Observed-data DIC for Spatial Panel Data Models*

Osman Doğan† Ye Yang‡ Süleyman Taşpınar§

December 25, 2023

Abstract

In spatial panel data modeling, researchers often need to choose a spatial weights matrix from a pool of candidates, and decide between static and dynamic specifications. We propose observed-data deviance information criteria to resolve these specification problems in a Bayesian setting. The presence of high dimensional latent variables (i.e., the individual and time fixed effects) in spatial panel data models invalidates the use of a deviance information criterion (DIC) formulated with the conditional and the complete-data likelihood functions of spatial panel data models. We first show how to analytically integrate out these latent variables from the complete-data likelihood functions to obtain integrated likelihood functions. We then use the integrated likelihood functions to formulate observed-data DIC measures for both static and dynamic spatial panel data models. Our simulation analysis indicates that the observed-data DIC measures perform satisfactorily to resolve specification problems in spatial panel data modeling. We also illustrate the usefulness of the proposed observed-data DIC measures using an application from the literature on spatial modeling of the house price changes in the US.

JEL-Classification: C13, C21, C31.
Keywords: Spatial panel data models, Bayesian inference, MCMC, Deviance information criterion, DIC, Bayesian model comparison, Model selection.

*This research was supported, in part, by a grant of computer time from the City University of New York High Performance Computing Center under NSF Grants CNS-0855217 and CNS-0958379.

†Department of Economics, University of Illinois at Urbana-Champaign, U.S., email: odogan@illinois.edu.

‡School of Accounting, Capital University of Economics and Business, Beijing, China, email: yeyang557@hotmail.com.

§Department of Economics, Queens College CUNY, New York, U.S., email: staspinar@qc.cuny.edu.

1 Introduction

Spatial panel data models have found applications in recent literature varying from urban economics to growth economics (Aquaro et al. 2021; Baltagi et al. 2015; LeSage and Sheng 2014; Parent and LeSage 2010; Yang 2021a); environmental and energy economics to agriculture and land use (Baldoni and Esposti 2021; Billé and Rogna 2022; Chen et al. 2020; Gori Maia et al. 2021; Jiang et al. 2021; Krisztin et al. 2021; Kuschnig et al. 2021; Li et al. 2021; Song et al. 2021); health economics to store sales and sports (Ehlert and Oberschachtsiek 2014; Guliyev 2020; Hunneman et al. 2021; Krisztin et al. 2020; Lee and Lee 2021). Two common problems applied researchers face in modeling are (i) how to choose a spatial weights matrix (or the connectivity matrix) from a pool of candidates, and (ii) how to choose between the static and dynamic specifications. Often it is the case that the model is specified in an ad-hoc manner and there is no guidance from an underlying structural model to deal with these two problems. In this paper, we focus on these model specification problems for spatial panel data models and propose observed-data deviance information criteria to resolve them.

The deviance information criterion (DIC) was suggested by Spiegelhalter et al. (2002) for the parametric model comparison exercises. The DIC measure is defined as the sum of the posterior mean deviance and the effective number of parameters, and balances a trade-off between model fit and complexity. Using a decision-theoretical perspective, Li et al. (2017, 2020) show that the DIC provides an asymptotically unbiased estimator of the expected Kullback-Leibler (KL) divergence between the true data generating process (DGP) and a suitable plug-in predictive distribution of hypothetically replicate data. This result indicates that the DIC selects the candidate model that has the better predictive performance, i.e., the smaller the DIC value is, the better the predictive performance of the candidate model is.

There can be alternative definitions of the DIC measure depending on the type of likelihood function used in the deviance term (Celeux et al. 2006). For example, in the latent variable models, there can be three types of likelihood functions: (i) the conditional likelihood function, i.e., the likelihood function obtained by conditioning on the latent variables, (ii) the complete data likelihood function, i.e, the joint likelihood function of data and the latent variables, (iii) the integrated likelihood function (or the observed-data likelihood function) obtained from the complete likelihood function by integrating out the latent variables.

In the literature, there are numerous simulations studies showing that the DIC measure formulated with the conditional likelihood function can perform unsatisfactorily in selecting the true model, e.g., see Chan and Grant (2016a,b) and Millar (2009). Moreover, the DIC measures formulated with the conditional likelihood and the complete data likelihood functions undermine the assumptions required for the decision-theoretical perspective of Li et al. (2017, 2020). That is, these measures provide biased estimators of the expected Kullback-Leibler (KL) divergence between the true data generating process (DGP) and a suitable plug-in predictive distribution of hypothetically replicate data. On the other hand, the simulation studies in Chan and Grant (2016a,b) indicate that the DIC measure based on the integrated likelihood function, i.e., the observed-data DIC,

performs well in selecting the true model among a pool of candidates.

In this study, we show how an observed-data DIC measure can be formulated in a Bayesian estimation setting for spatial panel data models that have both individual and time fixed effects. We consider static and dynamic spatial panel data models that have high order spatial lags in both the dependent variable and the disturbance terms.[1] In these models, the individual and time fixed effects can be regarded as the high-dimensional latent variables. Though the conditional likelihood functions are readily available, it is not clear how to formulate the integrated likelihood functions for these models. A possible approach can be based on a transformation approach that can be used to wipe out the latent variables from the models. For example, in the case of static spatial panel data models, Lee and Yu (2010b) suggest two orthonormal transformations to wipe out both individual and time effects from the model.[2] However, this approach is possible only when all spatial weights matrices are row-normalized. Moreover, in the case of spatial dynamic panel data models, there is no such transformation that eliminates both individual and time effects, and preserve the structure of specifications (see Lee and Yu (2010a,b) and Han et al. (2017) for the details).

Instead of a transformation approach, we suggest a direct approach to integrate out both the individual and time fixed effects from the complete data likelihood functions of these models. To that end, we assume multivariate normal prior distributions for the latent variables, and then use some properties of multivariate normal distribution to analytically integrate out these variables from the complete data likelihood functions of the models. We then use our closed-form integrated likelihood functions to formulate the observed-data DIC measures for both static and dynamic spatial panel data models. We design an extensive simulation study to assess the finite sample performance of the observed-data DICs in selecting the true specification. Using a recent application from the literature on spatial modeling of the house price changes in the US, we illustrate how the proposed observed-data DICs can be used to choose a spatial weights matrix from a pool of candidates, and to choose between the static and dynamic specifications. Our results from both simulation study and empirical illustration indicate that the observed-data DICs can be useful for the specification problems in spatial panel data models.[3]

[1] There is a growing literature on the spatial panel data models dealing mainly with the estimation and testing approaches. The estimation methods considered in the literature include (i) the (quasi) maximum likelihood (ML) methods, (ii) the generalized method of moments (GMM) and IV based approaches, (iii) the M-estimation approach, and (iv) the Bayesian estimation approaches. Among others, see Baltagi et al. (2014), Elhorst (2005), Lee and Yu (2010a,b), Qu et al. (2017), and Yu et al. (2008) on the (quasi) ML based estimation approaches, see Fingleton (2008), Kapoor et al. (2007), and Lee and Yu (2014) on the GMM and IV based estimation approaches, see Li and Yang (2020, 2021) and Yang (2018) on the M-estimation approach, and see Han and Lee (2016), Han et al. (2017), LeSage (2014), and Parent and LeSage (2010, 2011, 2012) for the Bayesian estimation approaches. There is also a growing literature on the testing spatial parameters in spatial panel data models, e.g., among other, see Baltagi and Yang (2012, 2013), Baltagi et al. (2003, 2007a,b), Bera et al. (2019), Kelejian and Piras (2016), Robinson (2008), Taşpınar et al. (2017), and Yang (2021b). Our approach based on the observed-data DIC is an alternative to usual testing approaches.

[2] These transformations are based on the decomposition of $\mathbf{J}_n = \left(\mathbf{I}_n - \frac{1}{n}\boldsymbol{l}_n\boldsymbol{l}_n'\right)$ and $\mathbf{J}_T = \left(\mathbf{I}_T - \frac{1}{T}\boldsymbol{l}_T\boldsymbol{l}_T'\right)$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\boldsymbol{l}_n$ is the $n \times 1$ vector of ones. See Lee and Yu (2010b) for the details.

[3] There are also some other approaches for the specification problems in the spatial econometric literature. One strand of the literature focuses on the model averaging approaches to account for the model uncertainty associated with the choice of spatial weights matrices (Crespo and Feldkircher 2013; Debarsy and LeSage 2020; LeSage 2014;

The rest of this paper is organized as follows. In Section 2, we first specify static and dynamic spatial panel data models, and then discuss the stability conditions for each model. In Section 3, we discuss different variants of DIC and show how the observed-data likelihood function can be derived for static and dynamic spatial panel data models. In Section 4, we provide the details of our suggested Bayesian estimation approach. In Section 5, we investigate the performance of the observed-data DIC in an extensive simulation study. In Section 6, we illustrate the methodology with an application to spatial modeling of the US house price data. In Section 7, we present our conclusions. Some technical results are relegated to an Appendix.

2 Model Specifications

In this section, we first present the static and dynamic spatial panel data models, and then discuss the parameter space of spatial parameters ensuring stability of these models. We start with a static spatial panel data model that includes (i) individual effects, (ii) time effects, (iii) high order spatial lags of the dependent variable, and (iv) high order spatial lags of the disturbance terms. Let $\mathbf{Y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})^{'}$ be the $n \times 1$ vector of dependent variable at time t. We assume the following data generating process (DGP) for $\mathbf{Y}_t$:

$$\begin{aligned} \mathbf{Y}_t &= \sum_{r_1=1}^{p_1} \lambda_{r_1} \mathbf{W}_{1r_1} \mathbf{Y}_t + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{c}_n + \alpha_t \boldsymbol{l}_n + \mathbf{U}_t, \\ \mathbf{U}_t &= \sum_{r_2=1}^{p_2} \rho_{r_2} \mathbf{W}_{2r_2} \mathbf{U}_t + \mathbf{V}_t, \quad \text{for} \quad t = 1, \ldots, T, \end{aligned} \tag{2.1}$$

where $\mathbf{X}_t$ is the $n \times k$ matrix of non-stochastic time-varying regressors with a matching parameter vector $\boldsymbol{\beta}$, $\mathbf{c}_n = (c_1, c_2, \ldots, c_n)^{'}$ is the $n \times 1$ vector of time-invariant individual effects, $\alpha_t \boldsymbol{l}_n$ represents individual invariant time effect at period t, $\boldsymbol{l}_n$ is the $n \times 1$ vector of ones, $\mathbf{U}_t = (u_{1t}, u_{2t}, \ldots u_{nt})^{'}$ is the $n \times 1$ vector of regression disturbance term, and $\mathbf{V}_t = (v_{1t}, v_{2t}, \ldots, v_{nt})^{'}$ is the $n \times 1$ vectors of innovations. The high order spatial lag terms are $\sum_{r_1=1}^{p_1} \mathbf{W}_{1r_1} \mathbf{Y}_t$ and $\sum_{r_2=1}^{p_2} \mathbf{W}_{2r_2} \mathbf{U}_t$, where $\mathbf{W}_{1r_1}$ and $\mathbf{W}_{2r_2}$ are the $n \times n$ non-stochastic spatial weights matrices that have zero diagonal elements for $r_1 = 1, 2, \ldots, p_1$ and $r_2 = 1, 2, \ldots, p_2$. Thus, $\mathbf{U}_t$ is allowed to incorporate spatial autoregressive features. The scalar spatial parameters are denoted by λ_{r_1} and ρ_{r_2} for $r_1 = 1, 2, \ldots, p_1$ and $r_2 = 1, 2, \ldots, p_2$. We assume that v_{it}'s are independent and identically distributed normal random variables with mean zero and variance σ^2 across $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$.

Piribauer and Crespo 2016; Zhang and Yu 2018). Another strand of the literature focuses on how the elements of spatial weight matrices can be estimated (Kelejian and Piras 2014; Krisztin and Piribauer 2021; Lam and Souza 2020). Finally, instead of selecting a spatial weights matrix from a pool of candidates, there are some studies focusing on a modeling approach for specifying the elements of an endogenous spatial weight matrix (Han and Lee 2016; Qu and Lee 2015; Qu et al. 2017).

Next, we consider a dynamic version of (2.1), which can be in the following form:

$$\mathbf{Y}_t = \sum_{r_1=1}^{p_1} \lambda_{r_1} \mathbf{W}_{1r_1} \mathbf{Y}_t + \gamma \mathbf{Y}_{t-1} + \sum_{r_1=1}^{p_1} \eta_{r_1} \mathbf{W}_{1r_1} \mathbf{Y}_{t-1} + \mathbf{X}_t \beta + \mathbf{c}_n + \alpha_t \boldsymbol{l}_n + \mathbf{U}_t, \tag{2.2}$$

$$\mathbf{U}_t = \sum_{r_2=1}^{p_2} \rho_{r_2} \mathbf{W}_{2r_2} \mathbf{U}_t + \mathbf{V}_t, \quad \text{for} \quad t = 1, \ldots, T,$$

where the scalar parameter γ of time lag-term $\mathbf{Y}_{t-1}$ measures the persistence in $\mathbf{Y}_t$, and the scalar parameters η_{r_1}'s of spatial time lag terms $\mathbf{W}_{1r_1}\mathbf{Y}_{t-1}$ capture the dynamic diffusion effects. Note that, in (2.1) and (2.2), we have $c_i + \alpha_t = (c_i + \kappa) + (\alpha_t - \kappa)$ for arbitrary κ. Thus, the individual and time effects are not identified, and we need to impose normalization-type constraints to achieve identification. For this purpose, we may require that $\mathbf{c}_n^{'} l_n = 0$, or we may simply set $\alpha_1 = 0$.

