correlation between ρ and θ is higher than with the other approaches. Even if the joint sampling approach is used, there seems to be a high correlation among the parameters. Then, it is not appropriate to separately sample the parameters so that the dependence between these parameters slows the convergence of the MCMC chain. Moreover, the SDPD model using the TaB-MH works with fewer simulations than the experiments in Parent and LeSage (2012)[9].

Therefore, in the case of the SDPD model with an a priori relationship among parameters, we conclude that the approach based on the blocking and approximating posterior distribution is superior to the RW-MH algorithm from the viewpoint of efficiency.

3.2 Application to regional spillovers in Japan

In this empirical illustration, we apply the SDPD model and the TaB-MH algorithm to estimate spillovers of regional production in Japan. There are some empirical studies on Japanese regional data, for example, manufacturing productivity (Kakamu *et al*., 2012), prefectural productivity (Ohtsuka, 2015), and the effect of fiscal policy (Funashima and Ohtsuka, 2019). These studies show that it is important to consider the spatial model when analyzing the regional data set in Japan.

We rely on a growth regression as in Parent and LeSage (2012). Following Parent and LeSage (2012), we consider a Cobb-Douglas production function that depends on physical capital, human capital, and labor force. The empirical specification of the prefectural growth $\dot{y}_t$ is described by the following:

$$
\begin{aligned}
\dot{y}_t &= \rho W\dot{y}_t + \rho\dot{y}_{t-1} + \theta W\dot{y}_{t-1} + \iota_n\alpha + \beta_k\dot{x}_{k,t} + \beta_l\dot{x}_{l,t} + \beta_h\dot{x}_{h,t} + \eta_t, \\
\eta_t &= \mu + \sigma\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \Lambda)
\end{aligned}
$$

[9]Their simulations were that the burn-in period was $10,000$, and the total period was $20,000$.

where $\dot{x}_{k,t}$, $\dot{x}_{l,t}$, and $\dot{x}_{h,t}$ denote the growth of physical capital, labor, and human capital, respectively. In the empirical analysis, we measure $\dot{y}_{it}$ and $\dot{x}_{l,t}$ using the prefectural product and the total number of workers in the prefecture. For the measurement of physical capital stock, we use a proxy using a prefectural gross capital formation[10]. As in Parent and LeSage (2012), we substitute a four-year enrollment per college-age population for human capital. Our data set is composed of annual observations of the 47 prefectures in Japan from financial year 2007 to 2015. We obtain all data except human capital from the Annual Report on Prefectural Accounts. These data are based on the System of National Accounts 2008 (SNA2008). These data can be retrieved from the website of Japan's Cabinet Office. The four-year enrollment per college-age population is calculated using the School Basic Survey from the Ministry of Education, Culture, Sports, Science and Technology.

Regarding the weight matrix, we use the contiguity dummy variables. Excluding the Okinawa prefecture, Japan is composed of four major islands: Hokkaido, Honshu, Shikoku, and Kyushu. Although these four islands are geographically separated, we assume that they are connected by trains and roads (see Kakamu *et al*., 2010)[11]. The hyper-parameters of the prior distribution are the following

$$\beta_0 = 0_3,\ \Sigma_{\beta_0} = 10I_3,\ \alpha_0 = 0.0,\ \sigma^2_{\alpha_0} = 10,\ \delta_{\tau_0} = 0.05,\ s_{\tau_0} = 2.0,\ \delta_{\sigma_0} = 0.05,\ s_{\sigma_0} = 2.0,$$
$$\upsilon_0 = 0_3,\ \Sigma_{\upsilon_0} = 10I_3,\ \delta_{\nu_0} = 10.0,\ s_{\nu_0} = 1.0.$$

We perform the MCMC procedures in generating $6,000$ draws in a single sample path and discard the first $3,000$ as the initial burn-in. Table 2 summarizes the estimated results using two methods. From the CD statistics, the p-values of all parameters are greater than 0.01. These results show the convergence of all parameters. First, we evaluate the sampling efficiency of the RW-MH and the TaB-MH. Compared with the inefficiency factors of space and time parameters, the values by the TaB-MH algorithm are smaller than those by the RW-MH. Hereafter, we explain the results using the TaB-MH.

