Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparison of the blocked sampling approaches for the spatial dynamic panel data model

Yoshihiro Ohtsuka

ohtsuka@mail.tohoku-gakuin.ac.jp

Tohoku Gakuin University

Research Article

Keywords: Markov chain Monte Carlo method, No-U-Turn sampler, Spatial dynamic panel data model, Taylor approximation, and blocked Metropolis-Hastings algorithm

Posted Date: May 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9600276/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Comparison of the blocked sampling approaches for the spatial dynamic panel data model

Abstract

This study proposes a posterior sampling algorithm for a spatial dynamic panel data model using Bayesian inference. The model's stationarity depends on three parameters: simultaneous and lagged spatial dependencies, and serial correlation. Previous approaches sampled interdependent parameters individually using the random walk Metropolis-Hastings (RW-MH) algorithm, leading to high autocorrelation in the posterior samples and necessitating extensive simulations. To improve sampling efficiency, we developed a Bayesian estimation algorithm for these parameters that employs a Taylor approximation and a blocked MH (TaB-MH) algorithm. The TaB-MH algorithm demonstrates superior performance compared to the RW-MH algorithm and the No-U-Turn sampler, as shown by both simulated and empirical data.

Keywords: Markov chain Monte Carlo method, No-U-Turn sampler, Spatial dynamic panel data model, Taylor approximation, and blocked Metropolis-Hastings algorithm

JEL Classification: C4 , C33 , R15

Highlights.

- We examine the sampling efficiency for the spatial dynamic panel data model.
- This study develops an estimation algorithm based on a Taylor approximation and blocked sampling.
- Illustrative examples show that the proposed method is superior to the other algorithm.

Funding.

- No funding was received for conducting this study.

1 Introduction

Recent regional studies increasingly focus on specifying and estimating the spatial relationships between regions and temporal changes within them. Spatial econometrics originated from the analysis of cross-sectional data (Anselin 1988, 2001; Arbia 2006). With the proliferation of panel data analysis, spatial econometric models have been extended to spatial dynamic panel data (SDPD) models, which account for both spatial and temporal dependencies (LeSage and Pace 2009; Elhorst 2014). Because the SDPD model accommodates spatiotemporal interdependencies in the dependent variable—such as simultaneous and lagged spatial correlations, as well as serial correlation—it offers the advantage of evaluating how observations relate across both space and time. Consequently, the SDPD model is highly applicable to research in regional economies, industry, land prices, and transportation.

Three primary approaches have been developed for estimating SDPD models: the quasi-maximum likelihood method (Su and Yang 2015), the generalized method of moments (Lee and Yu 2014), and Bayesian inference (Parent and LeSage 2011, 2012). Following the seminal work of Anselin (1988) and LeSage and Pace (2009), we employ a Markov chain Monte Carlo (MCMC) method grounded in Bayesian inference. The MCMC method combines random sampling from the full conditional distribution, based on the Markov chain, with Monte Carlo approximation. We note that, in regional economic analysis using the prefectural data in Japan, the sample size of the time series is typically small. The MCMC method does not rely on the usual asymptotic properties, such as asymptotic normality and asymptotic efficiency, when evaluating the posterior distribution on the data. Thus, it is reasonable to utilize Bayesian inference. Given the characteristics of our data, we focus on the SDPD model with random effects rather than fixed effects. Furthermore, we propose an extension to the SDPD model that incorporates random effects, as in Parent and LeSage (2012), and introduces heteroskedasticity, as outlined in Mills and Parent (2019).

In Bayesian estimation of the SDPD model, a common approach is to sample three parameters individually using the random-walk Metropolis-Hastings (RW-MH) algorithm. However, this method is inefficient due to parameter correlation and autocorrelation within the posterior sample. As Parent and LeSage (2011, 2012) demonstrated, the stationarity condition for the SDPD model depends on three parameters that capture spatial and temporal dependence. Consequently, individually sampling parameters to satisfy the stationarity condition leads to high correlations because each parameter's value depends on the other two. Furthermore, the RW-MH algorithm, while simple, generates parameters from previously sampled values, leading to high autocorrelation in the posterior sample series. These issues impair Markov chain mixing and slow convergence to the target distribution, thereby requiring more simulations and incurring significantly higher computational costs. This necessitates the development of improved estimation methods.

Regarding sampling efficiency for spatial econometric models, some studies have examined the development of sampling methods, including griddy Gibbs sampling (Ohtsuka and Kakamu 2015), the Laplace approximation approach (Bivand et al. 2014), and the Hamiltonian Monte Carlo method (Wolf et al. 2018). The aforementioned studies focused on models without temporal structure, such as spatial

autoregressive models and spatial error models. They proposed methodologies for estimating only simultaneous spatial correlation. To date, no research has extended the sampling method to the three parameters of the SDPD model. In this study, we examine the following solutions to the two problems mentioned above. The first is blocked sampling of parameters. Liu (1994) mentioned that grouping highly correlated parameters into a single block improves sampling efficiency. Thus, we devise a strategy for jointly sampling three parameters related to spatiotemporal dependence. The second is the application of an alternative MH algorithm. Specifically, the independent-chain MH method is employed, in which the proposal distribution is defined by a Taylor approximation to the full conditional distribution of these parameters. This approach has been used in econometric analysis, such as a stochastic volatility model (Watanabe and Omori 2004) and a dynamic stochastic general equilibrium model (Chib and Ramamurthy 2010). In this work, we propose the Taylor approximation and blocking MH (hereafter, TaB-MH) algorithm, which combines two strategies. Moreover, we employ the No-U-Turn sampler (NUTS) method as the third approach. This method is based on the Hamiltonian Monte Carlo method, which has been used since its proposal in Duane et al. (1987). It has been incorporated into Bayesian analysis packages for the programming languages R and Python, such as Stan and PyMC. The NUTS method has advantages, including automatic adjustment of the step-size and number-of-steps parameters.

In illustrative examples, the properties of the algorithms are first demonstrated using simulated data, with two datasets and several parameter values for space and time dependence. The results indicate that the TaB-MH algorithm is selected based on inefficiency. When the RW-MH algorithm is extended to blocked sampling, no improvement in estimation efficiency is observed, and the parameters do not converge. Despite the NUTS method demonstrating higher estimation efficiency than the RW-MH algorithm, it is less efficient than the TaB-MH algorithm. Furthermore, as part of our empirical analysis, we apply the SDPD model to estimate the effects of fiscal policy using prefectural panel data, following Funashima and Ohtsuka (2019). The estimation results indicate that private consumption exhibits stronger spatial correlation than serial correlation, and private investment does not. For private consumption, the TaB-MH algorithm has been shown to be a practical method, owing to its excellent properties in both mixing quality and computational efficiency. In addition, the marginal effects of government expenditures and government consumption, used as explanatory variables, were estimated, and both direct and indirect effects were shown to inhibit private-sector activity.

The remainder of this paper is organized as follows. Section 2 introduces the SDPD model and MCMC methods. Section 3 demonstrates the Monte Carlo experiments using simulated data and applied work using Japanese prefectural panel data. Finally, we provide concluding remarks.

2 Econometric model and estimation procedure

We first introduce the spatial dynamic panel data model[1] with heteroskedasticity among regions in Mills and Parent (2019). Let $\mathbf{y}_t$ define the $n \times 1$ vector of observations at time t, and $\mathbf{X}_t$ denotes the $n \times k$ matrix of exogenous variables for $t = 1, \ldots, T$. The dynamic spatial panel model is defined as follows:

$$\mathbf{y}_t = \rho \mathbf{W}\mathbf{y}_t + \phi \mathbf{y}_{t-1} + \theta \mathbf{W}\mathbf{y}_{t-1} + \alpha \boldsymbol{\iota}_n + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\eta}_t, \tag{1}$$

$$\boldsymbol{\eta}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \sigma^2 \boldsymbol{\Lambda}), \quad \boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_n), \tag{2}$$

where ρ captures the degree of simultaneous spatial interaction, the parameters of the strength of the first-order serial and spatial dependence are denoted by the scalars ϕ and θ. $\mathbf{W}$ 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[2]. $\boldsymbol{\iota}_n$ is the $n \times 1$ vector of ones associated with the constant term α.