Next, we discuss the stability conditions. Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_{p_1})^{'}$, $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_{p_2})^{'}$, $\mathbf{S}_1(\boldsymbol{\lambda}) = \mathbf{I}_n - \sum_{r_1=1}^{p_1} \lambda_{r_1} \mathbf{W}_{1r_1}$ and $\mathbf{S}_2(\boldsymbol{\rho}) = \mathbf{I}_n - \sum_{r_2=1}^{p_2} \rho_{r_2} \mathbf{W}_{2r_2}$. Since the spatial autoregressive models represent equilibrium relationships, we require that $\mathbf{S}_1(\boldsymbol{\lambda})$ and $\mathbf{S}_2(\boldsymbol{\rho})$ are invertible for all $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ values. Let $\vartheta_i(\mathbf{A})$ be the ith eigenvalue of the $n \times n$ matrix $\mathbf{A}$. Then, by Lemma 1 in Appendix, $\mathbf{S}_1(\boldsymbol{\lambda})$ and $\mathbf{S}_2(\boldsymbol{\rho})$ are invertible under the following respective sufficient conditions: (1^a) $\max_{1 \le i \le n} \left\{ \left| \vartheta_i \left(\sum_{r_1=1}^{p_1} \lambda_{r_1} \mathbf{W}_{1r_1} \right) \right| \right\} < 1$, and (2^a) $\max_{1 \le i \le n} \left\{ \left| \vartheta_i (\sum_{r_2=1}^{p_2} \rho_{r_2} \mathbf{W}_{2r_2}) \right| \right\} < 1$. The proof of Lemma 1 indicates that the necessary and sufficient condition for the invertibility of $\mathbf{S}_1(\boldsymbol{\lambda})$ is $\vartheta_i(\sum_{r_1=1}^{p_1} \lambda_{r_1} \mathbf{W}_{1r_1}) \neq 1$ for $i = 1, \ldots, n$ (the analogous result also applies to $\mathbf{S}_2(\boldsymbol{\rho})$). Thus, the sufficient conditions (1^a) and (2^a) may lead to restrictive parameter spaces for $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$. However, by requiring that all eigenvalues of $\sum_{r_1=1}^{p_1} \lambda_{r_1} \mathbf{W}_{1r_1}$ are less than one in absolute value, we rule out the unstable Nash equilibria.[4]

By Lemmas 1 and 2 in Appendix, we may alternatively specify the parameter space restrictions for stability using the following relatively restrictive sufficient conditions: (1^{*a}) $\left\| \sum_{r_1=1}^{p_1} \lambda_{r_1} \mathbf{W}_{1r_1} \right\| < 1$ and (2^{*a}) $\left\| \sum_{r_2=1}^{p_2} \rho_{r_2} \mathbf{W}_{2r_2} \right\| < 1$, where $\| \cdot \|$ is any matrix norm.[5] Also, note that

$$\left\| \sum_{j=1}^{p} a_j \mathbf{W}_j \right\| \le |a_1| \cdot \|\mathbf{W}_1\| + \ldots + |a_p| \cdot \|\mathbf{W}_p\| \le \left(\sum_{j=1}^{p} |a_j| \right) \times \max_{1 \le j \le p} \|\mathbf{W}_j\|,$$

where a_j's are scalar parameters and $\mathbf{W}_j$'s are $n \times n$ matrices. Thus, we can further restrict the parameter space by requiring the following easy-to-check sufficient conditions: (1^{*a}) $\left(\sum_{r_1=1}^{p_1} |\lambda_{r_1}| \right) \times \max_{1 \le r_1 \le p_1} \|\mathbf{W}_{1r_1}\| < 1$, and (2^{*a}) $\left(\sum_{r_2=1}^{p_2} |\rho_{r_2}| \right) \times \max_{1 \le r_2 \le p_2} \|\mathbf{W}_{2r_2}\| < 1$. If we choose the matrix row sum norm, which is denoted by $\| \cdot \|_\infty$, the above conditions further simplifies for the row-normalized spatial weights matrices: (1^{*a}) $\sum_{r_1=1}^{p_1} |\lambda_{r_1}| < 1$, and (2^{*a}) $\sum_{r_2=1}^{p_2} |\rho_{r_2}| < 1$.

The stability conditions for the dynamic specification in (2.2) can be investigated from its

[4] See Kelejian and Prucha (2010) for the details.

[5] Note that these conditions are relatively restrictive since they provide upper bounds on the conditions in (1^a) and (2^a) respectively.

reduced form, which can be expressed as

$$\mathbf{Y}_t = \mathbf{S}_1^{-1}(\boldsymbol{\lambda}) \left(\gamma \mathbf{I}_n + \sum_{r_1=1}^{p_1} \eta_{r_1} \mathbf{W}_{1r_1} \right) \mathbf{Y}_{t-1} + \mathbf{S}_1^{-1}(\boldsymbol{\lambda}) \left(\mathbf{X}_t \boldsymbol{\beta} + \mathbf{c}_n + \alpha_t \boldsymbol{l}_n + \mathbf{S}_2^{-1}(\boldsymbol{\rho}) \mathbf{V}_t \right). \tag{2.3}$$

Let $\mathbf{A}(\boldsymbol{\lambda}, \gamma, \boldsymbol{\eta}) = \mathbf{S}_1^{-1}(\boldsymbol{\lambda})(\gamma \mathbf{I}_n + \sum_{r_1=1}^{p_1} \eta_{r_1} \mathbf{W}_{1r_1})$, where $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_{p_1})'$. If all eigenvalues of $\mathbf{A}(\boldsymbol{\lambda}, \gamma, \boldsymbol{\eta})$ lie inside the unit circle, then $\mathbf{Y}_t$ in (2.3) is a stable process (Hamilton 1994, Proposition 10.1). Therefore, in the case of the dynamic specification, the required sufficient conditions for stability are (1^b) $\max_{1 \leq i \leq n} \left\{ \left| \vartheta_i (\sum_{r_1=1}^{p_1} \lambda_{r_1} \mathbf{W}_{1r_1}) \right| \right\} < 1$, (2^b) $\max_{1 \leq i \leq n} \left\{ \left| \vartheta_i (\sum_{r_2=1}^{p_2} \rho_{r_2} \mathbf{W}_{2r_2}) \right| \right\} < 1$, and (3^b) $\max_{1 \leq i \leq n} \{ |\vartheta_i (\mathbf{A}(\boldsymbol{\lambda}, \gamma, \boldsymbol{\eta}))| \} < 1$. By Lemma 2 in Appendix, the relatively restrictive sufficient conditions for (3^b) is given by (3^{*b}) $\| \mathbf{A}(\boldsymbol{\lambda}, \gamma, \boldsymbol{\eta}) \| < 1$. We can also investigate upper bounds for these conditions to formulate further restrictive, but easy-to-check conditions. Note that

$$\begin{aligned}
\| \mathbf{A}(\boldsymbol{\lambda}, \gamma, \boldsymbol{\eta}) \| &\leq \left\| \mathbf{S}_1^{-1}(\boldsymbol{\lambda}) \right\| \times \left\| \gamma \mathbf{I}_n + \sum_{r_1=1}^{p_1} \eta_{r_1} \mathbf{W}_{1r_1} \right\| \\
&= \left\| \mathbf{I}_n + \left(\sum_{r_1=1}^{p_1} \lambda_{r_1} \mathbf{W}_{1r_1} \right) + \left(\sum_{r_1=1}^{p_1} \lambda_{r_1} \mathbf{W}_{1r_1} \right)^2 + \ldots \right\| \times \left\| \gamma \mathbf{I}_n + \sum_{r_1=1}^{p_1} \eta_{r_1} \mathbf{W}_{1r_1} \right\| \\
&\leq \frac{1}{1 - \tau_1} \times \left(|\gamma| + \left(\sum_{r_1=1}^{p_1} |\eta_{r_1}| \right) \times \max_{1 \leq r_1 \leq p_1} \| \mathbf{W}_{1r_1} \| \right),
\end{aligned}$$

where $\tau_1 = \left(\sum_{r_1=1}^{p_1} |\lambda_{r_1}| \right) \times \max_{1 \leq r_1 \leq p_1} \| \mathbf{W}_{1r_1} \| < 1$. Then, the more restrictive sufficient conditions for the stability of the dynamic specification in (2.2) can be expressed as (1^{*b}) $\left(\sum_{r_1=1}^{p_1} |\lambda_{r_1}| \right) \times \max_{1 \leq r_1 \leq p_1} \| \mathbf{W}_{1r_1} \| < 1$, (2^{*b}) $\left(\sum_{r_2=1}^{p_2} |\rho_{r_2}| \right) \times \max_{1 \leq r_2 \leq p_2} \| \mathbf{W}_{2r_2} \| < 1$, and (3^{*b}) $\frac{1}{1-\tau_1} \times \left(|\gamma| + \left(\sum_{r_1=1}^{p_1} |\eta_{r_1}| \right) \times \max_{1 \leq r_1 \leq p_1} \| \mathbf{W}_{1r_1} \| \right) < 1$. Note that when spatial weights matrices are row-normalized, the above conditions further simplify. For example, using the matrix row sum norm, the conditions in $(1^{*b}) - (3^{*b})$ reduce to (1^{*b}) $\sum_{r_1=1}^{p_1} |\lambda_{r_1}| < 1$, (2^{*b}) $\sum_{r_2=1}^{p_2} |\rho_{r_2}| < 1$ and (3^{*b}) $\sum_{r_1=1}^{p_1} |\lambda_{r_1}| + \sum_{r_1=1}^{p_1} |\eta_{r_1}| + |\gamma| < 1$.

3 Observed-data DIC for Spatial Panel Data Models

In this section, we show how the DIC suggested by Spiegelhalter et al. (2002) can be used in model selection exercises involving the static and dynamic spatial panel data specifications in Section 2. In latent variable models, there can be alternative definitions of the DIC depending on the type of likelihood function considered in the formulation of the DIC (Celeux et al. 2006). Note that for the static and dynamic spatial panel data specifications in Section 2, the latent high-dimensional variables are $\mathbf{c}_n$ and $\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)'$ is the $T \times 1$ vector of time effects. Let $\boldsymbol{\Phi} = (\boldsymbol{\beta}', \sigma^2)'$, and $\boldsymbol{\Psi} = (\boldsymbol{\lambda}', \boldsymbol{\rho}')'$ be the $(p_1 + p_2) \times 1$ vector of spatial parameters in the case of (2.1), and $\boldsymbol{\Psi} = (\boldsymbol{\lambda}', \gamma, \boldsymbol{\eta}', \boldsymbol{\rho}')'$ be the $(2p_1 + p_2 + 1) \times 1$ vector of parameters in the case of (2.2). We

use $p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi},\mathbf{c}_n,\boldsymbol{\alpha})$ to denote the conditional likelihood function, and $p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi})$ to denote the observed data likelihood function (or the integrated likelihood function), where $\mathbf{Y}=(\mathbf{Y}_1^{'},\ldots,\mathbf{Y}_T^{'})^{'}$. Accordingly, we call the DIC formulated with $p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi},\mathbf{c}_n,\boldsymbol{\alpha})$ "the conditional DIC", and the DIC formulated with $p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi})$ "the observed-data DIC."

The conditional likelihood functions are readily available. We start with the conditional likelihood function of (2.1). Define $\mathbf{Y}_L=(\mathbf{I}_T\otimes\mathbf{S}_1(\boldsymbol{\lambda}))\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}-\boldsymbol{l}_T\otimes\mathbf{c}_n-\boldsymbol{\alpha}\otimes\boldsymbol{l}_n$, where $\mathbf{X}=(\mathbf{X}_1^{'},\ldots,\mathbf{X}_T^{'})^{'}$. Then, the conditional likelihood function of (2.1) can be written as

$$p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi},\mathbf{c}_n,\boldsymbol{\alpha})=(2\pi)^{-nT/2}(\sigma^2)^{-nT/2}\cdot|\mathbf{S}_1(\boldsymbol{\lambda})|^T\cdot|\mathbf{S}_2(\boldsymbol{\rho})|^T\times\exp\left(-\frac{1}{2}\mathbf{Y}_L^{'}\left(\mathbf{I}_T\otimes\mathbf{\Omega}(\mathbf{\Upsilon})\right)\mathbf{Y}_L\right), \tag{3.1}$$

where $\mathbf{\Omega}(\mathbf{\Upsilon})=\mathbf{S}_2^{'}(\boldsymbol{\rho})\mathbf{S}_2(\boldsymbol{\rho})/\sigma^2$ and $\mathbf{\Upsilon}=(\boldsymbol{\rho}^{'},\sigma^2)^{'}$. Next, we derive the conditional likelihood function of (2.2). To that end, we assume that the initial values at time $t=0$ in $\mathbf{Y}_0$ are exogenously given. Let $\mathbf{Y}^f=(\mathbf{I}_T\otimes\mathbf{S}_1(\boldsymbol{\lambda}))\mathbf{Y}-(\mathbf{I}_T\otimes\mathbf{R}(\gamma,\boldsymbol{\eta}))\mathbf{Y}_{-1}$, where $\mathbf{R}(\gamma,\boldsymbol{\eta})=\gamma\mathbf{I}_n+\sum_{r_1=1}^{p_1}\eta_{r_1}\mathbf{W}_{1r_1}$, $\mathbf{Y}=(\mathbf{Y}_1^{'},\ldots,\mathbf{Y}_T^{'})^{'}$ and $\mathbf{Y}_{-1}=(\mathbf{Y}_0^{'},\ldots,\mathbf{Y}_{T-1}^{'})^{'}$. Then, the conditional likelihood function of (2.2) is

$$\begin{aligned}p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi},\mathbf{c}_n,\boldsymbol{\alpha})&=(2\pi)^{-nT/2}(\sigma^2)^{-nT/2}\cdot|\mathbf{S}_1(\boldsymbol{\lambda})|^T\cdot|\mathbf{S}_2(\boldsymbol{\rho})|^T \\ &\times\exp\left(-\frac{1}{2}\left(\mathbf{Y}^f-\mathbf{X}\boldsymbol{\beta}-\boldsymbol{l}_T\otimes\mathbf{c}_n-\boldsymbol{\alpha}\otimes\boldsymbol{l}_n\right)^{'}\left(\mathbf{I}_T\otimes\mathbf{\Omega}(\mathbf{\Upsilon})\right)\left(\mathbf{Y}^f-\mathbf{X}\boldsymbol{\beta}-\boldsymbol{l}_T\otimes\mathbf{c}_n-\boldsymbol{\alpha}\otimes\boldsymbol{l}_n\right)\right),\end{aligned} \tag{3.2}$$

where $\mathbf{\Omega}(\mathbf{\Upsilon})=\mathbf{S}_2^{'}(\boldsymbol{\rho})\mathbf{S}_2(\boldsymbol{\rho})/\sigma^2$ and $\mathbf{\Upsilon}=(\boldsymbol{\rho}^{'},\sigma^2)$.