Focusing on the estimated results of the spatial and serial dependence parameters, we found that the parameters except θ do not include zero in the 95% credible intervals. Since the posterior mean of the spatial dependence is positive, this implies that the prefectural growth rates of productivity are similar to neighboring areas. On the other hand, the posterior mean of serial correlation is negative. Negative autocorrelation is not generally found in economic data. However, our results resemble the literature using growth rates of Japanese macro-economic data series (Ohtsuka, 2018). When the sum of the posterior means $\rho + \phi + \theta$ is

[10]Gross capital formation is measured by total gross fixed capital formation, which is composed of resident producers' investments, deducting disposals from fixed assets during a given period, changes in inventories, and acquisitions less disposals of valuables for a prefectural economy. Capital stock, which is widely used in economic theory, is not published in Japan.

[11]For example, we consider Hokkaido to be contiguous with Honshu by the Seikan Railway Tunnel. Honshu and Shikoku are considered contiguous by the Awaji and Seto Bridges, and Kyushu is considered contiguous with Honshu by the Kanmon Tunnel and Bridge.

within unity, and the value of $\phi - (\rho - \theta)$ is more than negative one, we confirm that the process satisfies the space-time stationary condition.

The posterior means of β_k, β_l, and β_h are 0.102, 0.259, and 0.024, respectively. The credible intervals except β_h do not contain zero. Thus, we find that the effect of labor on the growth rate of production is larger than the effect of labor on the growth rate of physical capital. Regarding human capital, the posterior mean of β_h using Japanese data is different to mean in the results of US data in Parent and LeSage (2012). The assumption that an area with a higher enrollment rate has skilled workers is not applicable to Japanese data. Finally, the degree of freedom ν is around 10 and shows that the errors deviate substantially from normal, supporting our flexible modeling for heteroskedasticity across the regions of the error term.

4 Conclusion

Although the SDPD model is flexible enough to estimate the structure of spatial and temporal dependencies, there are few Bayesian estimations developed for the model. For sampling spatiotemporal parameters, empirical analyses frequently use the existing approach with individual and RM-MH algorithms. Since a single sampling for correlated parameters due to the space-time stationary condition and the RW-MH lead to slow convergence and a bad mixing of posterior simulations, researchers require an enormous number of iterations. Moreover, this computational cost is a serious problem when handling the large set of dimensional data. Thus, this work proposes an efficient MCMC algorithm for the SDPD model. To sample parameters that relate to the space and serial correlation simultaneously, we constructed a proposal density function for the MH algorithm based on the approximate normal distribution using a Taylor expansion of the logarithm of the target posterior density. We compared the two algorithms using simulated data and real data. With respect to the mixing and the efficiency of the MCMC, we found that blocked sampling the three parameters on the space-time stationary condition is preferable to the previous sampling approach.

Finally, there are some remaining issues. In this study, we focused on the approach based on blocking and the approximation of a target distribution. Recently, various approaches have been proposed, such as the Hamiltonian Monte Carlo (Wolf *et al.*, 2018) approach for spatial econometric models and the variational Bayesian inference (Gefang *et al.*, 2019) for large data analysis. This subject will be discussed in our future research. However, it is important to show how the sampling methods are efficient or inefficient We consider that, in this respect, our experiment represents a benchmark in applied works.

References

[1] Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht, The Netherlands: Kluwer.

[2] Anselin, L. (2001). Spatial econometrics. In B. H. Baltagi (Ed.), *A companion to theoretical econometrics*. Massachusetts: Blackwell Publishers Ltd.

[3] Arbia, G. (2014). *A primer for spatial econometrics with application in R*. Palgrave Macmillan.

[4] Bivand, R. S., Gómez-Rubio, V., & Rue, H. (2014). Approximate Bayesian inference for spatial econometrics models. *Spatial Statistics*, **9**, 146–165.

[5] Chib, S. (2001). Markov chain Monte Carlo methods: Computation and inference. In J. J. Heckman & E. Leamer (Eds.), *Handbook of econometrics*, **5**. (pp. 3569—3649) Amsterdam, North Holland.

[6] Chib, S., & Ramamurthy, S. (2010). Tailored randomized block MCMC methods with application to DSGE models. *Journal of Econometrics*, **155**, 19–38.

[7] Elhorst, J. P. (2014). *Spatial econometrics from cross-sectional data to spatial panels*. Springer Briefs in Regional Science.

[8] Doornik, J. A. (2006). *Ox: Object-oriented matrix programming language*. London: Timberlake Consultants Press.