With respect to the vector of the constant term $\boldsymbol{\mu}$ in the equation (2), the $n \times 1$ vector $\boldsymbol{\mu}$ is assumed to be the random effect. Thus, $\boldsymbol{\mu}$ is assumed to be the following normal distribution, such as $\boldsymbol{\mu} \sim \mathcal{N}(0, \tau^2)$ 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 zero and variance $\sigma^2 \Lambda$. λ_i denotes the auxiliary parameter that captures heteroskedasticity in the error distribution across regions. As Mills and Parent (2019) notes, extending the model to include the regional variance-scaling parameter λ_i leads to slow convergence due to the identification problem between the variance σ^2 and Λ. Thus, we assume that the error term $\boldsymbol{\epsilon}_t$ follows the standardized Student's t-distribution. The heteroskedasticity variance scale is assumed to be independent and identically distributed, following a 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 $\mathbf{y}_0$ 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 stationarity condition of the SDPD model. The stationary condition is more complex than that of the conventional dynamic panel data model because it allows for spatial interaction at both simultaneous and serial dynamics. Let us define the stationarity region like $\{\psi : \ \psi \in S(\psi)\}$, and Parent and LeSage (2011,

[1] This work discusses spatial autoregressive dynamic panel data models. Refer to Parent and LeSage (2011) and Mills and Parent (2019) for other model types, such as the spatial error model and the spatial Durbin model.

[2] Following Stakhovych and Bijmolt (2009), who recommend using the contiguity dummy in their numerical experiments, we adopt the same approach.

2012) present the stationarity 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 $\boldsymbol{\psi} = (\rho, \phi, \theta)'$ and $\varpi_{\min}$ ($\varpi_{\max}$) is the minimum (maximum) eigenvalue of $\mathbf{W}$. The stationary condition of the SDPD model, as shown above, depends on three parameters: simultaneous spatial dependence, lagged spatial dependence, and serial clustering. Thus, the space-time stationarity constraints must be imposed on the parameters ρ, ϕ, and θ to estimate the model, consistent with the condition (3).

To simplify the notations, let $\boldsymbol{\vartheta} = (\boldsymbol{\psi}, \alpha, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \tau^2, \lambda, \nu)$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$, $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$, $\mathbf{y}_- = (\mathbf{y}_0', \ldots, \mathbf{y}_{T-1}')'$, and $\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_T')'$. Therefore, the log likelihood function is the following:

$$\ln p(\mathbf{y}|\boldsymbol{\vartheta}, \mathbf{y}_0, \mathbf{X}, \mathbf{W}) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Omega}| + T\sum_{j=1}^{n}\ln(1-\rho\varpi_j) - \frac{1}{2}\mathbf{e}'\boldsymbol{\Omega}^{-1}\mathbf{e}, \tag{4}$$

where $\mathbf{e} = \{\boldsymbol{\iota}_T \otimes (\mathbf{I}_n - \rho\mathbf{W})\}\mathbf{y} - \{\boldsymbol{\iota}_T \otimes (\phi\mathbf{I}_n + \theta\mathbf{W})\}\mathbf{y}_- - \boldsymbol{\iota}_{nT}\alpha - \mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{\Omega} = \tau^2(\mathbf{J}_T \otimes \mathbf{I}_n) + \sigma^2(\mathbf{I}_T \otimes \boldsymbol{\Lambda})$, $\mathbf{J}_T = \boldsymbol{\iota}_T\boldsymbol{\iota}_T'$, and $\otimes$ denotes the Kronecker product.

2.1 Bayesian analysis

2.1.1 Markov chain Monte Carlo method

Our estimation procedure relies on Bayesian inference, such as MCMC. Empirical studies using the SDPD model are mostly based on the QML estimation via the concentrated log-likelihood approach (Yu et al. 2008; Su and Yang 2015)[3]. The QML method relies on asymptotic normalities, whereas the MCMC method does not. The MCMC is employed to evaluate the posterior distributions of the parameters, with the understanding that these are conditioned on the data. Furthermore, the number of time series for regional economic data is limited. Consequently, it is reasonable to estimate the SDPD model using MCMC.

Since we adopt Bayesian inference, we evaluate the posterior distribution given the prior distribution and the likelihood. The joint posterior distribution is given by

$$p(\boldsymbol{\vartheta}|\mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W}) \propto p(\boldsymbol{\vartheta})p(\mathbf{y}|\boldsymbol{\vartheta}, \mathbf{y}_0, \mathbf{X}, \mathbf{W}), \tag{5}$$

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

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

follows:

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

We assume that the prior distributions for the parameters (α, $\boldsymbol{\beta}$, τ^2, σ^2, $\boldsymbol{\psi}$, and ν) are all independent. Since the joint posterior distribution is given by (5), we can now use MCMC. This approach requires multiple iterations to evaluate the marginal posterior distribution of a complex joint posterior distribution. Evaluating the marginal posterior distribution is analytically difficult when the joint posterior distribution is complicated. Then, we draw parameters from the full conditional distributions via Markov sampling and Monte Carlo integration to approximate the full conditional. For the SDPD model, we implement the MCMC algorithm, which involves the following steps

Step 0. Initial ϑ
Step 1: Generate β from $p(\beta|\vartheta_{-\beta}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$
Step 2: Generate σ^2 from $p(\sigma^2|\vartheta_{-\sigma^2}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$
Step 3: Generate μ from $p(\mu|\vartheta_{-\mu}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$
Step 4: Generate τ^2 from $p(\tau^2|\vartheta_{-\tau^2}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$
Step 5: Generate α from $p(\alpha|\vartheta_{-\alpha}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$
Step 6: Generate ψ from $p(\psi|\vartheta_{-\psi}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$
Step 7: Generate $\boldsymbol{\lambda}$ from $p(\boldsymbol{\lambda}|\vartheta_{-\boldsymbol{\lambda}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$
Step 8: Generate ν from $p(\nu|\vartheta_{-\nu}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$
Step 9: Repeat Step 1:. to Step 8:.

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

The Gibbs sampling method is a powerful technique widely employed 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 $\boldsymbol{\psi}$ in step Step 6:. The RW-MH algorithm, which creates only a single parameter at a time, is widely used for estimating these parameters. This method produces highly correlated sample sequences between parameters and exhibits slow convergence to the target marginal distribution. Therefore, to acquire independent posterior samples, the sampling process must be repeated for a substantial number of iterations. In this study, we propose a more efficient estimation method and, in the following subsections, explain the sampling algorithm for the SDPD model. The sampling methods for the other parameters are summarized in the Supplementary Material.

2.1.2 Random walk Metropolis-Hastings algorithm

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

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

The individual prior distribution (ρ, ϕ, θ) is assumed as following uniform distribution:

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

Moreover, they assumed that the largest (smallest) eigenvalue is 1 (-1). Thus, the joint prior distribution is uniform and equal to 0.5 over the parameter space defined as in the stationarity region (3). The full conditional distribution of ρ is as follows:

$$p(\rho|\boldsymbol{\vartheta}_{-\rho}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W}) \propto |I_n - \rho\mathbf{W}|^T \exp\left(-\frac{1}{2}\mathbf{e}'\boldsymbol{\Omega}^{-1}\mathbf{e}\right). \tag{6}$$

When directly sampling a parameter is difficult, we employ the MH algorithm with a 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 (Holloway et al. 2002). Next, we evaluate the acceptance probability

$$Pr_1(\rho_{old}, \rho_{new}) = \min\left[\frac{p(\rho_{new}|\boldsymbol{\vartheta}_{-\rho}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})}{p(\rho_{old}|\boldsymbol{\vartheta}_{-\rho}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})}, 1\right],$$

and finally set $\rho = \rho_{new}$ with probability $Pr_1(\rho_{old}, \rho_{new})$; otherwise, $\rho = \rho_{old}$. The most beneficial aspect of this methodology is its ease of implementation. However, because random numbers are generated from previously sampled values, the posterior samples exhibit high autocorrelation.