The conditional likelihood functions in (3.1) and (3.2) are useful in designing Gibbs samplers for the estimation of models, but they cannot be used to formulate DIC measures as they undermine the assumptions required for the decision-theoretical perspective of Li et al. (2017, 2020). Also, see Chan and Grant (2016a,b) and Millar (2009) for the simulation evidence on the poor performance of the DIC measure formulated with the conditional likelihood functions. To formulate the observed-data DIC measures, we need to derive the integrated likelihood functions. To that end, we assume the following independent prior distributions for $\mathbf{c}_n$ and $\boldsymbol{\alpha}$: $\mathbf{c}_n\sim N(\boldsymbol{\mu}_c,\mathbf{V}_c)$, $\alpha_t\sim N(\mu_\alpha,V_\alpha)$ for $t=1,\ldots,T$. Under these prior distributions, we can obtain the integrated likelihood functions as

$$p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi})=\int p(\mathbf{Y},\mathbf{c}_n,\boldsymbol{\alpha}|\mathbf{\Phi},\mathbf{\Psi})\,\mathrm{d}\mathbf{c}_n\mathrm{d}\boldsymbol{\alpha}=\int p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi},\mathbf{c}_n,\boldsymbol{\alpha})p(\mathbf{c}_n)p(\boldsymbol{\alpha})\,\mathrm{d}\mathbf{c}_n\mathrm{d}\boldsymbol{\alpha}, \tag{3.3}$$

where $p(\mathbf{Y},\mathbf{c}_n,\boldsymbol{\alpha}|\mathbf{\Phi},\mathbf{\Psi})$ is the complete-data likelihood function, $p(\mathbf{c}_n)$ is the prior distribution of $\mathbf{c}_n$, and $p(\boldsymbol{\alpha})$ is the prior distribution of $\boldsymbol{\alpha}$. In the following proposition, we provide $p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi})$ in the case of each model.

Proposition 1. *Let* $\mathbf{B}_n=(\mathbf{I}_n,\,\boldsymbol{l}_n)$, $\boldsymbol{\mu}_f=(\boldsymbol{\mu}_c^{'},\mu_\alpha)^{'}$ *and*

$$\mathbf{V}_f=\begin{pmatrix}\mathbf{V}_c & \mathbf{0}_{n\times1}\\ \mathbf{0}_{1\times n} & V_\alpha\end{pmatrix}.$$

Define $\mathbf{B} = \mathbf{I}_T \otimes \mathbf{B}_n$, $\boldsymbol{\mu}_F = \boldsymbol{l}_T \otimes \boldsymbol{\mu}_f$ *and* $\mathbf{V}_F = \mathbf{I}_T \otimes \mathbf{V}_f$.

(i) The observed-data likelihood function of the static specification in (2.1) *is*

$$p(\mathbf{Y}|\boldsymbol{\Phi},\boldsymbol{\Psi}) = (2\pi)^{-nT/2}(\sigma^2)^{-nT/2}\cdot|\mathbf{S}_1(\boldsymbol{\lambda})|^T\cdot|\mathbf{S}_2(\boldsymbol{\rho})|^T\cdot|\mathbf{V}_F|^{-1/2}\cdot|\mathbf{K}_F|^{-1/2} \tag{3.4}$$
$$\times\exp\left(-\frac{1}{2}\left(\mathbf{Y}^{F'}\left(\mathbf{I}_T\otimes\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\mathbf{Y}^F+\boldsymbol{\mu}_F^{'}\mathbf{V}_F^{-1}\boldsymbol{\mu}_F-\mathbf{k}_F^{'}\mathbf{K}_F^{-1}\mathbf{k}_F\right)\right),$$

where $\mathbf{Y}^F = (\mathbf{I}_T\otimes\mathbf{S}_1(\boldsymbol{\lambda}))\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$, $\mathbf{K}_F = \mathbf{B}^{'}\left(\mathbf{I}_T\otimes\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\mathbf{B} + \mathbf{V}_F^{-1}$ *and* $\mathbf{k}_F = \mathbf{B}^{'}\left(\mathbf{I}_T\otimes\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\mathbf{Y}^F + \mathbf{V}_F^{-1}\boldsymbol{\mu}_F$.

(ii) The observed-data likelihood function of the dynamic specification in (2.2) *is*

$$p(\mathbf{Y}|\boldsymbol{\Phi},\boldsymbol{\Psi}) = (2\pi)^{-nT/2}(\sigma^2)^{-nT/2}\cdot|\mathbf{S}_1(\boldsymbol{\lambda})|^T\cdot|\mathbf{S}_2(\boldsymbol{\rho})|^T\cdot|\mathbf{V}_F|^{-1/2}\cdot|\mathbf{K}_F|^{-1/2} \tag{3.5}$$
$$\times\exp\left(-\frac{1}{2}\left(\left(\mathbf{Y}^f-\mathbf{X}\boldsymbol{\beta}\right)^{'}\left(\mathbf{I}_T\otimes\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\left(\mathbf{Y}^f-\mathbf{X}\boldsymbol{\beta}\right)+\boldsymbol{\mu}_F^{'}\mathbf{V}_F^{-1}\boldsymbol{\mu}_F-\mathbf{k}_F^{'}\mathbf{K}_F^{-1}\mathbf{k}_F\right)\right),$$

where $\mathbf{K}_F = \mathbf{B}^{'}(\mathbf{I}_T\otimes\boldsymbol{\Omega}(\boldsymbol{\Upsilon}))\mathbf{B} + \mathbf{V}_F^{-1}$ *and* $\mathbf{k}_F = \mathbf{B}^{'}(\mathbf{I}_T\otimes\boldsymbol{\Omega}(\boldsymbol{\Upsilon}))(\mathbf{Y}^f-\mathbf{X}\boldsymbol{\beta}) + \mathbf{V}_F^{-1}\boldsymbol{\mu}_F$.

Proof. See Appendix B. □

Next, we use $p(\mathbf{Y}|\boldsymbol{\Phi},\boldsymbol{\Psi})$ to formulate the observed-data DIC measures. The Bayesian deviance term defined in Spiegelhalter et al. (2002) is given by $D(\boldsymbol{\Phi},\boldsymbol{\Psi}) = -2\log p(\mathbf{Y}|\boldsymbol{\Phi},\boldsymbol{\Psi}) + 2\log f(\mathbf{Y})$, where $f(\mathbf{Y})$ is some fully specified standardizing term which is a function of data alone. For model comparison exercises, we set $f(\mathbf{Y}) = 1$ (Berg et al. 2004). Then, the observed-data DIC takes the following form:

$$\text{DIC} = \overline{D}(\boldsymbol{\Phi},\boldsymbol{\Psi}) + P_D, \tag{3.6}$$

where $\overline{D}(\boldsymbol{\Phi},\boldsymbol{\Psi}) = -2\mathbb{E}\left(\log p(\mathbf{Y}|\boldsymbol{\Phi},\boldsymbol{\Psi})|\mathbf{Y}\right)$ is the posterior mean deviance and serves as a Bayesian measure of model fit. When the model is compatible with data, the likelihood attains larger values, leading to smaller values for $\overline{D}(\boldsymbol{\Phi},\boldsymbol{\Psi})$. Thus, the better the model fit is, the smaller $\overline{D}(\boldsymbol{\Phi},\boldsymbol{\Psi})$ is. The second term P_D is defined as the difference between the posterior mean deviance and the deviance at the estimated parameters, and is given by

$$P_D = \overline{D}(\boldsymbol{\Phi},\boldsymbol{\Psi}) - D(\tilde{\boldsymbol{\Phi}},\tilde{\boldsymbol{\Psi}}) = -2\mathbb{E}\left(\log p(\mathbf{Y}|\boldsymbol{\Phi},\boldsymbol{\Psi})|\mathbf{Y}\right) + 2\log p(\mathbf{Y}|\tilde{\boldsymbol{\Phi}},\tilde{\boldsymbol{\Psi}}),$$

where $\tilde{\boldsymbol{\Phi}}$ and $\tilde{\boldsymbol{\Psi}}$ are the joint maximum a posterior (MAP) estimators (Celeux et al. 2006).[6] P_D can be understood as a measure of the effective number of parameters in the model, i.e. it is a measure of model complexity. Gelman et al. (2003) consider a related way to measure model complexity

[6]The MAP estimates can be approximated by the posterior draws of tuples $(\boldsymbol{\Phi},\boldsymbol{\Psi})$ that yield the largest value for $p(\mathbf{Y}|\boldsymbol{\Phi},\boldsymbol{\Psi})p(\boldsymbol{\Phi})p(\boldsymbol{\Psi})$, where $p(\boldsymbol{\Phi})$ and $p(\boldsymbol{\Psi})$ are the prior density functions.

defined as two times the posterior variance of the likelihood function. This alternative always takes positive values and is given by $P_{DV} = 2 \times \text{Var}(\log p(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Psi})|\mathbf{Y})$. Thus, depending on the measure of model complexity, we have the following observed-data DIC measures:

$$\text{DIC}_1 = -4\mathbb{E}\left(\log p(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Psi})|\mathbf{Y}\right) + 2\log p(\mathbf{Y}|\tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}}), \tag{3.7}$$

$$\text{DIC}_2 = -2\mathbb{E}\left(\log p(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Psi})|\mathbf{Y}\right) + 2\text{Var}(\log p(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Psi})|\mathbf{Y}). \tag{3.8}$$

The posterior term $\mathbb{E}\left(\log p(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Psi})|\mathbf{Y}\right)$ in DIC_1 and DIC_2 can be estimated by averaging the log-integrated likelihood function $\log p(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Psi})$ over the posterior draws of $\mathbf{\Phi}$ and $\mathbf{\Psi}$. The term $\log p(\mathbf{Y}|\tilde{\mathbf{\Phi}}, \tilde{\mathbf{\Psi}})$ in DIC_1 is obtained by evaluating $\log p(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Psi})$ at $\tilde{\mathbf{\Phi}}$ and $\tilde{\mathbf{\Psi}}$, and the term $\text{Var}(\log p(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Psi})|\mathbf{Y})$ of DIC_2 is computed by taking the variance of $\log p(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Psi})$ over the posterior draws of $\mathbf{\Phi}$ and $\mathbf{\Psi}$.

4 Posterior Analysis

In this section, we provide the posterior analysis for the static and dynamic spatial panel data specifications in Section 2. We assume the following independent prior distributions:

$$\begin{aligned}
&\lambda_{r_1} \sim \text{Uniform}(-1, 1), \quad \eta_{r_1} \sim \text{Uniform}(-1, 1), \quad r_1 = 1, 2, \ldots, p_1, \\
&\rho_{r_2} \sim \text{Uniform}(-1, 1), \quad r_2 = 1, 2, \ldots, p_2, \quad \gamma \sim \text{Uniform}(-1, 1), \\
&\mathbf{c}_n \sim N(\boldsymbol{\mu}_c, \mathbf{V}_c), \quad \alpha_t \sim N(\mu_\alpha, V_\alpha), \quad t = 2, \ldots, T, \\
&\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta), \quad \sigma^2 \sim \text{IG}(a_0, b_0),
\end{aligned}$$

where $\text{Uniform}(-1, 1)$ denotes the uniform distribution over the interval $(-1, 1)$ and $\text{IG}(a_0, b_0)$ denotes the inverse gamma distribution with shape parameter a_0 and scale parameter b_0. The hyper-parameters can take the following values: $\boldsymbol{\mu}_c = \mathbf{0}$, $\mathbf{V}_c = 10\mathbf{I}_n$, $\mu_\alpha = 0$, $V_\alpha = 10$, $\boldsymbol{\mu}_\beta = \mathbf{0}$, $\mathbf{V}_\beta = 10\mathbf{I}_k$, $a_0 = 0.01$ and $b_0 = 0.01$ (Han et al. 2017).

Since we assume the conjugate priors for the common parameters $\boldsymbol{\beta}$, σ^2, $\mathbf{c}_n$ and $\boldsymbol{\alpha}$, the conditional posterior distributions of these parameters take known forms, indicating direct sampling for these parameters in each model. We leave the details of the sampler for these parameters to the Appendix C. However, the conditional posterior distribution of $\mathbf{\Psi}$ does not result in a known form in the context of each model. Therefore, we use the adaptive Metropolis (AM) algorithm suggested in Haario et al. (2001) and Roberts and Rosenthal (2009) to generate draws from the conditional posterior distribution $p(\mathbf{\Psi}|\mathbf{Y}, \mathbf{\Phi}, \mathbf{c}_n, \boldsymbol{\alpha})$.[7] In the AM algorithm, the historical MCMC draws of $\mathbf{\Psi}$ are used to determine the covariance matrix of proposal distributions. We summarize the steps required for this algorithm below.

Algorithm 1 (AM Algorithm for $\mathbf{\Psi}$)**.**

[7] Han and Lee (2016) and Han et al. (2017) use similar AM algorithms to generate draws for the spatial parameters in spatial panel data models. Their results show that the AM algorithm can perform satisfactorily.

1. Draw a candidate $\tilde{\boldsymbol{\Psi}}$ as follows: At the iteration g for $g = 1, 2, \ldots, G$,

   (a) if $g \leq 2\bar{p}$, propose $\tilde{\boldsymbol{\Psi}} \sim N\left(\boldsymbol{\Psi}^{(g-1)}, \frac{(0.1)^2}{\bar{p}} \times \mathbf{I}_{\bar{p}}\right)$, where $\bar{p} = p_1 + p_2$ for the static specification and $\bar{p} = 2p_1 + p_2 + 1$ for the dynamic specification,

   (b) if $g > 2\bar{p}$, propose $\tilde{\boldsymbol{\Psi}} \sim 0.95 \times N\left(\boldsymbol{\Psi}^{(g-1)}, \kappa\frac{(2.38)^2}{\bar{p}} \times \text{Cov}\left(\boldsymbol{\Psi}^{(0)}, \ldots, \boldsymbol{\Psi}^{(g-1)}\right)\right) + 0.05 \times N\left(\boldsymbol{\Psi}^{(g-1)}, \frac{(0.1)^2}{\bar{p}} \times \mathbf{I}_{\bar{p}}\right)$, which is a mixture of two normal distributions.

2. Check whether $\tilde{\boldsymbol{\Psi}}$ satisfies the stability conditions in Section 2. If not, draw a new candidate $\tilde{\boldsymbol{\Psi}}$ until it meets the stability conditions.