[9] Funashima, Y., & Ohtsuka, Y. (2019). Spatial crowding-out and crowding-in effects of government spending on the private sector in Japan. *Regional Science and Urban Economics*, **75**, 35–48.

[10] Gefang, D., Koop, G., & Poon, A. (2019). Variational Bayesian inference in large vector autoregressions with hierarchical shrinkage. CAMA Working Papers 2019-08, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

[11] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J.M. Bernardo, J.O. Berger, A.P. Dawid, & A.F.M. Smith (eds.) *Bayesian statistics 4* (pp. 169–193). Oxford: Oxford University Press.

[12] Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London/New York: Chapman&Hall.

[13] Holloway, G., Shankar, B., & Rahmanb, S. (2002). Bayesian spatial probit estimation: A primer and an application to HYV rice adoption. *Agricultural Economics*, **27**, 383–402.

[14] Kakamu, K., Wago, H., & Tanizaki, H. (2010). Estimation of regional business cycles in Japan using Bayesian panel spatial autoregressive approach. In Nolin, T. P. (Ed.). *Handbook of regional economics* (pp. 555—571). New York: Nova Science Publishers.

[15] Kakamu, K., Polasek, W., & Wago, H. (2012). Production technology and agglomeration for Japanese prefectures during 1991–2000. *Papers in Regional Science*, **91**, 29–41.

[16] Lacombe, D. J., & McIntyre, S. G. (2016). Local and global spatial effects in hierarchical models. *Applied Economics Letters*, **23**(16), 1168–1172.

[17] Lee, L. F., & Yu, J. (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics*, **180**(2), 174–197.

[18] LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton, FL: CRC press.

[19] Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, **89**, 958–966.

[20] Mills, J A., & Parent, O. (2014). Bayesian MCMC estimation. In M. M. Fischer, & P. Nijkamp (Eds.), *Handbook of regional science*. Berlin Heidelberg: Springer-Verlag.

[21] Nocedal, J., & Wright, S. J. (2000). *Numerical optimization*, Second Edition. New York: Springer.

[22] Ohtsuka, Y. (2015). Estimation of marginal effects and their spillover effects in productivity of prefectures. *Journal of the Japan Statistical Society: Japanese Issue*, **41**(1), 41–58.

[23] Ohtsuka, Y. (2018). Large shocks and the business cycle: The effect of outlier adjustments. *Journal of Business Cycle Research*, **14**, 143–178.

[24] Ohtsuka, Y., & Kakamu, K. (2015). Comparison of the sampling efficiency in spatial autoregressive model. *Open Journal of Statistics*, **5**, 10–20.

[25] Ohtsuka, Y., & Kakamu, K. (2009). Estimation of electric demand in Japan: A Bayesian spatial autoregressive AR(p) approach. In L.V. Schwartz (ed.) *Inflation: Causes and effects* (pp. 156—178). Nova Science Publishers, Inc.

[26] Pace, R. K., & LeSage, J. P. (2009). A sampling approach to estimate the log determinant used in spatial likelihood problems. *Journal of Geographical System*, **11**, 209–225.

[27] Parent, O., & LeSage, J P. (2011). A space-time filter for panel data models containing random effects. *Computational Statistics and Data Analysis*, **55**, 475–490.

[28] Parent, O., & J P. LeSage (2012). Spatial dynamic panel data models with random effects. *Regional Science and Urban Economics*, **42**, 727–738.

[29] Smirnov, O., & Anselin, L. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach. *Computational Statistics & Data Analysis*, **35**, 301–319.

[30] Stakhovych, S., & Bijmolt, T. H. A. (2009). Specification of spatial models: A simulation study on weights matrices. *Papers in Regional Science*, **88**, 389–408.

[31] Su, L., & Yang, Z. (2015). QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics*, **185**(1), 230–258.

[32] Watanabe, T. (2001). On sampling the degree-of-freedom of student's-t disturbances. *Statistics & Probability Letters*, **52**, 177–181.

[33] Watanabe, T., & Omori, Y. (2004). A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard & Pitt (1997). *Biometrika*, **91**(1), 246–248.

[34] Wolf, L. J., Anselin, L., & Arribas-Bel, D. (2018). Stochastic efficiency of Bayesian Markov chain Monte Carlo in spatial econometric models: An empirical comparison of exact sampling methods. *Geographical Analysis*, **50**, 97–119.