In addition, because the prior distributions of the parameters are constrained by the values of other parameters, correlations among the parameters arise. Liu (1994) suggests that lumping the highly correlated parameters together improves the sample mixing. In this study, we investigate blocking multiple parameters and therefore introduce a blocking method for the aforementioned RW-MH method. We assume the prior distribution for the parameters ψ as the truncated multivariate normal distribution:

$$\boldsymbol{\psi} \sim \mathcal{N}(\boldsymbol{\psi}_0, \boldsymbol{\Sigma}_{\boldsymbol{\psi}_0}) I[S(\boldsymbol{\psi})], \tag{7}$$

where $\boldsymbol{\psi}_0$ and $\boldsymbol{\Sigma}_{\psi_0}$ denote the hyperparameters. $I[\cdot]$ is an indicator function that takes one if the condition in parentheses is satisfied and zero otherwise. Using this prior

distribution and the likelihood function, the full conditional posterior distribution of the parameter $\boldsymbol{\psi}$ is derived. The sampling of $\boldsymbol{\psi}$ is generated using the aforementioned RW-MH method as follows:

$$\boldsymbol{\psi}_{new} \sim \mathcal{N}(\boldsymbol{\psi}_{old}, s^2_{\boldsymbol{\psi}}),$$

where $s^2_{\boldsymbol{\psi}}$ denote the tuning parameters. Next, we evaluate the acceptance probability

$$Pr_2(\boldsymbol{\psi}_{old}, \boldsymbol{\psi}_{new}) = \min\left[\frac{p(\boldsymbol{\psi}_{new}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})}{p(\boldsymbol{\psi}_{old}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})}, 1\right],$$

and finally set $\boldsymbol{\psi} = \boldsymbol{\psi}_{new}$ with probability $Pr_2(\boldsymbol{\psi}_{old}, \boldsymbol{\psi}_{new})$; otherwise, $\boldsymbol{\psi} = \boldsymbol{\psi}_{old}$. The proposal value of $\boldsymbol{\psi}$ is not truncated to satisfy the space-time stationarity condition $S(\boldsymbol{\psi})$. Thus, if the proposed value of $\boldsymbol{\psi}$ is not within the condition, the conditional posterior is zero, and the proposal value is rejected with probability one.

2.1.3 Taylor approximation blocking MH algorithm

In addition to the RW-MH method, another frequently employed approach for constructing proposal distributions is the independence MH algorithm, which uses proposal distributions that are independent of previously sampled values. In the independence MH algorithm, the full conditional posterior distribution is often approximated by a normal distribution. Thus, we develop an estimation approach for the SDPD model by incorporating a tailored approximation and blocking MH algorithm (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 model (Chib and Ramamurthy 2010). Our estimation algorithm is based on Chib and Ramamurthy (2010)[4].

At first, we assume the prior distribution of (7). Let $f(\boldsymbol{\psi})$ denote the full conditional distribution of $\boldsymbol{\psi}$ with

$$p(\boldsymbol{\psi}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W}) \propto p(\mathbf{y}|\boldsymbol{\vartheta}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})p(\boldsymbol{\psi}). \tag{8}$$

Next, applying the second-order Taylor expansion to the logarithm of the density (8) around the mode $\boldsymbol{\psi} = \boldsymbol{\psi}^*$, we have an approximate density as follows:

$$\begin{aligned}\ln p(\boldsymbol{\psi}^*|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W}) \approx\ & \ln p(\boldsymbol{\psi}^*|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W}) \\ & + (\boldsymbol{\psi} - \boldsymbol{\psi}^*)'\nabla p(\boldsymbol{\psi}^*) + \frac{1}{2}(\boldsymbol{\psi} - \boldsymbol{\psi}^*)'\nabla^2 p(\boldsymbol{\psi}^*)(\boldsymbol{\psi} - \boldsymbol{\psi}^*) \\ =\ & \ln g(\boldsymbol{\psi}^*|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W}),\end{aligned} \tag{9}$$

[4]The authors proposed randomized blocking for parameters because the parameter dimension in the DSGE model is large. For the SDPD model, we propose an approach that does not randomize parameters, since we should block the three parameters.

where

$$\nabla p(\boldsymbol{\psi}^*) = \left.\frac{\partial \ln p(\boldsymbol{\psi}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}=\boldsymbol{\psi}^*},$$
$$\nabla^2 p(\boldsymbol{\psi}^*) = \left.\frac{\partial^2 \ln p(\boldsymbol{\psi}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}=\boldsymbol{\psi}^*},$$

denoting the numerical gradient vector and Hessian matrix[5] of $p(\boldsymbol{\psi}^*|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$. The mode $\boldsymbol{\psi}^*$ can be obtained by selecting in a way that maximizes the conditional posterior density $p(\boldsymbol{\psi}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$.

Thus, we generate a candidate value $\boldsymbol{\psi}_{new}$ from a multivariate student's-t distribution $T(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\psi}}, \nu_{\boldsymbol{\psi}})$ where $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi}^*$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\psi}} = (-\nabla^2 p(\boldsymbol{\psi}^*))^{-1}$ and $\nu_{\boldsymbol{\psi}}$ denotes degrees of freedom[6]. Finally, we evaluate the acceptance probability

$$Pr_3(\boldsymbol{\psi}_{old}, \boldsymbol{\psi}_{new}) = \min\left[\frac{p(\boldsymbol{\psi}_{new}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W}) g(\boldsymbol{\psi}_{old}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})}{g(\boldsymbol{\psi}_{new}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W}) p(\boldsymbol{\psi}_{old}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})}, \ 1\right],$$

and set $\boldsymbol{\psi} = \boldsymbol{\psi}_{new}$ with probability $Pr_3(\boldsymbol{\psi}_{old}, \boldsymbol{\psi}_{new})$; otherwise, $\boldsymbol{\psi} = \boldsymbol{\psi}_{old}$. When the $S(\boldsymbol{\psi})$ from the proposed values of $\boldsymbol{\psi}$ does not satisfy the space-time stationary condition, the conditional posterior is zero, and the proposal values are rejected with probability one.

2.1.4 Hamiltonian Monte Carlo method

We use the No-U-Turn Sampler (NUTS) based on the Hamiltonian Monte Carlo (HMC) method as the third sampling strategy. Since the seminal work in Duane et al. (1987), the HMC has been developed for estimating hierarchical and latent variable models. HMC employs a sampling method from a target distribution to simulate Hamiltonian dynamics. In spatial econometric analysis, Wolf et al. (2018) applied the HMC method to a spatial autoregressive model and demonstrated its effectiveness.

In this work, we explain the sampling scheme of $p(\boldsymbol{\psi}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$. The HMC introduces auxiliary momentum vector $\mathbf{r}$, which is set to be the dimension of the interest parameters $\boldsymbol{\psi}$, and defines a Hamiltonian

$$H(\boldsymbol{\psi}, \mathbf{r}) = U(\boldsymbol{\psi}) + \frac{1}{2}\mathbf{r}'\mathbf{r}, \tag{10}$$

where $U(\boldsymbol{\psi}) = -\ln p(\boldsymbol{\psi}|\boldsymbol{\vartheta}_{-\boldsymbol{\psi}}, \mathbf{y}, \mathbf{y}_0, \mathbf{X}, \mathbf{W})$ denotes the logarithm of the full conditional posterior distribution of $\boldsymbol{\psi}$, and the distribution of momentum vector $\mathbf{r}$ is multivariate normal distribution. The function $U(\boldsymbol{\psi})$ is called potential energy, and the second term on the right side of equation (10) is called kinetic energy. The joint

[5] Chib and Ramamurthy (2010) notes that the negative Hessian need not be guaranteed. In this case, we should use the modified Cholesky decomposition in Nocedal and Wright (2000).