3. Set the acceptance probability to

$$\mathbb{P}(\boldsymbol{\Psi}^{(g-1)}, \tilde{\boldsymbol{\Psi}}) = \min\left(\frac{p(\mathbf{Y}|\boldsymbol{\Phi}^{(g-1)}, \tilde{\boldsymbol{\Psi}}, \mathbf{c}_n^{(g-1)}\boldsymbol{\alpha}^{(g-1)})}{p(\mathbf{Y}|\boldsymbol{\Phi}^{(g-1)}, \boldsymbol{\Psi}^{(g-1)}, \mathbf{c}_n^{(g-1)}\boldsymbol{\alpha}^{(g-1)})}, 1\right).$$

   Then, return $\tilde{\boldsymbol{\Psi}}$ with probability $\mathbb{P}(\boldsymbol{\Psi}^{(g-1)}, \tilde{\boldsymbol{\Psi}})$; otherwise return $\boldsymbol{\Psi}^{(g-1)}$.

4. Adjust the tuning parameter according to $\kappa^{(g)} = \kappa^{(g-1)}/1.05$ if the acceptance rate is less than 40 percent, and $\kappa^{(g)} = \kappa^{(g-1)} \times 1.05$ if the acceptance rate is greater than 60 percent.

Remark 1. *Note that the covariance of the proposal distribution in the AM algorithm is formulated with the historical draws of* $\boldsymbol{\Psi}$*. The proposal distribution consists of two components. The first component is* $N\left(\boldsymbol{\Psi}^{(g-1)}, \frac{(0.1)^2}{\bar{p}} \times \mathbf{I}_{\bar{p}}\right)$*, and is used to generate the candidate draws when* $g \leq 2\bar{p}$*. The second component is a mixture of two normal distributions, and is used when* $g > 2\bar{p}$*. The first mixture component is chosen with probability* 0.95*, and is given by* $N\left(\boldsymbol{\Psi}^{(g-1)}, \kappa\frac{(2.38)^2}{\bar{p}} \times Cov\left(\boldsymbol{\Psi}^{(0)}, \ldots, \boldsymbol{\Psi}^{(g-1)}\right)\right)$*, where* $Cov\left(\boldsymbol{\Psi}^{(0)}, \ldots, \boldsymbol{\Psi}^{(g-1)}\right) = \frac{1}{g}\sum_{j=0}^{g-1}\boldsymbol{\Psi}^{(j)}\boldsymbol{\Psi}^{(j)\prime} - \overline{\boldsymbol{\Psi}}^{(g-1)}\overline{\boldsymbol{\Psi}}^{(g-1)\prime}$ *with* $\overline{\boldsymbol{\Psi}}^{(g-1)} = \frac{1}{g}\sum_{j=0}^{g-1}\boldsymbol{\Psi}^{(j)}$*, and* κ *is a tuning parameter. The second mixture component* $N\left(\boldsymbol{\Psi}^{(g-1)}, \frac{(0.1)^2}{\bar{p}} \times \mathbf{I}_{\bar{p}}\right)$ *has a probability of* 0.05*, and prevents the algorithm to generate a singular covariance matrix. The candidate values generated in Step 1 are subject to the stability conditions in Section 2 as stated in Step 2. In Step 3, we choose the candidate* $\tilde{\boldsymbol{\Psi}}$ *generated in Step 1 with probability* $\mathbb{P}(\boldsymbol{\Psi}^{(g-1)}, \tilde{\boldsymbol{\Psi}})$*. Finally, in Step 4, we adjust the tuning parameter* κ *during the estimation such that the acceptance rate falls between* 40 *percent and* 60 *percent.*[8]

5 A Simulation Study

In this section, we examine the performance of the observed-data DIC measures in an extensive simulation study. To this end, we are interested in the performance of the observed-data DIC measures in two scenarios an empiricist is likely to face: (i) a weights matrix selection problem,

[8] Alternatively, in Step 2, after rejecting the candidate $\tilde{\boldsymbol{\Psi}}$ that does not satisfy the stability condition, the tuning parameter can be adjusted before drawing a new candidate.

and (ii) a static vs. dynamic model selection problem. We consider the following DGPs:

$$
\begin{aligned}
\text{M1}: \quad & \mathbf{Y}_t = \lambda_1 \mathbf{W}_{11}\mathbf{Y}_t + \mathbf{X}_t\boldsymbol{\beta} + \mathbf{c}_n + \alpha_t + \mathbf{U}_t, \quad \mathbf{U}_t = \rho_1 \mathbf{W}_{21}\mathbf{U}_t + \mathbf{V}_t, \\
\text{M2}: \quad & \mathbf{Y}_t = \lambda_1 \mathbf{W}_{11}\mathbf{Y}_t + \lambda_2 \mathbf{W}_{12}\mathbf{Y}_t + \mathbf{X}_t\boldsymbol{\beta} + \mathbf{c}_n + \alpha_t + \mathbf{U}_t, \quad \mathbf{U}_t = \rho_1 \mathbf{W}_{21}\mathbf{U}_t + \mathbf{V}_t \\
\text{M3}: \quad & \mathbf{Y}_t = \lambda_1 \mathbf{W}_{11}\mathbf{Y}_t + \gamma \mathbf{Y}_{t-1} + \eta_1 \mathbf{W}_{11}\mathbf{Y}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \mathbf{c}_n + \alpha_t + \mathbf{U}_t, \\
& \mathbf{U}_t = \rho_1 \mathbf{W}_{21}\mathbf{U}_t + \mathbf{V}_t, \\
\text{M4}: \quad & \mathbf{Y}_t = \lambda_1 \mathbf{W}_{11}\mathbf{Y}_t + \lambda_2 \mathbf{W}_{12}\mathbf{Y}_t + \gamma \mathbf{Y}_{t-1} + \eta_1 \mathbf{W}_{11}\mathbf{Y}_{t-1} + \eta_2 \mathbf{W}_{12}\mathbf{Y}_{t-1} \\
& \qquad + \mathbf{X}_t\boldsymbol{\beta} + \mathbf{c}_n + \alpha_t + \mathbf{U}_t, \quad \mathbf{U}_t = \rho_1 \mathbf{W}_{21}\mathbf{U}_t + \mathbf{V}_t,
\end{aligned}
$$

for $t = 1, 2, \ldots, T$. We let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, where X_{1it}'s and X_{2it}'s are drawn independently from $N(0,\, 2)$. For the parameters, we consider the following true values: $\lambda_1 = 0.6$, $\lambda_2 = 0.1$, $\gamma = 0.1$, $\eta_1 = -0.1$, $\eta_2 = -0.1$, $\rho_1 = 0.2$, $\beta_1 = \beta_2 = 1$, and $\sigma^2 = 1$. The time fixed effects are drawn independently from the standard normal distribution. The individual fixed effects are generated according to Mundlak (1978) as follows, $\mathbf{c}_n = 0.3\overline{\mathbf{X}}_1 + 0.3\overline{\mathbf{X}}_2 + \boldsymbol{\epsilon}$, where $\overline{\mathbf{X}}_1$ and $\overline{\mathbf{X}}_2$ are the empirical time-averages of $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively, and ϵ_i's are drawn independently from $N(0, 0.05)$.

For the spatial weights matrices, we consider rook and queen contiguity cases. We first generate a vector containing a random permutation of the integers from 1 to n without repeating elements. Then, we reshape this vector into an $k \times m$ rectangular lattice, where $m = n/k$. In the case of rook contiguity, we set $w_{ij} = 1$ if the j'th observation is adjacent (left/right/above or below) to the i'th observation on the lattice. In the case of queen contiguity, we set $w_{ij} = 1$ if the j'th observation is adjacent to, or shares a border with the i'th observation. We set $k = 10$, and row-normalize all spatial weights matrices.[9] We consider two sample sizes: $\{n = 50, T = 18\}$ and $\{n = 100, T = 6\}$ to represent the long and short panel cases, respectively. The MCMC estimation is based on 10000 draws with 5000 draws discarded as burn-ins. The number of resamples is set to 200 in each experiment. We set the initial value of the tuning parameter κ to 1.5 in the AM algorithm.

To explore the performance of the observed-data DICs in each scenario, we use histogram plots as suggested by Chan and Grant (2016b). More specifically, these histogram plots will present the empirical distribution of the difference between the observed-data DIC of the misspecified model and the observed-data DIC of the correct model, over 200 resamples. Since the model with the smallest DIC value is the preferred model, a non-negative value for the difference of the observed-data DICs would indicate that the correct model is favored. In other words, if the observed-data DIC performs satisfactorily, the majority of the mass should be placed on the non-negative values.

For the weights matrix selection problem, we consider four experiments. In the first experiment, we use M1 (a static spatial panel data model) with both $\mathbf{W}_{11}$ and $\mathbf{W}_{21}$ generated according to the queen contiguity case to generate 200 samples. In this experiment, the empiricist who is unaware of the true process has the dilemma to choose between the rook and queen contiguity based weights

[9] In Section D of Appendix, we provide some additional simulation results by considering a wider set of spatial weights matrices. We generate the candidate weights matrices based on (i) the 5-nearest neighbors scheme (K5), (ii) the 10-nearest neighbors scheme (K10), (iii) the group interaction scheme (Group), (iv) the rook scheme (Rook), and (v) the queen scheme (Queen).

matrices. We use each sample to fit M1 with true weights matrices (queen contiguity) and compute the observed-data DICs. Similarly, we use the same sample to fit M1 with the wrong weights matrices (rook contiguity) and compute the observed-data DICs. We then compute the difference between the observed-data DIC of the misspecified model (rook contiguity) and the observed-data DIC of the correct model (queen contiguity) over 200 samples and present them in histogram plots in Figure 1. We find that the observed-data DICs choose the correct model in all of the samples for the long and short panels. In the second experiment, we use M2 (a high-order static spatial panel data model) with $\mathbf{W}_{11}$, $\mathbf{W}_{12}$ and $\mathbf{W}_{21}$ generated according to the rook contiguity case to generate 200 samples. We again use each sample to fit M2 with the true weight matrices and the wrong weight matrices. Figure 2 shows the histograms of the difference between the observed-data DIC of the misspecified model (queen contiguity) and the observed-data DIC of the correct model (rook contiguity) over 200 samples. Again, we observed that for all samples, the observed-data DICs choose the correct model under both sample sizes.

In the third experiment, the true process is M3 (a dynamic spatial panel data model), and $\mathbf{W}_{11}$ and $\mathbf{W}_{21}$ are generated according to the rook contiguity case. Again, we generate 200 samples and fit M3 with both true weights matrices and the wrong weights matrices. The histogram plots in Figure 3 show the difference between the observed-data DIC of the misspecified model (queen contiguity) and the observed-data DIC of the correct model (rook contiguity) over 200 samples. For the long panel case $(n, T) = (50, 18)$, for almost all of the samples, the observed-data DICs report positive values (99.5%), choosing the correct specification. In the case of the short panel case $(n, T) = (100, 6)$, the the observed-data DICs favor the correct model in all samples. In the last experiment, we use M4 (a high order dynamic spatial panel data model) with $\mathbf{W}_{11}$, $\mathbf{W}_{12}$ and $\mathbf{W}_{21}$ generated according to the queen contiguity case. The histogram plots in Figure 4 show the difference between the observed-data DIC of the misspecified model (rook contiguity) and the observed-data DIC of the correct model (queen contiguity) over 200 samples. The observed-data DICs favor the correct model in all resamples for the long and short panels.

For the static vs. dynamic specification problem, we again consider four experiments. In the first experiment, the true process follows M1 (static specification) with both $\mathbf{W}_{11}$ and $\mathbf{W}_{21}$ generated according to the rook contiguity case. The empiricist who is unaware of the true process has the dilemma to choose between M1 and M3 (dynamic specification) for the empirical analysis. We again compute the difference between the observed-data DIC of the misspecified model (M3) and the observed-data DIC of the correct model (M1) over 200 samples, and present them in the histogram plots. The top panels in Figure 5 presents the result for the short panel case. We observe that the difference is positive for the majority of samples, 71% in the case of DIC_1 and 92% in the case of DIC_2, indicating that the observed-data DICs generally choose the correct model. In the second experiment, we use M2 (a static specification) with $\mathbf{W}_{11}$, $\mathbf{W}_{12}$ and $\mathbf{W}_{21}$ generated according to the queen contiguity case to generate 200 samples. The empiricist who is unaware of the true process has the dilemma to choose between M2 and M4 (a dynamic specification) for the empirical analysis. The bottom panels in Figure 5 presents the result for the short panel case. Again we find

that the observed-data DICs choose the correct model in majority of the samples, 93% in the case of DIC_1 and 88.5% in the case of DIC_2.

In the third experiment, we use M3 (a dynamic specification) with both $\mathbf{W}_{11}$ and $\mathbf{W}_{21}$ generated according to the queen contiguity to generate 200 samples. In this experiment, the empiricist needs to choose between M1 (a static specification) and M3 for the empirical analysis. The results in Figure 6 show the difference between the observed-data DIC of the misspecified model (M1) and the observed-data DIC of the correct model (M3) over 200 samples. In the long panel case, the observed-data DICs favor the correct model for in all samples. In the short panel case, for the majority of samples, 90% in the case of DIC_1 and 96.5% in the case of DIC_2, the observed-data DICs favor the correct model. In the final experiment, the true process follows M4 (a dynamic specification) with $\mathbf{W}_{11}$, $\mathbf{W}_{12}$ and $\mathbf{W}_{21}$ generated according to the rook contiguity case. In this case, the empiricist needs to choose between M4 and M2 (static specification) for the empirical analysis. The histograms in Figure 7 show the difference between the observed-data DIC of the misspecified model (M2) and the observed-data DIC of the correct model (M4) over 200 samples. In the long panel, we observe again that the observed-data DICs choose the correct model in almost all of the samples. In the short panel case, for the majority of samples, 98.5% in the case of DIC_1 and 97% in the case of DIC_2, the observed-data DICs favor the correct model.

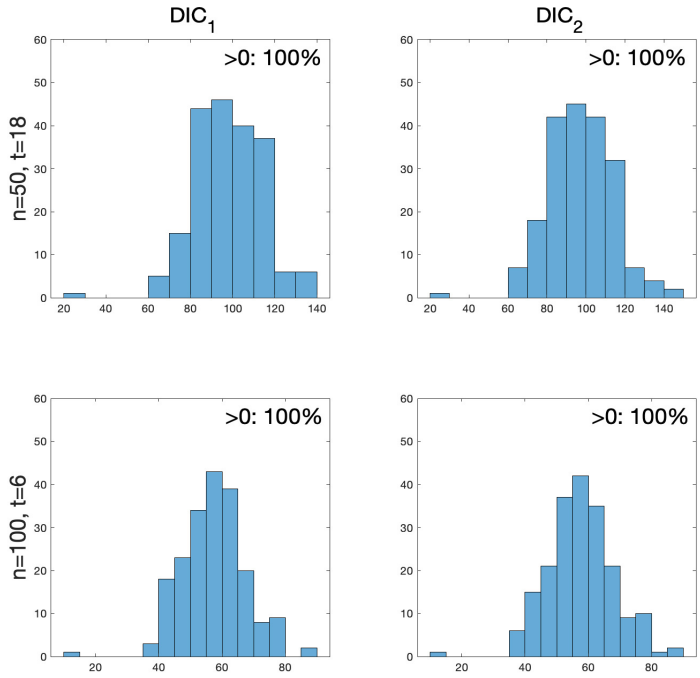

Figure 1: Histogram plots for the DIC of M1 with the rook case minus the DIC of M1 with the queen case.