[35] Yu, J., de Jong, R., & Lee, L. F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when n and T are large. *Journal of Econometrics*, **146**(1), 118–134.

A Sampling procedure for the parameters except space-time correlation ψ

First, we assume the following prior distributions,

$$\beta \sim \mathcal{N}(\beta_0, \Sigma_{\beta_0}), \quad \alpha \sim \mathcal{N}(\alpha_0, \sigma^2_{\alpha 0}), \quad \tau^2 \sim \mathcal{IG}(\delta_{\tau 0}/2, s_{\tau 0}/2),$$
$$\sigma^2 \sim \mathcal{IG}(\delta_{\sigma 0}/2, s_{\sigma 0}/2), \quad \nu \sim \mathcal{G}(\delta_{\nu 0}, s_{\nu 0}) I[\nu > 2]$$

where $\mathcal{G}$ denotes a gamma distribution, and we assume the truncated gamma distribution to satisfy a finite variance.

Let $y_\beta = \{\iota_T \otimes (I_n - \rho W)\}y - \alpha \iota_{nT}$. The full conditional distribution of β and σ^2 are given by

$$\beta | \vartheta_{-\beta}, y, y_0, X, W \sim \mathcal{N}(\hat{\mu}_\beta, \hat{\Sigma}_\beta), \quad \sigma^2 | \vartheta_{-\sigma^2}, y, y_0, X, W \sim \mathcal{IG}(\hat{\delta}_\sigma/2,\ \hat{s}_\sigma/2),$$

where $\hat{\mu}_\beta = \hat{\Sigma}_\beta(X'\Omega^{-1}y_\beta + \Sigma^{-1}_{\beta_0}\beta_0)$, $\hat{\Sigma}_\beta = (X'\Omega^{-1}X + \Sigma^{-1}_{\beta_0})^{-1}$, $\hat{\delta}_\sigma = e'(I_T \otimes \Lambda)e + \hat{\delta}_{\sigma 0}$ and $\hat{s}_\sigma = nT + s_{\sigma 0}$.

Next, we focus on sampling procedures for μ and τ^2. Define μ_i and x_i for $i = 1, \ldots, n$ as i the component of the vector μ and X_i, respectively. Let $y_i = (y_{i1}, \ldots, y_{iT})'$ and $X_i = (x'_{i1}, \ldots, x'_{iT})'$. The model (1) and (2) becomes

$$\bar{y}_i = \mu_i \iota_T + \sigma \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \lambda_i I_T)$$

where $\bar{y}_i = (\bar{y}_{i1}, \ldots, \bar{y}_{iT})'$, $\bar{y}_{it} = y_{it} - \rho \sum_{j=1}^n w_{ij} y_{jt} - \phi y_{i,t-1} - \theta \sum_{j=1}^n w_{ij} y_{j,t-1} - \alpha - x_{it}\beta$, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$. The full conditional distribution of μ_i and τ^2 are written as

$$\mu_i | \vartheta_{-\mu_i}, y, y_0, X, W \sim \mathcal{N}(\hat{\mu}_i, \hat{\sigma}^2_{\mu_i}), \quad \tau^2 | \vartheta_{-\tau^2} \sim \mathcal{IG}(\hat{\delta}_\tau/2, \hat{s}_\tau/2),$$

where $\hat{\mu}_i = \hat{\sigma}^2_{\mu_i}(\sigma^{-2}\lambda_i^{-1}\bar{y}'_i \iota_T)$ and $\hat{\sigma}^2_{\mu_i} = (\sigma^{-2}\lambda_i^{-1}T + \tau^{-2})^{-1}$, $\hat{\delta}_\tau = \mu'\mu + \delta_{\tau 0}$ and $\hat{s}_\tau = n + s_{\tau 0}$.

For the constant term α, let $y_\alpha = \{\iota_T \otimes (I_n - \rho W)\}y - \{\iota_T \otimes (\phi I_n + \theta W)\}y_- - X\beta$, and the full conditional distribution of α is given as

$$\alpha | \vartheta_{-\alpha}, y_0, y, W, X \sim \mathcal{N}(\hat{\alpha}, \hat{\sigma}^2_\alpha),$$

where $\hat{\alpha} = \hat{\sigma}^2_\alpha(\iota'_{nT}\Omega^{-1}y_\alpha + \sigma^{-2}_{\alpha 0}\alpha_0)$ and $\hat{\sigma}^2_\alpha = (\iota'_{nT}\Omega^{-1}\iota_{nT} + \sigma^{-2}_{\alpha 0})^{-1}$.