[6] As in Chib and Ramamurthy (2010), numerical and empirical analyses sets the degrees of freedom $\nu_{\boldsymbol{\psi}}$ to 15.

distribution $p(\boldsymbol{\psi}, \mathbf{r})$ is following as:

$$p(\boldsymbol{\psi}, \mathbf{r}) \propto \exp\left[-H(\boldsymbol{\psi}, \mathbf{r})\right].$$

In the next step of the simulation, the HMC method is used to update the momentum and parameter vectors of the candidates $\boldsymbol{\psi}$ and $\mathbf{r}$. This adjustment is achieved using a leapfrog integrator, a time-integration technique. The leapfrog integrator proceeds according to the updates

$$\begin{aligned}
\mathbf{r}\left(k+\frac{\xi}{2}\right) &= \mathbf{r}(k) - \frac{\xi}{2}\nabla U(\psi(k)), \\
\boldsymbol{\psi}(k+\xi) &= \boldsymbol{\psi}(k) + \xi\mathbf{r}\left(k+\frac{\xi}{2}\right), \\
\mathbf{r}(k+\xi) &= \mathbf{r}\left(k+\frac{\xi}{2}\right) - \frac{\xi}{2}\nabla U(\boldsymbol{\psi}(k+\xi))
\end{aligned}$$

where $\mathbf{r}(k)$ and $\psi(k)$ $(k = 1, \ldots, K)$ denote the values of the momentum and parameters $\mathbf{r}$ and $\boldsymbol{\psi}$ at step k and $\nabla U(\boldsymbol{\psi}(k))$ denotes the gradient with respect to ψ. Starting from the current values of these parameters $\boldsymbol{\psi}_{old}$ and drawing $\mathbf{r}(0)$ from the distribution $\mathcal{N}(0, \mathbf{I}_3)$, we can update the pair ($\mathbf{r}$, $\boldsymbol{\psi}$) in each step k by taking $K-1$ discrete steps of size ξ. The resulting state at the end of the iteration will be denoted by the candidate parameter $\boldsymbol{\psi}_{new}$. When generating via HMC, the parameter can be moved in steps larger than those of the MH algorithm.

In the MH step, the acceptance probability is given by

$$Pr_4(\boldsymbol{\psi}_{old}, \boldsymbol{\psi}_{new}) = \min\left[\frac{p(\boldsymbol{\psi}_{new}, \mathbf{r}_{new})}{p(\boldsymbol{\psi}_{old}, \mathbf{r}_{old})}, 1\right],$$

and set $\boldsymbol{\psi} = \boldsymbol{\psi}_{new}$ with probability $Pr_4(\boldsymbol{\psi}_{old}, \boldsymbol{\psi}_{new})$; otherwise, $\boldsymbol{\psi} = \boldsymbol{\psi}_{old}$. In the same manner as the TaB-MH algorithm, if the $S(\boldsymbol{\psi}_{new})$ does not satisfy the stationarity, then the candidate values are rejected with probability one.

The performance of the sampling scheme by using the HMC depends on the setting of the discretization time ξ and the number of steps K. At a glance, ξ controls the step size to turn up how large each $\nabla U(\boldsymbol{\psi})$ and $\mathbf{r}$ should be. If the number of iterations and the step size are too small, the algorithm may not explore sufficiently many candidates. These values must be tuned. Thus, the NUTS approach proposed by Hoffman and Gelman (2014) automatically selects appropriate ξ and K. This method is implemented in computation packages such as Stan and PyMC. A comprehensive elucidation of the NUTS model lies beyond the purview of this study; for a thorough exposition of NUTS, refer to Hoffman and Gelman (2014).

3 Illustrative examples

3.1 Numerical experiments

We conduct a simulation study to evaluate the performance of three approaches. The desirable properties for sampling methods in MCMC are efficiency and good mixing, which yield fast convergence. In addition to these properties, computational requirements are important for empirical analysis.

Following previous literature (Chib and Ramamurthy 2010; Ohtsuka and Kakamu 2015; Wolf et al. 2018), we also measure these efficiencies using an inefficiency factor (IF) and effective sample size (ESS). The inefficiency factor is defined as

$$IF = 1 + 2\sum_{s=1}^{L} r_s,$$

where r_s is the sample autocorrelation at lag s calculated from the simulated values. The inefficiency factor measures how well the chain mixes and is defined as the ratio of the numerical variance of the sample posterior mean to the variance of the sample mean under hypothetical uncorrelated draws (Chib 2001). The ESS is the ratio of the number of iterations in the posterior simulation to the inefficiency factor. This measure is defined as $ESS = M/IF$, where M is the iteration number. The ESS is a significant statistical measure that quantifies the extent to which independent information is encapsulated within the samples generated by the MCMC. If the samples are independent, the ESS equals the sample size. Conversely, when the chain's samples are strongly correlated, the ESS value approaches zero. In the TaB-MH algorithm, the posterior mode and the approximate posterior covariance matrix must be computed at each iteration, which involves optimization. Our approach may require more computational time than other methods. Therefore, this study evaluates the computational efficiency of each sampling method in terms of ESS per second (ESS/s). A higher ESS/s indicates better computational efficiency, as it enables a more accurate approximation of the posterior distribution with less computation time.

We conduct a Monte Carlo experiment to evaluate the performance of our sampling method. The data are simulated using the SDPD model presented in Eq. (11),

$$\begin{aligned} \mathbf{y}_t &= \rho\mathbf{W}\mathbf{y}_t + \phi\mathbf{y}_{t-1} + \theta\mathbf{W}\mathbf{y}_{t-1} + \alpha\boldsymbol{\iota}_n + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\eta}_t, \\ \boldsymbol{\eta}_t &= \boldsymbol{\mu} + \sigma\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Lambda}). \end{aligned} \tag{11}$$

The weight matrix $\mathbf{W}$ is constructed 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. Values for $\mathbf{X}_t$ and $\mathbf{x}_{it} = (x_{1it}, x_{2it})$ were generated using standard normal distribution. Concerning the parameters, we implement the same settings as in Parent and LeSage (2012). We implement different settings for spatio-temporal parameters ψ and sample size. With these parameters, various settings are implemented; $(\rho, \phi, \theta) = (0.9, 0.9, -0.85)$, which corresponds to a case of strong interaction in both spatial and serial parameters, $(\rho, \phi, \theta) = (0.4, 0.9, -0.40)$ and $(0.9, 0.4, -0.40)$ which in the cases of either spatial or

serial correlation is strong. Moreover, we varied $T = 5, 30$ and $N = 50$. The other parameters are set to be $\boldsymbol{\beta}' = (\beta_1, \beta_2) = (2, 2)$, $\alpha = 2$; $\boldsymbol{\mu}$ generates from $\mathcal{N}(0, \tau^2 \mathbf{I}_n)$ with $\tau^2 = 0.05$; λ_i $(i = 1, \ldots, n)$ generates the inverse of $\chi^2(\nu)/(\nu - 2)$ with $\nu = 7$. A pre-sample $\mathbf{y}_0$ draws from the initial distribution $\mathcal{N}(\mathbf{A}\alpha\iota_n, \mathbf{A}'\mathbf{B}\mathbf{A})$ with $\mathbf{A} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}$ and $\mathbf{B} = \tau^2\mathbf{I}_n + \sigma^2\boldsymbol{\Lambda}$.

The hyperparameters of the prior distribution are the following.

$$\boldsymbol{\beta}_0 = \mathbf{0}_2,\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}_0} = 100\mathbf{I}_2,\ \alpha_0 = 0.0,\ \sigma^2_{\alpha_0} = 100,\ \delta_{\tau_0} = 0.05,\ s_{\tau_0} = 2.0,\ \delta_{\sigma_0} = 0.05,$$
$$s_{\sigma_0} = 2.0,\ \boldsymbol{\psi}_0 = \mathbf{0}_3,\ \boldsymbol{\Sigma}_{\boldsymbol{\psi}_0} = 100\mathbf{I}_3,\ \delta_{\nu_0} = 6.0,\ s_{\nu_0} = 1.0.$$

In the numerical experiment using the RW-MH, we set the tuning parameter $s^2_{\boldsymbol{\psi}}$ to 1.0. For each simulation, we use $15,000$ posterior iterations after a burn-in of $5,000$ iterations. All the results in this work were calculated using MATLAB version R2025a.