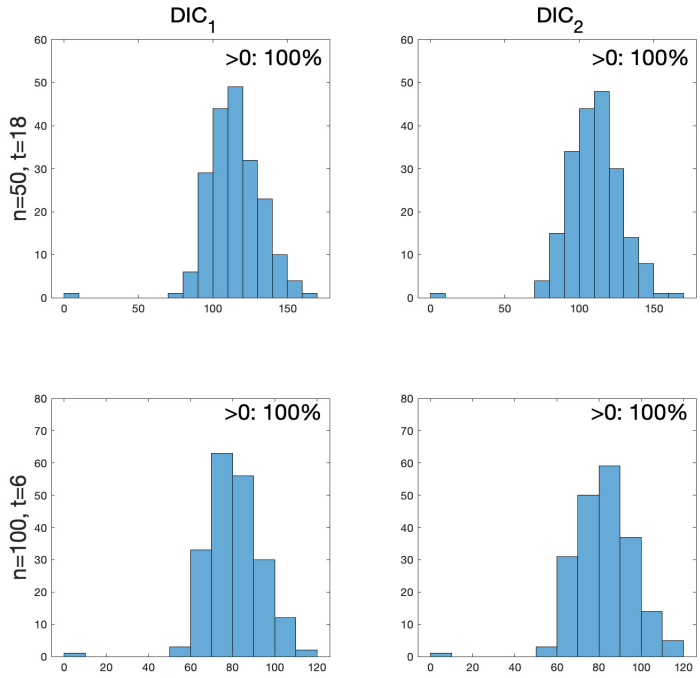

Figure 2: Histogram plots for the DIC of M2 with the queen case minus the DIC of M2 with the rook case.

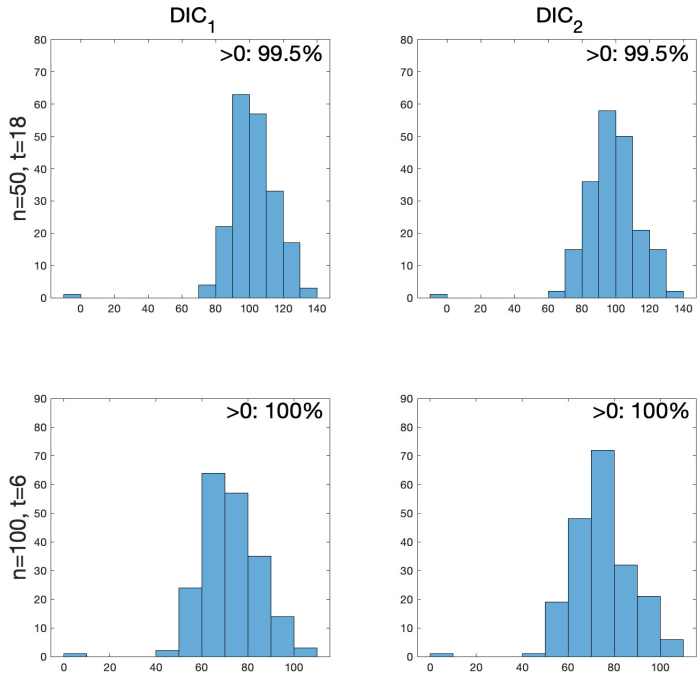

Figure 3: Histogram plots for the DIC of M3 with the queen case minus the DIC of M3 with the rook case.

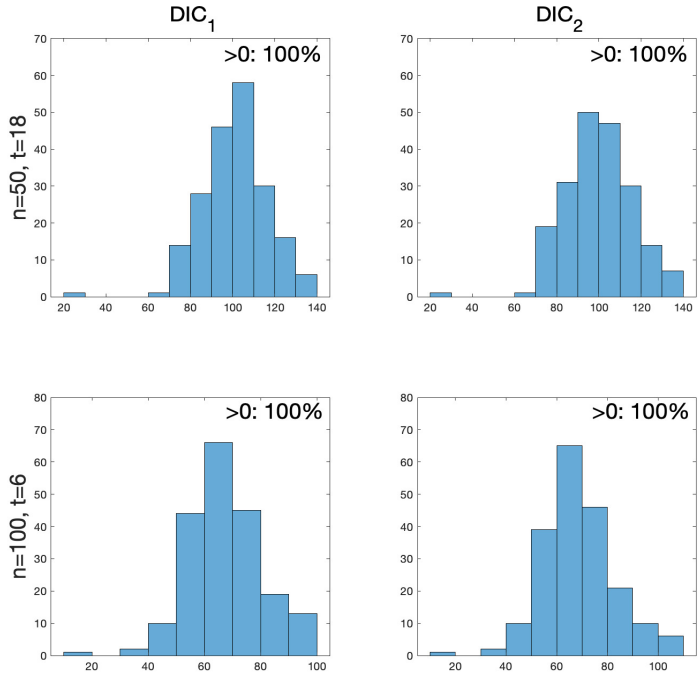

Figure 4: Histogram plots for the DIC of M4 with the rook case minus the DIC of M4 with the queen case.

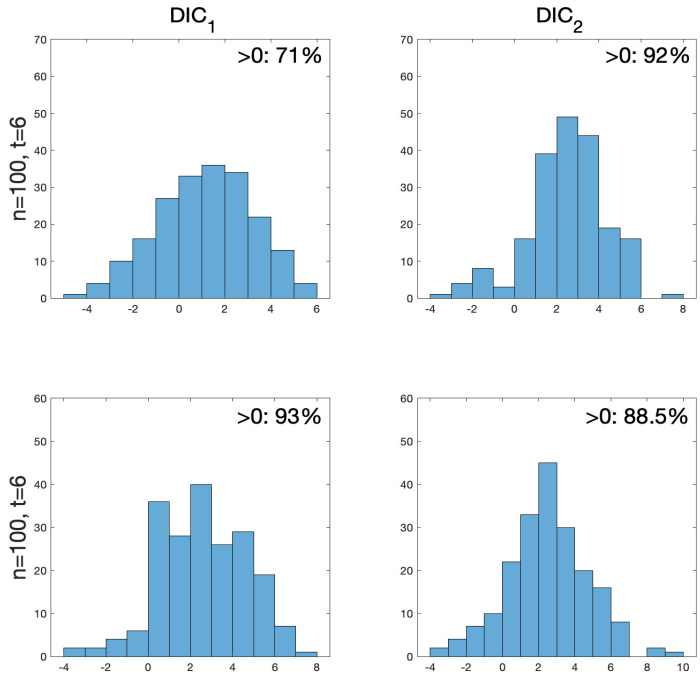

Figure 5: Top panel: Histogram plots for the DIC of M3 minus the DIC of M1. Bottom panel: Histogram plots for the DIC of M4 minus the DIC of M2

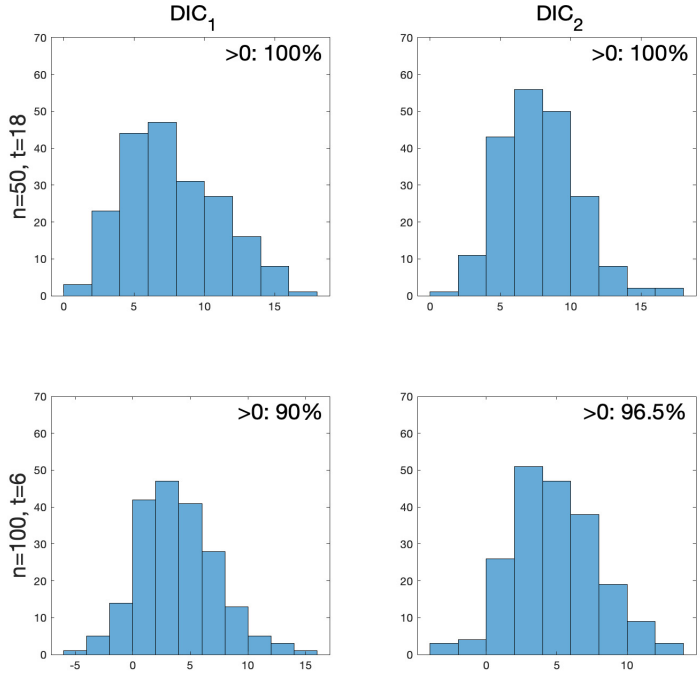

Figure 6: Histogram plots for the DIC of M1 minus the DIC of M3.

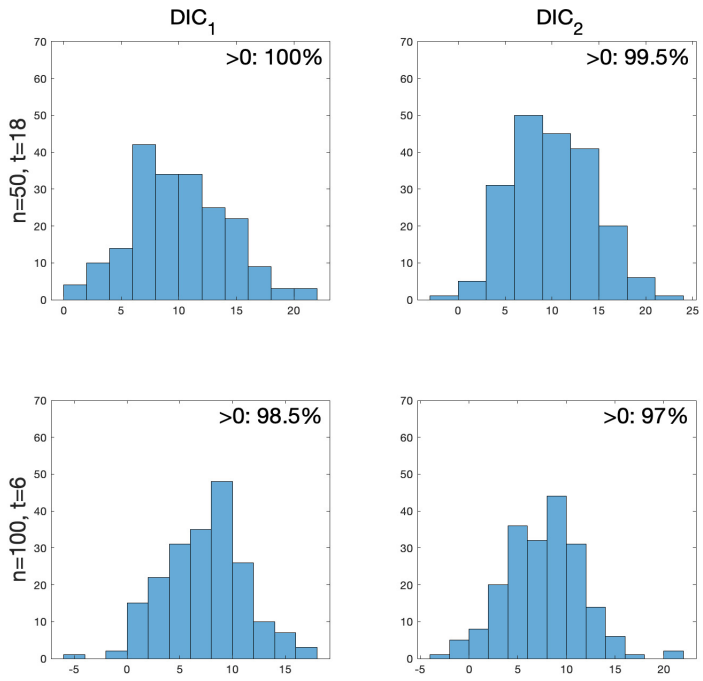

Figure 7: Histogram plots for the DIC of M2 minus the DIC of M4.

6 An Empirical Illustration

In this section, we provide an empirical illustration to flesh out how the observed-data DIC can be useful in model selection problems for spatial panel data models. To this end, we consider the empirical application on the US house prices in Aquaro et al. (2021). These authors put together a dataset for the US house price changes at the Metropolitan Statistical Areas (MSA) level between 1975 and 2014. Their estimation specification is a dynamic spatial panel data model which also allows for heterogenous coefficients and an unknown form of heteroskedasticity in the error terms. For the spatial weights matrix, they consider three candidates that are based on the distance between the MSAs. The geodesic distances between each pair of MSAs are computed, and the radius threshold of 75, 100 and 125 miles are used to generate the spatial weights matrices. We will denote them by $\mathbf{W}_{75}, \mathbf{W}_{100}$ and $\mathbf{W}_{125}$ respectively.

In this illustration, we are interested in utilizing the observed-data DIC to resolve (i) the spatial weights matrix selection problem, and (ii) the static vs. dynamic model selection problem. Following Yang (2021a) and Aquaro et al. (2021), we consider the following static and dynamic specifications:

$$y_{it} = \lambda \sum_{j=1}^{n} w_{ij} y_{jt} + \beta_1 \text{gpop}_{it} + \beta_2 \text{ginc}_{it} + c_i + \alpha_t + v_{it}, \tag{6.1}$$

$$y_{it} = \lambda \sum_{j=1}^{n} w_{ij} y_{jt} + \gamma y_{i,t-1} + \eta \sum_{j=1}^{n} w_{ij} y_{j,t-1} + \beta_1 \text{gpop}_{it} + \beta_2 \text{ginc}_{it} + c_i + \alpha_t + v_{it}, \tag{6.2}$$

where y_{it} is the percent quarterly rate of change of real house price in the MSA i in quarter t, gpop_{it} is the percent quarterly rate of change of population, and ginc_{it} is the percent quarterly rate of change in real capita income. The data set consists of 160 quarters and 377 MSAs.

We use Algorithms 2 and 3 given in Appendix C to estimate the static and dynamic specifications.[10] We present the estimation results for both specifications in Table 1. All parameter estimates are statistically significant and provide the same set of inference with those reported in Yang (2021a) and Aquaro et al. (2021). Our reported parameter estimates can be compared to those from the first two columns in Table 3 of Yang (2021a) for the static specification, and those from the last panel in Table 3 of Aquaro et al. (2021) for the dynamic specification. For example, for the static specification using $\mathbf{W}_{100}$, Yang (2021a) reports 0.730, 0.380 and 0.099 for λ, β_1 and β_2, respectively. For the dynamic specification using $\mathbf{W}_{75}$, Aquaro et al. (2021) reports 0.603, 0.667, -0.515, 0.250 and 0.050 for λ, γ, η, β_1 and β_2, respectively. Our estimates reported in Table 1 are consistent with these estimates.

The observed-data DIC estimates are presented in Table 2. We find that the smallest DIC values are observed in the case of $\mathbf{W}_{125}$ for both the static and dynamic specifications. Also, for a given spatial weights matrix, the observed-data DIC favors the dynamic specification over the

[10] We use the hyperparameter values given in Section 4. The estimation results are based on 10000 draws with 5000 draws discarded as burn-ins.