For the parameters of heteroskedasticity, we use the Gibbs sampler and acceptance rejection-MH (AR-MH) to generate λ and ν, respectively. Define $e_i = (e_{i1}, \ldots, e_{iT})'$ and $e_{it} = y_{it} - \rho \sum_{j=1}^n w_{ij} y_{jt} - \phi y_{i,t-1} - \theta \sum_{j=1}^n w_{ij} y_{j,t-1} - \alpha - \mu_i - x_{it}\beta$. Conditional on ν, the $\hat{\lambda}_i = (e'_i e_i + \nu - 2)/\lambda_i$ follows an independent

$\chi^2(\nu+T)$ distribution. Hence, λ_i^* is sampled from $\chi^2(\nu+T)$ distribution. We calculate λ_i using λ_i^* and $\hat{\lambda}_i$. Finally, the log of the full conditional distribution of ν is given by

$$\ln p(\nu|\vartheta_{-\nu}, y_0, y, W, X) \propto \frac{T\nu}{2}\ln\left(\frac{\nu-2}{2}\right) - T\ln\Gamma\left(\frac{\nu}{2}\right) - \xi\nu + (\delta_{\nu_0})\ln(\nu),$$

where

$$\xi = \frac{1}{2}\sum_{i=1}^{n}\left\{\ln(\lambda_i) + \frac{1}{\lambda_i}\right\} + s_{\nu_0}.$$

It is difficult to generate the parameter from the distribution. Thus, we employ the AR-MH algorithm extended by Watanabe (2001). This algorithm generates the parameter from a proposal density by applying the second-order Taylor expansion to the log of $p(\nu|\vartheta_{-\nu}, y_0, y, W, X)$ and evaluating the sampled parameter with AR and MH steps. See the detail of AR-MH algorithm for the degree of freedom in Watanabe (2001).

Table 1: Posterior summary of the simulated data set ($T = 5$, $n = 50$)

Parameter	True value	Mean	SD	95%CI	CD	IF
		RW-MH algorithm				
ρ	0.9	0.892	0.01	$[0.872, 0.913]$	0.00	121.66
ϕ	0.9	0.886	0.01	$[0.864, 0.911]$	0.00	220.25
θ	-0.85	-0.828	0.02	$[-0.859, -0.803]$	0.01	241.65
σ^2	1.0	1.030	0.16	$[0.748, 1.395]$	0.52	5.00
τ^2	0.05	0.043	0.04	$[0.008, 0.138]$	0.69	25.17
α	2.0	2.037	0.30	$[1.444, 2.652]$	0.04	49.81
β_1	2.0	2.043	0.06	$[1.939, 2.153]$	0.37	3.44
β_2	2.0	1.935	0.06	$[1.809, 2.054]$	0.21	4.61
β_3	2.0	2.045	0.06	$[1.936, 2.159]$	0.27	3.20
ν	6.0	5.049	0.93	$[4.034, 7.436]$	0.08	16.92
		TaB-MH algorithm				
ρ	0.9	0.892	0.01	$[0.863, 0.918]$	0.72	11.93
ϕ	0.9	0.893	0.01	$[0.867, 0.919]$	0.41	1.68
θ	-0.85	-0.836	0.02	$[-0.869, -0.802]$	0.32	1.33
σ^2	1.0	1.039	0.17	$[0.755, 1.404]$	0.90	3.62
τ^2	0.05	0.047	0.04	$[0.008, 0.143]$	0.62	22.14
α	2.0	2.083	0.29	$[1.528, 2.730]$	0.55	29.63
β_1	2.0	2.046	0.06	$[1.936, 2.155]$	0.88	2.04
β_2	2.0	1.939	0.06	$[1.809, 2.063]$	0.56	1.91
β_3	2.0	2.042	0.06	$[1.929, 2.155]$	0.01	3.17
ν	6.0	4.987	0.88	$[4.033, 7.197]$	0.46	4.28

Note: CI, CD, and IF are the credible interval, the p-value of the convergence diagnostic (CD) test in Geweke (1992), and the inefficiency factor, respectively. The acceptance rates are as follows: around 50% in the RW-MH, and around 99% in the TaB-MH.