Across all datasets examined, the true values were contained within the 95% credible intervals. Thus, we focus only on relevant issues to compare the sampling efficiency for the parameter $\boldsymbol{\psi}$. Table 1 summarizes the inefficiency factors, effective sample sizes per second, Geweke (1992)'s convergence diagnostics[7] and computation time, including discarded and rejected draws for each data set and estimation method. Regarding mixing quality, evaluated using the inefficiency factor, the TaB-MH method consistently yields lower values than the other methods in all cases. To improve upon the existing method, RW-MH was extended to simultaneously sample three parameters. However, the IFs often had a three-digit value, suggesting that more iterations are required for empirical analysis. Despite the NUTS method exhibiting higher estimation efficiency than the RW-MH method, its results are inferior to those of the TaB-MH algorithm.

Next, we will look at the results by the number of time series. First, from the results for the small data set, we find that the CD values for the RW and NUTS methods are below 5%, and the parameter sequence has not converged. On the other hand, the TaB-MH method performs well even with a small number of time-series samples. Second, the results for the medium-sized dataset ($T = 30$) showed that the NUTS method improved estimation efficiency, whereas the RW-MH algorithm showed no significant improvement.

Moreover, an investigation will be conducted into the estimated efficiency results, with consideration given to computation time. The TaB-MH method requires substantially more computational time than the other methods for each dataset. However, when the ESS per second for each parameter was examined, the TaB method demonstrated higher values than those of other methods. In summary, the TaB-MH method was shown to be at least 1.3 times, and up to six times, more computationally efficient than the NUTS method, based on the average ESS per second. This finding suggests that the proposed method is more computationally efficient than NUTS.

Figure 1 and 2 display the sample draws from the posterior distribution and their autocorrelation functions (ACF) to illustrate the differences in mixing between the

[7] 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 those generated from the target distribution when the CD is less than 0.05 (see Geweke, 1992, for a detailed discussion of the CD).

three methods in the cases of $\rho = 0.9$, $\phi = 0.9$, $\theta = -0.85$, $n = 50$, and $T = 5$. Figure 1 shows the sample paths obtained from the TaB-MH (left panels), the RW-MH (middle panels), and the NUTS (right panels). From the sample-path movements, it is found that the TaB-MH algorithm mixes more effectively than the other methods. As with the sample paths, the ACFs exhibit the same tendency. The figure of the ACFs shows that the TaB-MH algorithm performs better for autocorrelations that decay rapidly to zero than the other methods. The result for TaB-MH, as indicated by the autocorrelation figure, is consistent with that of the ACF, which disappears. On the contrary, the results of ACF for the RW-MH algorithm indicate that there is serious autocorrelation for parameters at large lag lengths. The NUTS method is more efficient than the RW-MH algorithm but is inferior to the TaB-MH algorithm. Thus, our algorithm converges faster than the other methods.

Finally, we examine the correlations among the drawing parameters and discuss why the convergence of the posterior samples is slow. Figure 3 shows the scatter plots of drawing from the full conditional distribution of the parameter ψ using the TaB-MH approach[8]. From the figure, it is found that the negative correlations between ρ and θ, and between ϕ and θ, are higher than those with the other approaches. Even when using the joint-sampling approach, the parameters remain highly correlated. Then, it is not appropriate to sample the parameters separately, as this dependence between them slows the convergence of the MCMC chain. Therefore, for the SDPD model with an a priori relationship among parameters, we conclude that the approach based on blocking and approximating the posterior distribution is more efficient than the other algorithm.

3.2 Application to fiscal policy analysis by using prefectural panel data in Japan

In this empirical illustration, we apply the SDPD model and the TaB-MH algorithm to estimate government spending spillovers onto private consumption or investment in Japan, as in Funashima and Ohtsuka (2019). Previous literature has shown that spatial dependence is important when examining Japanese regional data, such as manufacturing productivity (Kakamu et al. 2010) and prefectural gross domestic product (Ohtsuka 2015).

The Japanese economy faces challenges, including a declining birth rate, an aging population, and an excessive concentration of population and industry in Tokyo. In addition, Japan's debt-to-GDP ratio is the highest among advanced economies, necessitating effective policies to address these problems. As mentioned above, Japan's domestic economy is highly concentrated in Tokyo, leading to significant disparities between urban and rural areas. Therefore, fiscal policy, such as government spending, must be formulated with consideration for the characteristics of regional economic activity. To evaluate the effects of government spending on the regional economy, policymakers must consider both spatial interaction and temporal structure. When evaluating the effects of government spending on the regional economy, policy-makers must consider not only spatial interaction but also the region's temporal structure.

[8]This scatter plot shows 1000 resamples from 10,000 posterior samples.

Funashima and Ohtsuka (2019) estimated these effects by using panel spatial autoregressive models. Thus, this empirical analysis shows that implementing spatial and temporal dependence is important for understanding spillovers to neighboring regions in economic activity.

In this study, the model setting follows Funashima and Ohtsuka (2019). Let $\dot{\mathbf{y}}_t = (\dot{y}_{1t}, \ldots, \dot{y}_{nt})'$ and $\dot{x}_t = (\dot{x}_{1t}, \ldots, \dot{x}_{nt})'$, which denotes the growth rate of private consumption (investment) in the ith prefecture at time t. Define $\mathbf{x}_{1t}$ and $\mathbf{x}_{2t}$ as the first differences of the share of government consumption and investment in prefectural product, respectively. The empirical specification is described by the following:

$$\begin{aligned}\dot{\mathbf{y}}_t &= \rho \mathbf{W}\dot{\mathbf{y}}_t + \rho \dot{\mathbf{y}}_{t-1} + \theta \mathbf{W}\dot{\mathbf{y}}_{t-1} + \alpha \boldsymbol{\iota}_n + \dot{\mathbf{x}}_t \boldsymbol{\beta} + \boldsymbol{\eta}_t, \\ \boldsymbol{\eta}_t &= \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \sigma^2 \boldsymbol{\Lambda})\end{aligned}$$

where β represents the government spending (government consumption) effect. If estimated β is positive (negative), then the government spending crowds in (out) the private sector. For instance, an increase in taxation to fund government expenditure would reduce disposable income, thereby decreasing consumption. This, in turn, reduces corporate sales and, consequently, private investment. Furthermore, if funds are raised by issuing government bonds, interest rates rise, suppressing both private consumption and investment. Government expenditure can be categorized into two distinct classifications: government consumption and public investment. The government provides services such as social security and education, which, if substituted, reduce private consumption. Therefore, it is imperative to demonstrate whether government expenditure stimulates or discourages private economic activity.

Our panel data set comprises annual observations of the 47 prefectures in Japan from the financial year 2011 to 2021.[9] All data are obtained from the Annual Report on Prefectural Accounts, and in real terms, based on the System of National Accounts 2008 (08SNA). These data can be retrieved from the Cabinet Office of Japan's website. Regarding the weight matrix, we use the contiguity dummy variables. Excluding Okinawa Prefecture, Japan comprises four major islands: Hokkaido, Honshu, Shikoku, and Kyushu. Although these four islands are geographically separated, we assume they are connected by train and road (Kakamu et al. 2010).[10]

Next, the hyperparameters of the prior distribution are the following

$$\begin{aligned}&\alpha_0 = 0.0,\ \sigma^2_{\alpha_0} = 100,\ \beta_0 = 0,\ \Sigma_{\beta_0} = 10,\ \delta_{\tau_0} = 0.05,\ s_{\tau_0} = 2.0, \\ &\delta_{\sigma_0} = 0.05,\ s_{\sigma_0} = 2.0,\ \upsilon_0 = \mathbf{0}_3,\ \Sigma_{\upsilon_0} = 10\mathbf{I}_3,\ \delta_{\nu_0} = 10.0,\ s_{\nu_0} = 1.0.\end{aligned}$$

We perform MCMC to generate $15,000$ draws in a single sample path and discard the first $5,000$ as the initial burn-in. To verify whether the estimation method proposed in this study is also applicable to empirical analysis, we performed estimates using three methods, as in the previous numerical experiments.