Table 1: Estimation results for the US house price models

	Static model			Dynamic model				
	λ	β_1	β_2	λ	γ	η	β_1	β_2
$\mathbf{W}_{75}$	0.588^{***}	0.412^{***}	0.134^{***}	0.629^{***}	0.652^{***}	-0.440^{***}	0.166^{***}	0.062^{***}
	(0.003)	(0.010)	(0.004)	(0.003)	(0.003)	(0.004)	(0.008)	(0.003)
$\mathbf{W}_{100}$	0.690^{***}	0.355^{***}	0.114^{***}	0.734^{***}	0.661^{***}	-0.520^{***}	0.145^{***}	0.052^{***}
	(0.002)	(0.009)	(0.004)	(0.002)	(0.003)	(0.004)	(0.007)	(0.003)
$\mathbf{W}_{125}$	0.748^{***}	0.350^{***}	0.111^{***}	0.796^{***}	0.664^{***}	-0.567^{***}	0.142^{***}	0.051^{***}
	(0.002)	(0.009)	(0.004)	(0.002)	(0.003)	(0.004)	(0.007)	(0.003)

Significance levels: ***, **, * denote respectively 1, 5 and 10 percent significance level.

static specification. Indeed, the estimates of γ and η are statistically significant in all cases. In lights of these points, we conclude that there is statistical evidence for choosing the dynamic spatial panel data model with $\mathbf{W}_{125}$ as the preferred model in this empirical exercise.

Table 2: Observed-data DICs for the US house price models

	DIC_1			DIC_2		
	$\mathbf{W}_{75}$	$\mathbf{W}_{100}$	$\mathbf{W}_{125}$	$\mathbf{W}_{75}$	$\mathbf{W}_{100}$	$\mathbf{W}_{125}$
Static model	270221	269465	267742	273247	272437	269952
Dynamic model	264917	265112	263124	269268	268802	266400

7 Conclusion

In this paper, we focused on two common specification problems in spatial panel data modeling, namely, choosing a spatial weights matrix from a set of candidates, and choosing between static and dynamic specifications. We proposed using observed-data DICs to settle these specification problems. Our approach has the advantage that the DIC measures are based on the integrated likelihood functions that result from analytically integrating out the latent variables from the complete data likelihood function. Therefore, there is no need for numerical schemes to obtain the integrated likelihood functions in our approach. Our simulation results attest that the resulting DIC measures, the observed-data DICs, perform well in finite samples. In future studies, our approach can be extended to more general specifications such as the spatial panel data model with interactive fixed effects, MESS-type specifications, specifications allowing for an unknown form of heteroskedasticity, and specifications allowing for endogeneity in weights matrices. We leave these extensions for future studies.

Appendix

A Some Useful Lemmas

The following lemmas are useful for the investigation of stability conditions.

Lemma 1. *Let* $\mathbf{S}(\boldsymbol{\lambda}) = (\mathbf{I}_n - \sum_{r=1}^{p} \lambda_r \mathbf{W}_r)$*, where* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^{'}$ *is the* $p \times 1$ *vector. Let* $\vartheta_i(\sum_{r=1}^{p} \lambda_r \mathbf{W}_r)$ *be the* i*th eigenvalue of* $(\sum_{r=1}^{p} \lambda_r \mathbf{W}_r)$ *for* $i = 1, \ldots, n$*. Then* $\mathbf{S}(\boldsymbol{\lambda})$ *is non-singular for all values of* $\boldsymbol{\lambda}$ *satisfying*

$$\max_{1 \leq i \leq n} \left\{ \left| \vartheta_i(\sum_{r=1}^{p} \lambda_r \mathbf{W}_r) \right| \right\} < 1. \tag{A.1}$$

Proof. Note that $\mathbf{S}(\boldsymbol{\lambda})$ is non-singular if and only if its determinant is non-zero, i.e., $|\mathbf{S}(\boldsymbol{\lambda})| \neq 0$. The determinant of $\mathbf{S}(\boldsymbol{\lambda})$ is

$$\begin{aligned} |\mathbf{S}(\boldsymbol{\lambda})| &= \left| \mathbf{I}_n - \sum_{r=1}^{p} \lambda_r \mathbf{W}_r \right| \\ &= \left(1 - \vartheta_1(\sum_{r=1}^{p} \lambda_r \mathbf{W}_r) \right) \times \left(1 - \vartheta_2(\sum_{r=1}^{p} \lambda_r \mathbf{W}_r) \right) \times \ldots \times \left(1 - \vartheta_n(\sum_{r=1}^{p} \lambda_r \mathbf{W}_r) \right). \end{aligned}$$

Thus, $|\mathbf{S}(\boldsymbol{\lambda})| \neq 0$ if and only if $\vartheta_i(\sum_{r=1}^{p} \lambda_r \mathbf{W}_r) \neq 1$ for $i = 1, 2, \ldots, n$. In particular, $\mathbf{S}(\boldsymbol{\lambda})$ is non-singular for all values of $\boldsymbol{\lambda}$ satisfying

$$\max \left\{ \left| \vartheta_1(\sum_{r=1}^{p} \lambda_r \mathbf{W}_r) \right|, \left| \vartheta_2(\sum_{r=1}^{p} \lambda_r \mathbf{W}_r) \right|, \ldots, \left| \vartheta_n(\sum_{r=1}^{p} \lambda_r \mathbf{W}_r) \right| \right\} < 1. \tag{A.2}$$

□

Lemma 2. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix and* $\vartheta_i(\mathbf{A})$ *be the* i*th eigenvalue of* $\mathbf{A}$*. Then,*

$$\max_{1 \leq i \leq n} \{|\vartheta_i(\mathbf{A})|\} \leq \|\mathbf{A}\|, \tag{A.3}$$

where $\|\cdot\|$ *denotes any matrix norm.*

Proof. See Horn and Johnson (2013, Theorem 5.6.9). □

B Proof of Proposition 1

In this section, we provide the proof of Proposition 1. We first introduce some notations. Let $\boldsymbol{\theta} = (\boldsymbol{\Phi}^{'}, \boldsymbol{\Psi}^{'}, \mathbf{c}_n^{'}, \boldsymbol{\alpha}^{'})^{'}$, $\mathbf{B}_n = (\mathbf{I}_n, \boldsymbol{l}_n)$, $\mathbf{B} = (\mathbf{I}_T \otimes \mathbf{B}_n)$ and $\mathbf{F}_t = (\mathbf{c}_n^{'}, \alpha_t)^{'}$. Then, the prior distribution

of $\mathbf{F}_t$ is $N(\boldsymbol{\mu}_f, \mathbf{V}_f)$, where

$$\boldsymbol{\mu}_f = \begin{pmatrix} \boldsymbol{\mu}_c \\ \mu_\alpha \end{pmatrix}, \quad \text{and} \quad \mathbf{V}_f = \begin{pmatrix} \mathbf{V}_c & \mathbf{0}_{n\times 1} \\ \mathbf{0}_{1\times n} & V_\alpha \end{pmatrix}.$$

Define $\mathbf{F} = (\mathbf{F}_1^{'}, \ldots, \mathbf{F}_T^{'})^{'}$, which has the prior distribution $N(\boldsymbol{\mu}_F, \mathbf{V}_F)$, where $\boldsymbol{\mu}_F = \boldsymbol{l}_T \otimes \boldsymbol{\mu}_f$ and $\mathbf{V}_F = \mathbf{I}_T \otimes \mathbf{V}_f$. Note that $\mathbf{BF} = \boldsymbol{l}_T \otimes \mathbf{c}_n + \boldsymbol{\alpha} \otimes \boldsymbol{l}_n$, which plays an important role in our derivation.

We start with showing the first part. Let $\mathbf{Y}^F = (\mathbf{I}_T \otimes \mathbf{S}_1(\boldsymbol{\lambda}))\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} = (\mathbf{Y}_1^{F'}, \ldots, \mathbf{Y}_T^{F'})'$, where $\mathbf{Y}_t^F = \mathbf{S}_1(\boldsymbol{\lambda})\mathbf{Y}_t - \mathbf{X}_t\boldsymbol{\beta}$. In terms of this new notation, the likelihood function of (2.1) can be expressed as

$$p(\mathbf{Y}|\boldsymbol{\theta}) = c_1 \times \exp\left(-\frac{1}{2}\left(\mathbf{Y}^F - \mathbf{BF}\right)^{'}\left(\mathbf{I}_T \otimes \boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\left(\mathbf{Y}^F - \mathbf{BF}\right)\right), \tag{B.1}$$

where $c_1 = (2\pi)^{-nT/2}(\sigma^2)^{-nT/2} \cdot |\mathbf{S}_1(\boldsymbol{\lambda})|^T \cdot |\mathbf{S}_2(\boldsymbol{\rho})|^T$. Then, the integrated likelihood function is

$$\begin{aligned} p(\mathbf{Y}|\boldsymbol{\Phi}, \boldsymbol{\Psi}) =& \int p(\mathbf{Y}|\boldsymbol{\Phi}, \boldsymbol{\Psi}, \mathbf{F})p(\mathbf{F})\mathrm{d}\mathbf{F} \\ =& c_2 \int \exp\left(-\frac{1}{2}\left(\mathbf{Y}^F - \mathbf{BF}\right)^{'}\left(\mathbf{I}_T \otimes \boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\left(\mathbf{Y}^F - \mathbf{BF}\right) - \frac{1}{2}(\mathbf{F} - \boldsymbol{\mu}_F)^{'}\mathbf{V}_F^{-1}(\mathbf{F} - \boldsymbol{\mu}_F)\right)\mathrm{d}\mathbf{F}, \end{aligned}$$

where $c_2 = (2\pi)^{-(nT+T/2)}(\sigma^2)^{-nT/2} \cdot |\mathbf{S}_1(\boldsymbol{\lambda})|^T \cdot |\mathbf{S}_2(\boldsymbol{\rho})|^T \cdot |\mathbf{V}_F|^{-1/2}$. Thus, our aim is to integrate out $\mathbf{F}$. Then, $p(\mathbf{Y}|\boldsymbol{\Phi}, \boldsymbol{\Psi})$ can be written as

$$p(\mathbf{Y}|\boldsymbol{\Phi}, \boldsymbol{\Psi}) = c_2 \int \exp\left(-\frac{1}{2}\left(\mathbf{Y}^{F'}\left(\mathbf{I}_T \otimes \boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\mathbf{Y}^F + \mathbf{F}^{'}\mathbf{K}_F\mathbf{F} - 2\mathbf{F}^{'}\mathbf{k}_F + \boldsymbol{\mu}_F^{'}\mathbf{V}_F^{-1}\boldsymbol{\mu}_F\right)\right)\mathrm{d}\mathbf{F},$$

where $\mathbf{K}_F = \mathbf{B}^{'}\left(\mathbf{I}_T \otimes \boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\mathbf{B} + \mathbf{V}_F^{-1}$ and $\mathbf{k}_F = \mathbf{B}^{'}\left(\mathbf{I}_T \otimes \boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\mathbf{Y}^F + \mathbf{V}_F^{-1}\boldsymbol{\mu}_F$. By completion of square, we have

$$\begin{aligned} \mathbf{F}^{'}\mathbf{K}_F\mathbf{F} - 2\mathbf{F}^{'}\mathbf{k}_F &= \mathbf{F}^{'}\mathbf{K}_F\mathbf{F} - 2\mathbf{F}^{'}\mathbf{k}_F + \mathbf{k}_F^{'}\mathbf{K}_F^{-1}\mathbf{k}_F - \mathbf{k}_F^{'}\mathbf{K}_F^{-1}\mathbf{k}_F \\ &= (\mathbf{F} - \mathbf{K}_F^{-1}\mathbf{k}_F)^{'}\mathbf{K}_F(\mathbf{F} - \mathbf{K}_F^{-1}\mathbf{k}_F) - \mathbf{k}_F^{'}\mathbf{K}_F^{-1}\mathbf{k}_F. \end{aligned} \tag{B.2}$$

Using (B.2) in $p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi})$, we derive the closed form of $p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi})$ in the following way:

$$\begin{aligned}
p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi}) &= c_2 \exp\left(-\frac{1}{2}\left(\mathbf{Y}^{F'}\left(\mathbf{I}_T \otimes \mathbf{\Omega}(\mathbf{\Upsilon})\right)\mathbf{Y}^F + \boldsymbol{\mu}_F^{'}\mathbf{V}_F^{-1}\boldsymbol{\mu}_F - \mathbf{k}_F^{'}\mathbf{K}_F^{-1}\mathbf{k}_F\right)\right) \\
&\times \int \exp\left(-\frac{1}{2}\left((\mathbf{F}-\mathbf{K}_F^{-1}\mathbf{k}_F)^{'}\mathbf{K}_F(\mathbf{F}-\mathbf{K}_F^{-1}\mathbf{k}_F)\right)\right)\mathrm{d}\mathbf{F} \\
&= c_2(2\pi)^{T(n+1)/2}|\mathbf{K}_F|^{-1/2}\exp\left(-\frac{1}{2}\left(\mathbf{Y}^{F'}\left(\mathbf{I}_T \otimes \mathbf{\Omega}(\mathbf{\Upsilon})\right)\mathbf{Y}^F + \boldsymbol{\mu}_F^{'}\mathbf{V}_F^{-1}\boldsymbol{\mu}_F - \mathbf{k}_F^{'}\mathbf{K}_F^{-1}\mathbf{k}_F\right)\right) \\
&= (2\pi)^{-nT/2}(\sigma^2)^{-nT/2}\cdot|\mathbf{S}_1(\boldsymbol{\lambda})|^T\cdot|\mathbf{S}_2(\boldsymbol{\rho})|^T\cdot|\mathbf{V}_F|^{-1/2}\cdot|\mathbf{K}_F|^{-1/2} \\
&\times \exp\left(-\frac{1}{2}\left(\sum_{t=1}^{T}\mathbf{Y}_t^{F'}\mathbf{\Omega}(\mathbf{\Upsilon})\mathbf{Y}_t^F + T\boldsymbol{\mu}_f^{'}\mathbf{V}_f^{-1}\boldsymbol{\mu}_f - \sum_{t=1}^{T}(\mathbf{B}_n^{'}\mathbf{\Omega}(\mathbf{\Upsilon})\mathbf{Y}_t^F + \mathbf{V}_f^{-1}\mu_f)^{'}\right.\right. \\
&\left.\left.(\mathbf{B}_n^{'}\mathbf{\Omega}(\mathbf{\Upsilon})\mathbf{B}_n + \mathbf{V}_f^{-1})^{-1}(\mathbf{B}_n^{'}\mathbf{\Omega}(\mathbf{\Upsilon})\mathbf{Y}_t^F + \mathbf{V}_f^{-1}\mu_f)\right)\right).
\end{aligned}$$