Table 2: Estimation results of the empirical data set

Parameter	Mean	SD	95%CI	CD	IF
	RW-MH algorithm				
ρ	0.518	0.04	$[0.433, 0.602]$	0.12	7.12
ϕ	-0.148	0.05	$[-0.254, -0.050]$	0.45	10.07
θ	0.044	0.06	$[-0.078, 0.177]$	0.75	11.80
σ^2	5.424	0.60	$[4.389, 6.740]$	0.89	9.48
τ^2	0.032	0.03	$[0.006, 0.113]$	0.74	22.69
α	0.107	0.12	$[-0.141, 0.341]$	0.66	1.43
β_0	0.102	0.02	$[0.071, 0.135]$	0.09	5.07
β_1	0.261	0.11	$[0.044, 0.470]$	0.54	1.06
β_2	0.025	0.03	$[-0.026, 0.076]$	0.92	1.03
ν	11.81	3.66	$[6.231, 20.290]$	0.59	12.90
	TaB-MH algorithm				
ρ	0.520	0.04	$[0.437, 0.605]$	0.27	2.73
ϕ	-0.142	0.06	$[-0.260, -0.028]$	0.78	1.02
θ	0.039	0.07	$[-0.091, 0.171]$	0.92	1.51
σ^2	5.386	0.62	$[4.361, 6.765]$	0.56	3.89
τ^2	0.037	0.03	$[0.007, 0.120]$	0.12	17.23
α	0.111	0.13	$[-0.138, 0.363]$	0.58	1.18
β_k	0.102	0.02	$[0.071, 0.134]$	0.40	1.98
β_l	0.259	0.11	$[0.041, 0.475]$	0.63	1.02
β_h	0.024	0.03	$[-0.028, 0.076]$	0.36	1.22
ν	11.95	3.62	$[6.068, 20.00]$	0.73	7.74

Note: CI, CD, and IF are the credible interval, the p-value of the convergence diagnostic (CD) test in Geweke (1992), and the inefficiency factor, respectively.