[9] Data converted to real terms using the latest inflation rates is only available up to 2021.

[10] 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.

Table 2 reports the estimated parameters, inefficiency factors, and relative inefficiency factors using the different data sets and the TaB-MH algorithm. Across all datasets, the CD values for the parameters were at least 0.05. The parameters, except the constant term, in the models using private consumption, and β_{inv} are the 95% credible intervals that did not include zero. Initially, the focus is on the estimated parameter outcomes. Row (1) of Table 2 indicates that the degrees of contemporaneous spatial correlation for private consumption movements are estimated to be about 0.8, which are positive and significantly strong. There are also positive but low time dependence effects indicated by the estimated values of ϕ. Row (2) of Table 2 indicates that the strengths of contemporaneous spatial correlation for private investment are also positive, but their values are lower than those for private consumption. For both cases of private consumption and investment, the values of $\phi - (\rho - \theta)$ are negative for all data sets. The findings indicate that spatial dependence exceeds temporal dependence, thereby substantiating the plausibility of spatially correlated spillovers across both space and time. Moreover, the posterior means of the degrees of freedom are in the single digits, indicating the need to account for heterogeneity across prefectures. Since all the posterior means of β_{cons} of the Government consumption are negative, it is expected that increases in government consumption depress the private sector. The subsequent sentence on marginal effects provides further details on this matter.

Next, we consider how differences in estimation methods affect sampling efficiency and computational cost. RIF is the ratio of the IF of the RW-MH (NUTS) divided by the IF of the TaB-MH algorithm, and is an indicator of how inefficient the RW-MH (NUTS) is compared to the TaB-MH algorithm. When focusing on the three parameters ρ, ϕ, and θ, which are the main objectives of this study, the RW-MH algorithm is the least efficient across all datasets. For private consumption, the RIF values for ϕ and θ exceed 1, indicating that the TaB-MH algorithm is more efficient than the NUTS method for sampling these parameters. Moreover, the mean ESS/s for the TaB method exceeded that of both the RW and NUTS methods, indicating that the TaB method is more efficient per unit time than the other two methods. On the other hand, when looking at the values of RIF and ESS/s in Row (2) of Table 2, the results showed that the NUTS method was either more efficient than the TaB-MH algorithm. It is implied that this may be because the posterior means of the parameters ρ, ϕ, and θ are closer to zero than those for private consumption. Thus, in this case, we recommend the NUTS method considering computational cost.

Finally, we evaluate the impacts of crowding out effects using posterior samples and marginal effects. Spatial models account for relationships with neighboring regions, enabling decomposition of the marginal effect into a direct and an indirect effect. The direct effect corresponds to the own prefecture's own partial derivatives, which measure the impact on the ith prefecture's private sector from a change in the government expenditure (consumption) of the own prefecture. On the other hand, the indirect effect is cumulated over neighboring spatially correlated prefectures and corresponds to the cross partial derivatives. To estimate these effects, we use the summary measures proposed by LeSage and Pace (2009). Detailed mathematical formulas for these measures in the SDPD model are summarized in the Supplementary material.

Table 3 reports the corresponding effect estimates of the SDPD models for all data sets. All effects except those of Government investment are negative, with their 95% credible intervals not including zero. Thus, it has been shown that increases in government consumption in prefectures not only suppress private economic activity within their own region but also reduce it in neighboring regions. The increase in social security costs has put pressure on national and local government finances, driven by Japan's declining birth rate and aging population. Consequently, the prospect of increased government expenditure may engender apprehension among households and businesses about potential increases in taxes and social insurance contributions, thereby prompting a decline in economic activity. In the case of private consumption, the indirect effect exceeded the direct effect, as the estimated results indicated that spatial dependence is stronger than time dependence. Since the strengths of the first-order time dependence are positive, the direct effects on the own prefecture increase, and the indirect effects weaken in the long-term. Turning to private investment, when it exhibits weaker spatial correlation than consumption, the indirect effects are lower than the direct effects. This result suggests that Japanese companies are concentrated in the surrounding areas, as noted by Nakajima et al. (2012). Conversely, a comparison of the long and short-term effects reveals an antithetical relationship to the results observed for consumption, attributable to negative estimated autocorrelation parameters. A decrease in investment by one company could affect other companies in the supply chain. The findings of this analysis indicate that a decline in government consumption may be a more effective strategy for stimulating regional economic activity in Japan.

4 Conclusion

In this study, the initial step is to provide a comprehensive overview of the SDPD model and of a Bayesian estimation method for it. The prevailing approach to estimating three parameters that characterize spatial and temporal correlations is to estimate them individually using the RW-MH method. This method is known to be inefficient because it generates random numbers that depend on the previously sampled value, leading to high autocorrelation in the posterior sample series. Furthermore, the SDPD model's time-lag structure implies stationarity. Given the interdependence of the three parameters for stationarity, sampling them individually would lead to high correlation among them. Consequently, empirical researchers are compelled to conduct more trials, thereby incurring substantially higher computational costs.

To enhance the usefulness of this model in empirical analysis, we propose a method that samples the three parameters jointly and is more efficient than the RW-MH method. The independent MH method is therefore proposed, in which the full conditional posterior distribution is approximated by a Taylor expansion and the resulting distribution is constructed. Moreover, to facilitate a comparative analysis, the NUT method will be employed. This approach has recently been adopted as the standard for Bayesian inference in R and Python. In illustrative examples, we first measured the efficiency of different estimation methods using artificial data. Our proposed algorithm

exhibits superior mixing and computational efficiency, and has been shown to outperform the NUTS method, which is the most efficient general-purpose method. Then, in an applied illustration, we demonstrated the fiscal policy effect, such as crowding out, using Japanese prefectural panel data.

In the analysis of private consumption, which exhibits strong spatial correlations, our proposed method proved more efficient than the NUTS method. However, for variables with weak temporal and spatial correlations, such as private investment, the NUTS method proved computationally less expensive and more efficient. Thus, the estimation method should vary with the data used.

Finally, some remaining issues remain. The model used in this analysis assumed that spatiotemporal parameters were constant. Future work should explore extending this model to one in which these parameters vary over time or across regions, as in Ohtsuka and Kakamu (2018) and Miyawaki (2013). This would enable us to quantitatively demonstrate that the strength of the spatial correlations identified by Funashima and Ohtsuka (2019) varies across regional blocks and to show changes in consumption dynamics due to the spread of COVID-19. However, because the SDPD model is highly versatile, it can be applied to a variety of empirical analyses. Therefore, comparing the estimation methods tested in this study could serve as a basis for future empirical analyses.