Note that the second equality above follows from the fact that $\int \exp\left(-\frac{1}{2}\left((\mathbf{F}-\mathbf{K}_F^{-1}\mathbf{k}_F)^{'}\mathbf{K}_F(\mathbf{F}-\mathbf{K}_F^{-1}\mathbf{k}_F)\right)\right)\mathrm{d}\mathbf{F} = (2\pi)^{T(n+1)/2}|\mathbf{K}_F|^{-1/2}$. Finally, in the case of (2.2), we use the likelihood function in (3.2), which can be expressed as

$$p(\mathbf{Y}|\boldsymbol{\theta}) = c_1 \times \exp\left(-\frac{1}{2}\left(\mathbf{Y}^F - \mathbf{BF}\right)^{'}\left(\mathbf{I}_T \otimes \mathbf{\Omega}(\mathbf{\Upsilon})\right)\left(\mathbf{Y}^F - \mathbf{BF}\right)\right) \tag{B.3}$$

where $c_1 = (2\pi)^{-nT/2}(\sigma^2)^{-nT/2}|\mathbf{S}_2(\boldsymbol{\rho})|^T$ and $\mathbf{Y}^F = (\mathbf{Y}^f - \mathbf{X}\boldsymbol{\beta}) = (\mathbf{I}_T \otimes \mathbf{S}_1(\boldsymbol{\lambda}))\mathbf{Y} - (\mathbf{I}_T \otimes \mathbf{R}(\gamma,\boldsymbol{\eta}))\mathbf{Y}_{-1} - \mathbf{X}\boldsymbol{\beta} = (\mathbf{Y}_1^F,\ldots,\mathbf{Y}_T^F)$, where $\mathbf{Y}_t^F = \mathbf{S}_1(\boldsymbol{\lambda})\mathbf{Y}_t - \mathbf{R}(\gamma,\boldsymbol{\eta})\mathbf{Y}_{-1,t} - \mathbf{X}_t\boldsymbol{\beta}$. Since (B.3) is in the form of (B.1), we can apply the same argument to show that

$$\begin{aligned}
p(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi}) &= (2\pi)^{-nT/2}(\sigma^2)^{-nT/2}\cdot|\mathbf{S}_1(\boldsymbol{\lambda})|^T\cdot|\mathbf{S}_2(\boldsymbol{\rho})|^T\cdot|\mathbf{V}_F|^{-1/2}\cdot|\mathbf{K}_F|^{-1/2} \\
&\times \exp\left(-\frac{1}{2}\left(\sum_{t=1}^{T}\mathbf{Y}_t^{F'}\mathbf{\Omega}(\mathbf{\Upsilon})\mathbf{Y}_t^F + T\boldsymbol{\mu}_f^{'}\mathbf{V}_f^{-1}\boldsymbol{\mu}_f - \sum_{t=1}^{T}(\mathbf{B}_n^{'}\mathbf{\Omega}(\mathbf{\Upsilon})\mathbf{Y}_t^F + \mathbf{V}_f^{-1}\mu_f)^{'}\right.\right. \\
&\left.\left.(\mathbf{B}_n^{'}\mathbf{\Omega}(\mathbf{\Upsilon})\mathbf{B}_n + \mathbf{V}_f^{-1})^{-1}(\mathbf{B}_n^{'}\mathbf{\Omega}(\mathbf{\Upsilon})\mathbf{Y}_t^F + \mathbf{V}_f^{-1}\mu_f)\right)\right).
\end{aligned}$$

C Details of the Posterior Analysis

In this section, we provide estimation algorithms for our main specifications.

Algorithm 2 (Estimation Algorithm for (2.1))**.**

1. Sampling step for $\mathbf{c}_n$: Let $\mathbf{Y}^c = (\mathbf{Y}_1^{c'},\ldots,\mathbf{Y}_T^{c'})^{'}$, where $\mathbf{Y}_t^c = \mathbf{S}_1(\boldsymbol{\lambda})\mathbf{Y}_t - \mathbf{X}_t\boldsymbol{\beta} - \alpha_t\boldsymbol{l}_n$ for $t = 1,2,\ldots,T$. Then,

$$\mathbf{c}_n|\mathbf{Y},\mathbf{\Phi},\mathbf{\Psi},\boldsymbol{\alpha} \sim N(\widehat{\boldsymbol{\mu}}_c,\widehat{\mathbf{V}}_c), \tag{C.1}$$

where $\widehat{\mathbf{V}}_c = \left(\mathbf{V}_c^{-1} + T\mathbf{\Omega}(\mathbf{\Upsilon})\right)^{-1}$, $\widehat{\boldsymbol{\mu}}_c = \widehat{\mathbf{V}}_c\left(\mathbf{V}_c^{-1}\boldsymbol{\mu}_c + \mathbf{\Omega}(\mathbf{\Upsilon})\sum_{t=1}^{T}\mathbf{Y}_t^c\right)$, and $\mathbf{\Omega}(\mathbf{\Upsilon}) = \mathbf{S}_2^{'}(\boldsymbol{\rho})\mathbf{S}_2(\boldsymbol{\rho})/\sigma^2$.

2. Sampling step for $\boldsymbol{\alpha}$: Let $\mathbf{Y}^{\alpha} = (\mathbf{Y}_1^{\alpha'}, \ldots, \mathbf{Y}_T^{\alpha'})'$, where $\mathbf{Y}_t^{\alpha} = \mathbf{S}_1(\boldsymbol{\lambda})Y_t - \mathbf{X}_t\boldsymbol{\beta} - \mathbf{c}_n$ for $t = 1, 2, \ldots, T$. Then,

$$\alpha_t|\mathbf{Y}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \mathbf{c}_n, \boldsymbol{\alpha}_{-t} \sim N(\widehat{\mu}_{\alpha_t}, \widehat{V}_{\alpha_t}), \quad t = 2, \ldots, T,$$

where $\widehat{V}_{\alpha_t} = \left(V_{\alpha}^{-1} + \boldsymbol{l}_n'\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\boldsymbol{l}_n\right)^{-1}$, $\widehat{\mu}_{\alpha_t} = \widehat{V}_{\alpha t}\left(V_{\alpha}^{-1}\mu_{\alpha} + \boldsymbol{l}_n'\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\mathbf{Y}_t^{\alpha}\right)$, and $\boldsymbol{\Omega}(\boldsymbol{\Upsilon}) = \mathbf{S}_2'(\boldsymbol{\rho})\mathbf{S}_2(\boldsymbol{\rho})/\sigma^2$.

3. Sampling step for $\boldsymbol{\beta}$: Let $\mathbf{Y}_t^{\beta} = \mathbf{S}_1(\boldsymbol{\lambda})\mathbf{Y}_t - \mathbf{c}_n - \alpha_t\boldsymbol{l}_n$. Then,

$$\boldsymbol{\beta}|\mathbf{Y}, \varrho, \sigma^2, \boldsymbol{\Psi}, \mathbf{c}_n, \boldsymbol{\alpha} \sim N(\widehat{\boldsymbol{\mu}}_{\beta}, \widehat{\mathbf{V}}_{\beta}),$$

where $\widehat{\mathbf{V}}_{\beta} = \left(\mathbf{V}_{\beta}^{-1} + \sum_{t=1}^{T}\mathbf{X}_t'\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\mathbf{X}_t\right)^{-1}$, $\widehat{\boldsymbol{\mu}}_{\beta} = \widehat{\mathbf{V}}_{\beta}\left(\mathbf{V}_{\beta}^{-1}\boldsymbol{\mu}_{\beta} + \sum_{t=1}^{T}\mathbf{X}_t'\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\mathbf{Y}_t^{\beta}\right)$, and $\boldsymbol{\Omega}(\boldsymbol{\Upsilon}) = \mathbf{S}_2'(\boldsymbol{\rho})\mathbf{S}_2(\boldsymbol{\rho})/\sigma^2$.

4. Sampling step for σ^2:

$$\sigma^2|\mathbf{Y}, \boldsymbol{\beta}, \varrho, \boldsymbol{\Psi}, \mathbf{c}_n, \boldsymbol{\alpha} \sim \mathrm{IG}\left(a, b\right), \tag{C.2}$$

where $a = a_0 + nT/2$, $b = b_0 + \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{Y}_t^{\beta} - \mathbf{X}_t\boldsymbol{\beta}\right)'\left(\sigma^2\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\left(\mathbf{Y}_t^{\beta} - \mathbf{X}_t\boldsymbol{\beta}\right)$, and $\boldsymbol{\Omega}(\boldsymbol{\Upsilon}) = \mathbf{S}_2'(\boldsymbol{\rho})\mathbf{S}_2(\boldsymbol{\rho})/\sigma^2$.

5. Sampling step for $\boldsymbol{\Psi}$: Use Algorithm 1 to sample $\boldsymbol{\Psi}$.

Algorithm 3 (Estimation Algorithm for (2.2))**.**

1. Sampling step for $\mathbf{c}_n$: Let $\mathbf{Y}_t^{c} = \mathbf{Y}_t^{f} - \mathbf{X}_t\boldsymbol{\beta} - \alpha_t\boldsymbol{l}_n$ for $t = 1, 2, \ldots, T$. Then,

$$\mathbf{c}_n|\mathbf{Y}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\alpha} \sim N(\widehat{\boldsymbol{\mu}}_c, \widehat{\mathbf{V}}_c), \tag{C.3}$$

where $\widehat{\mathbf{V}}_c = \left(\mathbf{V}_c^{-1} + T\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)^{-1}$, $\widehat{\boldsymbol{\mu}}_c = \widehat{\mathbf{V}}_c\left(\mathbf{V}_c^{-1}\boldsymbol{\mu}_c + \boldsymbol{\Omega}(\boldsymbol{\Upsilon})\sum_{t=1}^{T}\mathbf{Y}_t^{c}\right)$, and $\boldsymbol{\Omega}(\boldsymbol{\Upsilon}) = \mathbf{S}_2'(\boldsymbol{\rho})\mathbf{S}_2(\boldsymbol{\rho})/\sigma^2$.

2. Sampling step for $\boldsymbol{\alpha}$: Let $\mathbf{Y}_t^{\alpha} = \mathbf{Y}_t^{f} - \mathbf{X}_t\boldsymbol{\beta} - \mathbf{c}_n$ for $t = 1, 2, \ldots, T$. Then,

$$\alpha_t|\mathbf{Y}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \mathbf{c}_n \sim N(\widehat{\mu}_{\alpha}, \widehat{V}_{\alpha}), \quad t = 2, \ldots, T,$$

where $\widehat{V}_{\alpha} = \left(V_{\alpha}^{-1} + \boldsymbol{l}_n'\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\boldsymbol{l}_n\right)^{-1}$, $\widehat{\mu}_{\alpha} = \widehat{V}_{\alpha}\left(V_{\alpha}^{-1}\mu_{\alpha} + \boldsymbol{l}_n'\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\mathbf{Y}_t^{\alpha}\right)$, and $\boldsymbol{\Omega}(\boldsymbol{\Upsilon}) = \mathbf{S}_2'(\boldsymbol{\rho})\mathbf{S}_2(\boldsymbol{\rho})/\sigma^2$.

3. Sampling step for β: Let $\mathbf{Y}^{\beta} = \mathbf{Y}_t^{f} - \mathbf{c}_n - \alpha_t\boldsymbol{l}_n$. Then,

$$\boldsymbol{\beta}|\mathbf{Y}, \sigma^2, \boldsymbol{\Psi}, \mathbf{c}_n, \boldsymbol{\alpha} \sim N(\widehat{\boldsymbol{\mu}}_{\beta}, \widehat{\mathbf{V}}_{\beta}),$$

where $\widehat{\mathbf{V}}_\beta = \left(\mathbf{V}_\beta^{-1} + \sum_{t=1}^T \mathbf{X}_t'\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\mathbf{X}_t\right)^{-1}$, $\widehat{\boldsymbol{\mu}}_\beta = \widehat{\mathbf{V}}_\beta\left(\mathbf{V}_\beta^{-1}\boldsymbol{\mu}_\beta + \sum_{t=1}^T \mathbf{X}_t'\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\mathbf{Y}_t^\beta\right)$, and $\boldsymbol{\Omega}(\boldsymbol{\Upsilon}) = \mathbf{S}_2'(\boldsymbol{\rho})\mathbf{S}_2(\boldsymbol{\rho})/\sigma^2$.

4. Sampling step for σ^2:

$$\sigma^2|\mathbf{Y}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \mathbf{c}_n, \boldsymbol{\alpha} \sim \mathrm{IG}\left(a, b\right), \tag{C.4}$$

where $a = a_0 + nT/2$, $b = b_0 + \frac{1}{2}\sum_{t=1}^T \left(\mathbf{Y}_t^\beta - \mathbf{X}_t\boldsymbol{\beta}\right)' \left(\sigma^2\boldsymbol{\Omega}(\boldsymbol{\Upsilon})\right)\left(\mathbf{Y}_t^\beta - \mathbf{X}_t\boldsymbol{\beta}\right)$, and $\boldsymbol{\Omega}(\boldsymbol{\Upsilon}) = \mathbf{S}_2'(\boldsymbol{\rho})\mathbf{S}_2(\boldsymbol{\rho})/\sigma^2$.

5. Sampling step for $\boldsymbol{\Psi}$: Use Algorithm 1 to sample $\boldsymbol{\Psi}$.

D Some Additional Simulation Results

In this section, we provide some additional simulation results by considering a wider set of spatial weights matrices. The candidate weights matrices are generated based on the (i) the 5-nearest neighbors scheme (K5), (ii) the 10-nearest neighbors scheme (K10), (iii) the group interaction scheme (Group), (iv) the rook scheme (Rook), and (v) the queen scheme (Queen). We consider a data generating processes with the spatial weight matrix based on the 5-nearest neighbors scheme (K5). We provide the histogram plots that show the difference between the observed-data DIC of the misspecified model and that of the correct model. Figures E.1 and E.2 contain the results over 200 samples. All of differences are positive, indicating that the observed-data DIC measures select the model with the correct weights matrix over 200 samples.

E Trace Plots

In this section, we provide some trace plots to assess the convergence properties of the our samplers. To this end, we use the results for one of the sample out of 200 samples. Figure E.3 contains the trace plots for M1, Figure E.4 for M3, Figure E.5 for M2 and Figure E.6 for M4. These plots demonstrate the convergence of our samplers. In the context of empirical illustration, we also provide the trace plots to assess the convergence properties of our samplers. Since the observed-data DIC measures select $\mathbf{W}_{125}$, we only provide the trace plots for the static and the dynamic specifications based on $\mathbf{W}_{125}$. The results are shown in Figure E.7 and Figure E.8, respectively. These trace plots also indicate that the suggested sampler mixes properly and converges.