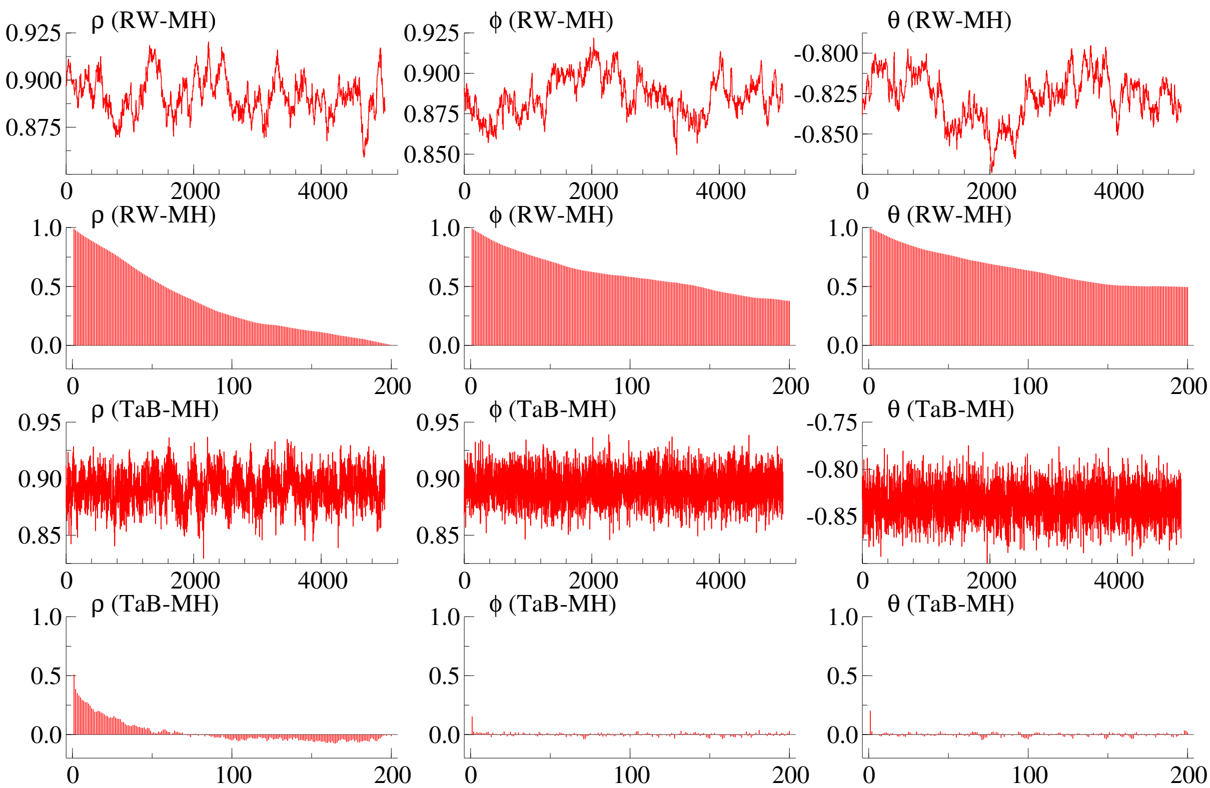

Figure 1: The simulation results of the autocorrelation and posterior sample

Note: The time-series plots of draws from the posterior and corresponding auto-correlation functions using the RW-MH (top panel) and the TaR-MH (bottom panel).

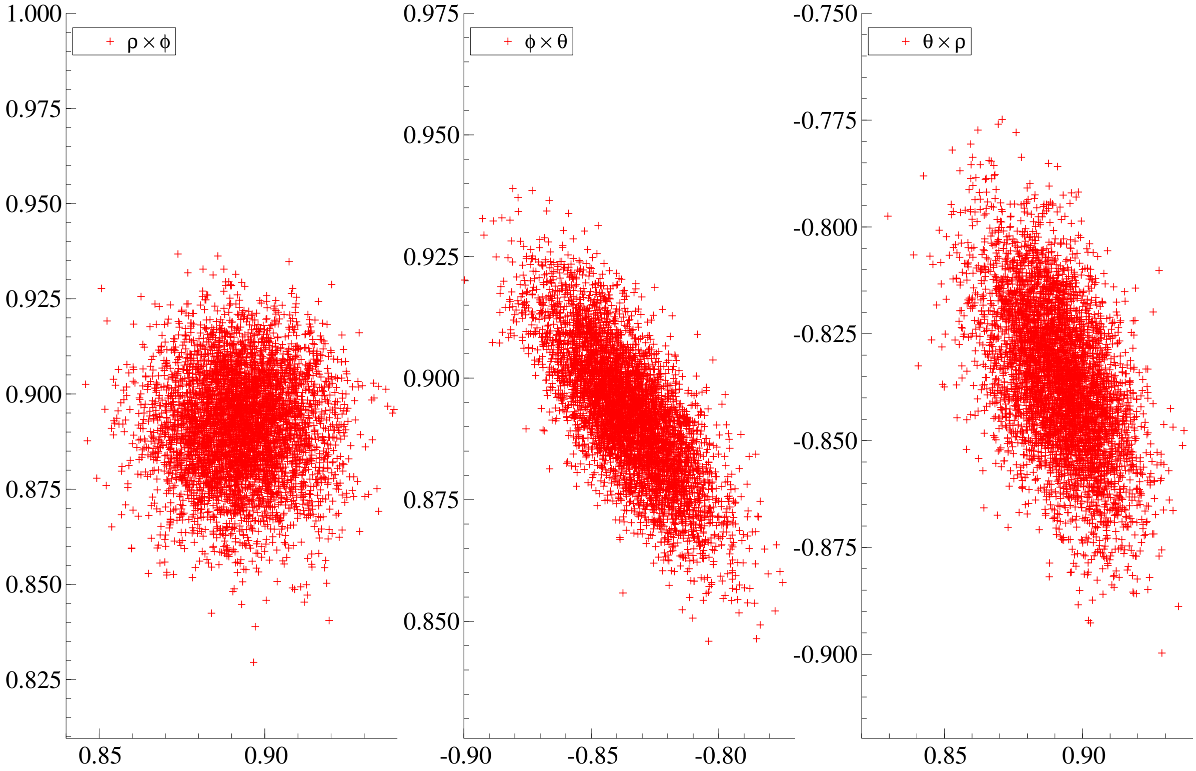

Figure 2: Scatter plots of space-time parameters