Table 1 Efficiency and convergence comparison results using experimental data

(1)Small data set: $n = 50$ and $T = 5$

Method	TaB-MH			RW-MH			NUTS		
Statistics	IF	ESS/s	CD	IF	ESS/s	CD	IF	ESS/s	CD
Pattern #1: $(\rho, \phi, \theta) = (0.9, 0.9, -0.85)$									
ρ	66.79	0.90	0.72	277.74	0.47	0.17	122.98	0.89	0.01
ϕ	6.20	9.65	0.88	290.96	0.90	0.02	68.19	1.59	0.00
θ	8.02	7.47	0.98	307.74	0.43	0.00	96.05	1.14	0.00
Time		167			76			91	
Ave. ESS/s		6.01			0.45			1.21	
Pattern #2: $(\rho, \phi, \theta) = (0.4, 0.9, -0.4)$									
ρ	37.49	1.58	0.71	308.80	0.43	0.00	243.01	0.46	0.00
ϕ	10.60	5.58	0.71	92.44	1.44	0.00	18.68	6.01	0.01
θ	25.94	2.28	0.91	288.49	0.46	0.20	225.66	0.50	0.00
Time		169			75			89	
Ave. ESS/s		3.15			0.78			2.32	
Pattern #3: $(\rho, \phi, \theta) = (0.9, 0.4, -0.4)$									
ρ	97.95	0.60	0.06	308.37	0.41	0.00	229.35	0.49	0.01
ϕ	12.38	4.78	0.88	207.77	0.60	0.00	38.86	2.89	0.62
θ	42.48	1.39	0.12	313.63	0.40	0.00	158.60	0.71	0.36
Time		169			80			89	
Ave. ESS/s		2.26			0.47			1.36	

(2)Medium data set: $n = 50$ and $T = 30$

Method	TaB-MH			RW-MH			NUTS		
Statistics	IF	ESS/s	CD	IF	ESS/s	CD	IF	ESS/s	CD
Pattern #1: $(\rho, \phi, \theta) = (0.9, 0.9, -0.85)$									
ρ	30.00	0.09	0.27	167.41	0.03	0.28	112.24	0.04	0.32
ϕ	5.39	0.58	0.07	148.08	0.09	0.82	31.69	0.13	0.06
θ	8.35	0.37	0.74	179.38	0.04	0.78	66.00	0.06	0.54
Time		3222			2366			2382	
Ave.ESS/s		0.35			0.03			0.08	
Pattern #2: $(\rho, \phi, \theta) = (0.4, 0.9, -0.4)$									
ρ	9.28	0.33	0.47	265.86	0.02	0.01	103.16	0.04	0.95
ϕ	1.80	1.71	0.67	35.66	0.12	0.50	12.40	0.30	0.01
θ	5.99	0.51	0.52	259.11	0.02	0.01	104.18	0.04	0.53
Time		3260			2378			2665	
Ave.ESS/s		0.85			0.05			0.13	
Pattern #3: $(\rho, \phi, \theta) = (0.9, 0.4, -0.4)$									
ρ	10.44	0.31	0.76	44.10	0.09	0.15	15.61	0.064	0.71
ϕ	3.06	1.07	0.31	128.12	0.03	0.19	17.42	0.057	0.06
θ	4.77	0.69	0.56	145.00	0.03	0.74	21.55	0.046	0.15
Time		3049			2414			2539	
Ave.ESS/s		0.69			0.05			0.22	

Note: IF, ESS/s, and CD are the inefficiency factor, the efficient sample size per second, and the p-value of the convergence diagnostic test in Geweke (1992), respectively. Time is the number of seconds required by the computation, including discarded and rejected draws. Ave. ESS/s denotes the average of the ESS/s.

Table 2 Estimation results of the empirical model using prefectural panel data

(1) Dependent variable: Private consumption

Method	TaB-MH			RW-MH		NUTS	
Parameter	Mean	IF	ESS/s	RIF	ESS/s	RIF	ESS/s
ρ	0.829*	6.56	5.1	5.8	1.4	2.7	2.6
ϕ	0.197*	8.57	3.9	34.5	0.2	2.1	2.6
θ	−0.227*	8.25	4.0	36.2	0.2	3.3	1.7
α	−0.038	1.47	22.6	1.5	23.5	3.3	9.5
β_{cons}	−0.368*	3.81	8.7	3.6	3.8	2.0	5.9
β_{inv}	0.009	1.11	29.8	3.9	12.1	1.3	30.7
τ^2	0.076*	10.25	3.2	13.0	0.4	2.7	1.6
σ^2	0.734*	7.98	4.2	1.8	3.7	1.1	5.5
ν	6.876*	7.15	4.6	1.3	5.8	1.0	6.6
Time		302			191		218
Ave.ESS/s		4.3			0.6		2.3

(2) Dependent variable: Private investment

Method	TaB-MH			RW-MH		NUTS	
Parameter	Mean	IF	ESS/s	RIF	ESS/s	RIF	ESS/s
ρ	0.222*	3.59	9.5	5.0	3.0	0.9	16.0
ϕ	−0.151*	3.70	9.5	4.8	3.1	0.8	16.4
θ	0.140*	3.62	9.4	7.5	2.0	1.2	11.5
α	1.066*	1.44	23.5	3.4	11.3	1.8	19.3
β_{cons}	−3.664*	2.05	16.5	3.8	7.1	1.2	19.7
β_{inv}	0.289	1.39	24.3	1.1	36.6	0.9	38.7
τ^2	0.054*	36.12	0.9	0.8	1.9	0.8	1.6
σ^2	24.58*	5.67	6.0	1.5	6.5	1.7	5.1
ν	5.353*	5.71	5.9	1.2	7.9	1.3	6.4
Time		295			183		203
Ave.ESS/s		9.3			2.7		14.6

Mean, IF, ESS/s, and RIF denote the posterior mean, inefficiency factor, effective sample size per second, and relative inefficiency factor, respectively. RIF is calculated by dividing the IF of the RW-MH (NUTS) by the IF of the TaB-MH algorithm. This ratio indicates the efficiency of the RW-MH (NUTS) algorithm relative to the TaB-MH algorithm. cons and inv are abbreviations for Government consumption and investment, respectively. $*$ indicates that the 95% credible interval does not include zero. Time is the number of seconds required by the computation, including discarded and rejected draws. Ave. ESS/s denotes the average of the ESS/s.

Table 3 Effects estimates of empirical models using prefectural panel data

Dependent variable	Private consumption		Private investment	
Explanatory variable	Govt. cons	Govt. inv	Govt. con	Govt. inv
Short-term direct	-0.532^*	0.013	-3.697^*	0.288
Short-term indirect	-1.569^*	0.039	-0.982^*	-0.934
Long-term direct	-1.213^*	0.027	-3.262^*	0.253
Long-term indirect	-1.213^*	0.004	-1.341^*	0.104

Govt. cons (inv) is an abbreviation for Government consumption (investment).
$*$ denotes that the 95% credible interval does not include zero.

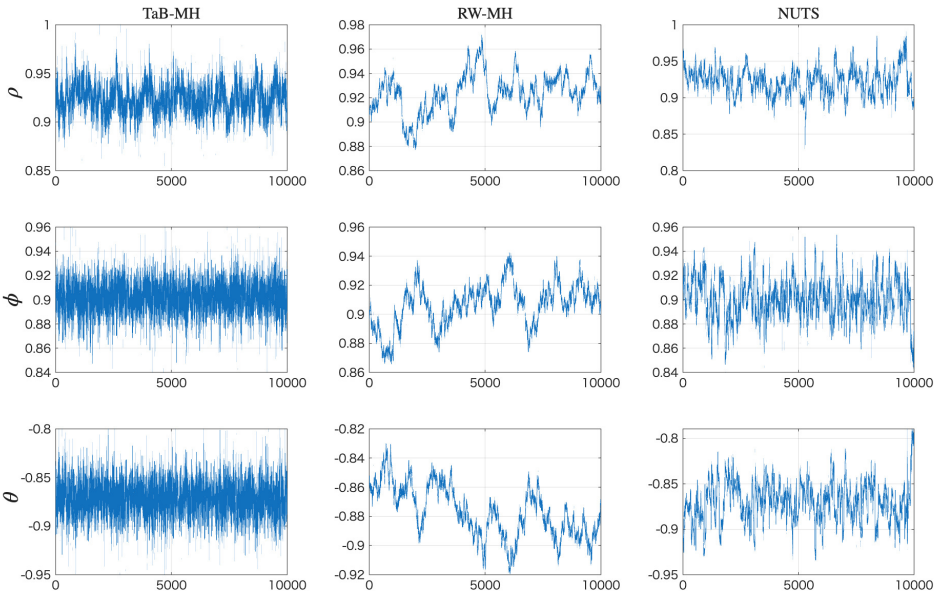


Fig. 1 Posterior sample path of simulation results: The time-series plots of draws using the TaR-MH (left panels), RW-MH (middle panels), and NUTS (right panels) in the case of $n = 50$, $T = 5$, $\rho = 0.9$, $\phi = 0.9$, and $\theta = -0.85$.

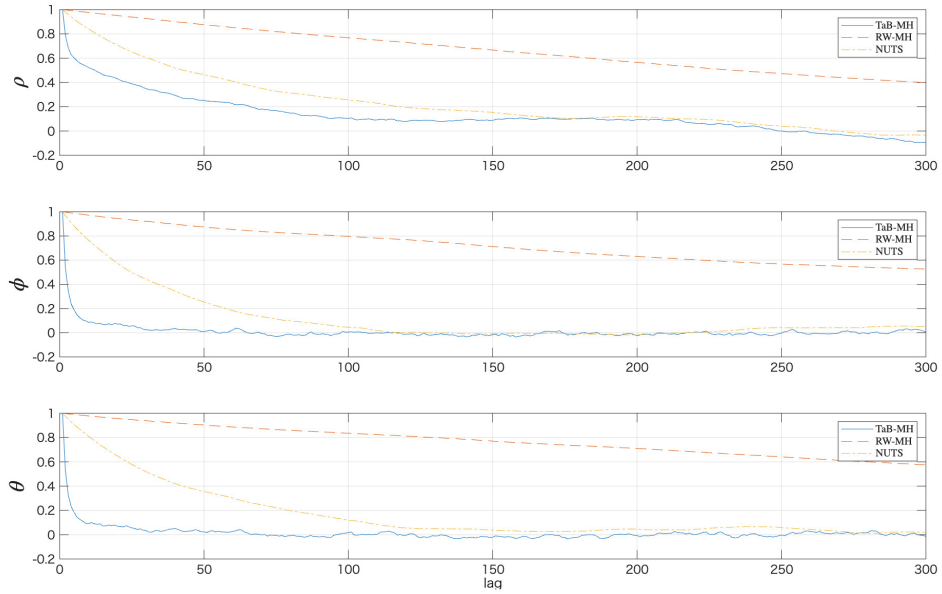


Fig. 2 Comparison of autocorrelation of simulation results: The figures show the autocorrelation function using the TaR-MH (solid lines), the RW-MH (dotted lines), and NUTS (dash-dot lines) in the case of $n = 50$, $T = 5$, $\rho = 0.9$, $\phi = 0.9$, and $\theta = -0.85$

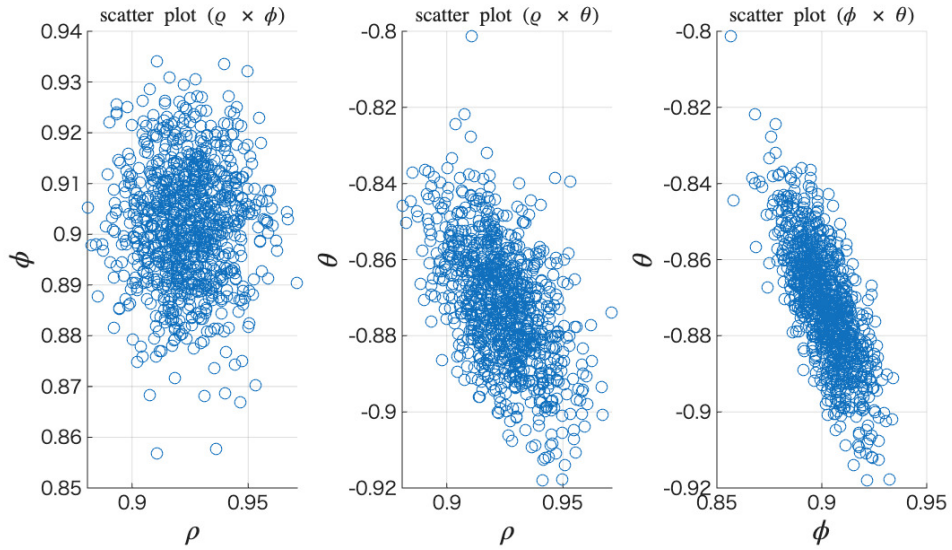


Fig. 3 Scatter plots among spatial and temporal parameters using the TaB-MH algorithm: Figures show in the case of $n = 50$, $T = 5$, $\rho = 0.9$, $\phi = 0.9$ and $\theta = -0.85$

References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht

Anselin L (2001) Spatial econometrics. In: Baltagi BH (ed) A companion to theoretical econometrics. Blackwell, Massachusetts, p 310–330

Arbia G (2006) Spatial econometrics: statistical foundations and applications to regional convergence. Springer, Berlin

Bivand RS, Gómez-Rubio V, Rue H (2014) Approximate Bayesian inference for spatial econometrics models. Spatial Statistics 9:146–165

Chib S (2001) Markov chain Monte Carlo methods: computation and inference. In: Heckman JJ, Leamer E (eds) Handbook of econometrics, vol 5. North Holland, Amsterdam, pp 3569–3649

Chib S, Ramamurthy S (2010) Tailored randomized block MCMC methods with application to DSGE models. Journal of Econometrics 155(1):19–38

Duane S, Kennedy AD, Pendleton BJ, et al (1987) Hybrid Monte Carlo. Physics Letters B 195(2):216–222

Elhorst JP (2014) Spatial econometrics: from cross-sectional data to spatial panels. Springer, Heidelberg

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 195(2):216–222

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo JM, Berger JO, Dawid AP, et al (eds) Bayesian statistics, vol 4. Oxford University Press, Oxford, pp 169–193

Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov chain Monte Carlo in practice. Chapman&Hallr, London/New York

Hoffman MD, Gelman A (2014) The No-U-Turn sample: Adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research 15(1):1351–1381.

Holloway G, Shankar B, Rahmanb S (2002) Bayesian spatial probit estimation: A primer and an application to HYV rice adoption. Agricultural Economics 27(3):383–402

Kakamu K, Wago H, Tanizaki H (2010) Estimation of regional business cycles in Japan is using the Bayesian panel spatial autoregressive approach. In: Nolin TP (ed) Handbook of regional economics. Nova Science Publishers, New York, p 555–571

Lee LF, Yu J (2014) Efficient GMM estimation of spatial dynamic panel data models with fixed effects. Journal of Econometrics 180(2):174–197

LeSage J, Pace RK (2009) Introduction to spatial econometrics. Chapman&Hall/CRC

Liu JS (1994) The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. Journal of the American Statistical Association 89(427):958–966

Mills JA, Parent O (2019) Bayesian MCMC estimation. In: Fischer MM, Nijkamp P (eds) Handbook of regional science. Springer, Berlin, Heidelberg, p 1–25

Miyawaki K (2013) Space-varying coefficient simultaneous autoregressive models for the structural analysis of residential water demand. Spatial Economic Analysis 8(4):498–518

Nakajima K, Saito YU, Uesugi I (2012) Measuring economic localization: Evidence from Japanese firm-level data. Journal of the Japanese and International Economies 26(2):201–220

Nocedal J, Wright SJ (2000) Numerical optimization. Springer, New York

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

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

Ohtsuka Y, Kakamu K (2018). Regional growth and business cycles in Japan. Review of Urban and Regional Development Studies 30(1):1–25

Parent O, LeSage JP (2011) A space-time filter for panel data models containing random effects. Computational Statistics and Data Analysis 55(1):475–490

Parent O, LeSage JP (2012) Spatial dynamic panel data models with random effects. Regional Science and Urban Economics 42(4):727–738

Stakhovych S, Bijmolt THA (2009) Specification of spatial models:A simulation study on weights matrices. Papers in Regional Science 88(2):389–409

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

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

Wolf LJ, 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(1):97–119

Yu J, de Jong R, Lee LF (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

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplmentaryMaterial.pdf