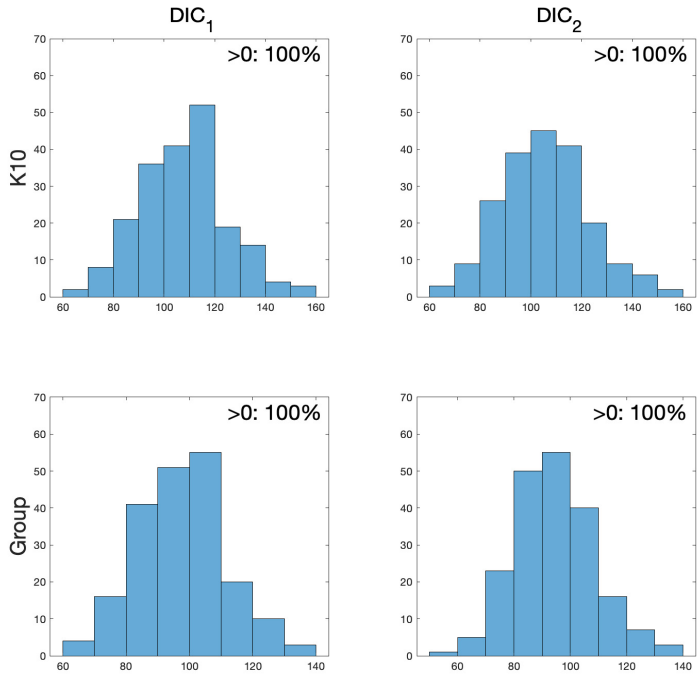

Figure E.1: Top panel: Histogram plots for the DIC of model based on K10 minus the DIC of model based on K5. Bottom panel: Histogram plots for the DIC of model based on Group minus the DIC of model based on K5

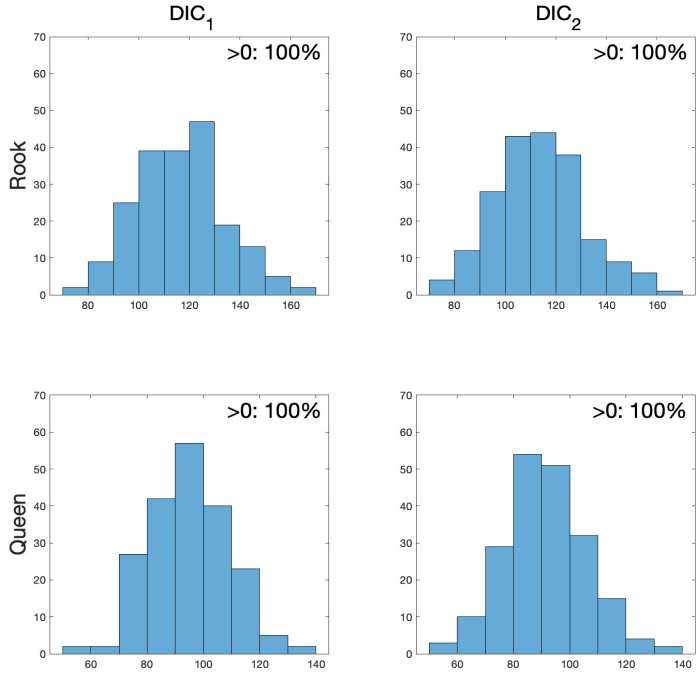

Figure E.2: Top panel: Histogram plots for the DIC of model based on Rook minus the DIC of model based on K5. Bottom panel: Histogram plots for the DIC of model based on Queen minus the DIC of model based on K5.

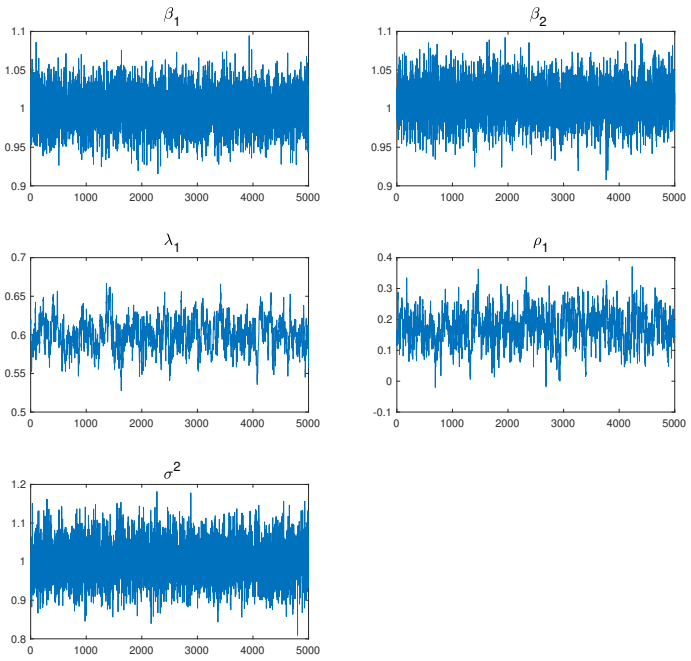

Figure E.3: Trace plots for M1

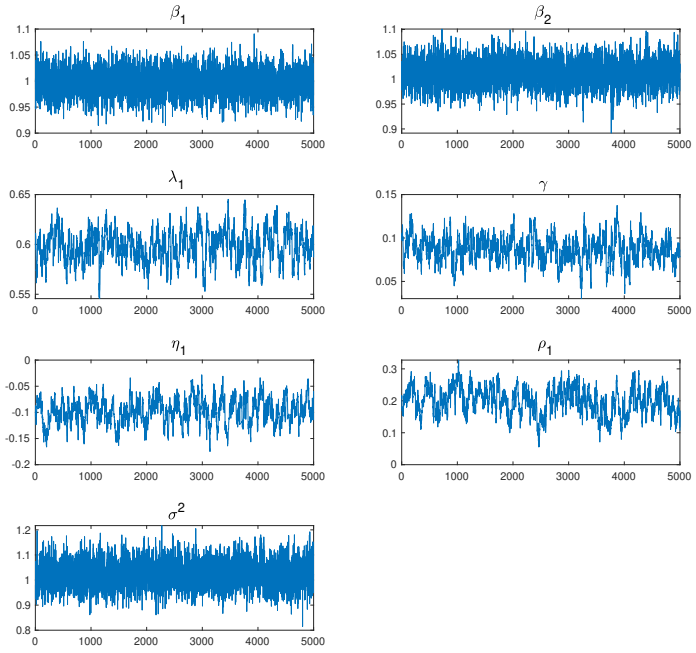

Figure E.4: Trace plots for M3

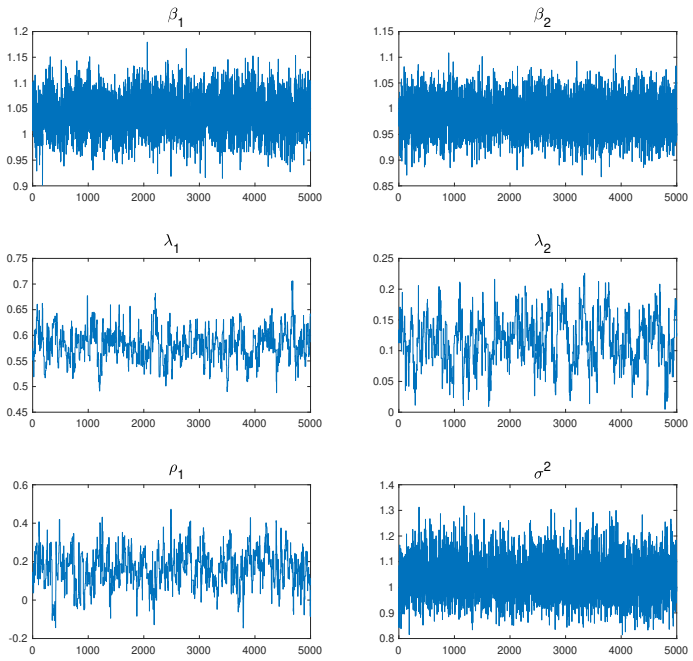

Figure E.5: Trace plots for M2

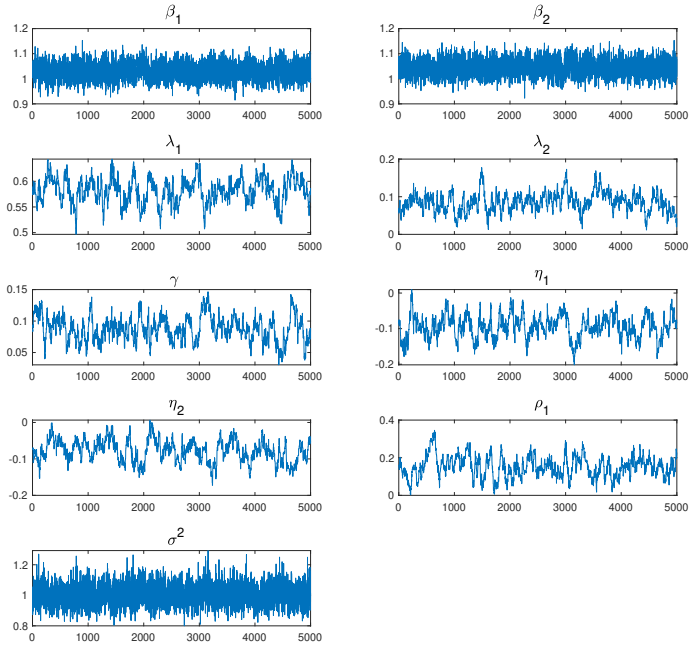

Figure E.6: Trace plots for M4

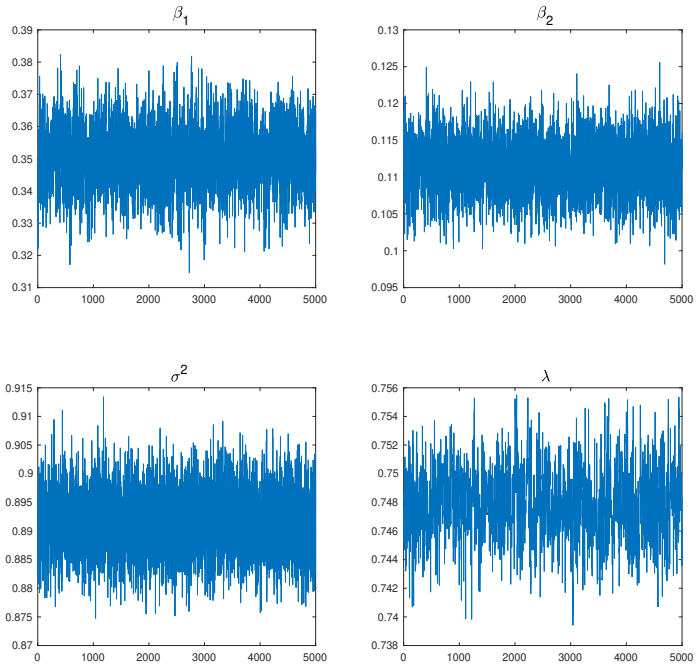

Figure E.7: Trace plots for the static model in the empirical illustration

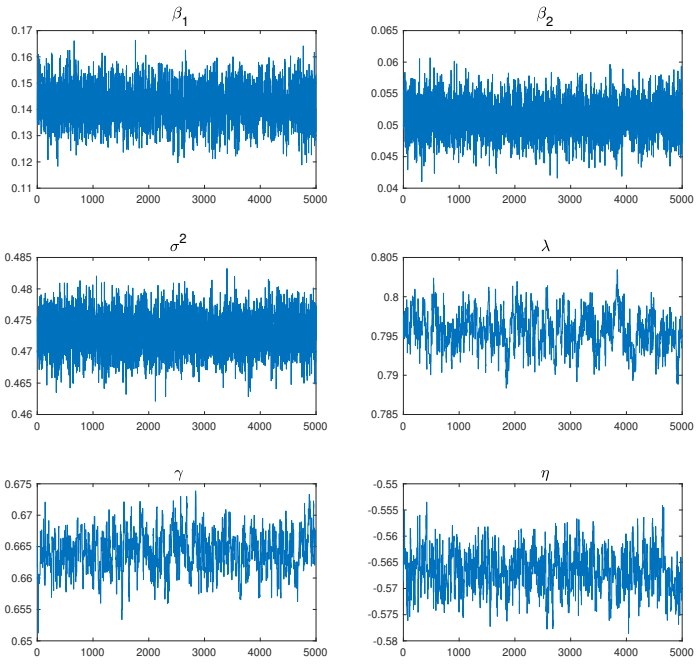

Figure E.8: Trace plots for the dynamic model in the empirical illustration

Compliance with ethical standards

Funding: This research was supported, in part, by a grant of computer time from the City University of New York High Performance Computing Center under NSF Grants CNS-0855217 and CNS-0958379.

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

Aquaro, Michele, Natalia Bailey, and M Hashem Pesaran (2021). "Estimation and inference for spatial models with heterogeneous coefficients: an application to US house prices". In: *Journal of Applied Econometrics* 36.1, pp. 18–44.

Baldoni, Edoardo and Roberto Esposti (2021). "Agricultural Productivity in Space: an Econometric Assessment Based on Farm-Level Data". In: *American Journal of Agricultural Economics* 103.4, pp. 1525–1544.

Baltagi, Badi H. and Zhenlin Yang (2012). "Standardized LM tests for spatial error dependence in linear or panel regressions". In: *The Econometrics Journal* 16.1, pp. 103–134.

— (2013). "Standardized LM tests for spatial error dependence in linear or panel regressions". In: *The Econometrics Journal* 16.1, pp. 103–134.

Baltagi, Badi H., Seuck Heun Song, and Won Koh (2003). "Testing panel data regression models with spatial error correlation". In: *Journal of Econometrics* 117.1, pp. 123 –150.

Baltagi, Badi H., Peter Egger, and Michael Pfaffermayr (2007a). "A Monte Carlo Study for Pure and Pretest Estimators of a Panel Data Model with Spatially Autocorrelated Disturbances". In: *Annales d'Economie et de Statistique* 87/88, pp. 11–38.

Baltagi, Badi H. et al. (2007b). "Testing for serial correlation, spatial autocorrelation and random effects using panel data". In: *Journal of Econometrics* 140.1, pp. 5 –51.

Baltagi, Badi H., Bernard Fingleton, and Alain Pirotte (2014). "Estimating and Forecasting with a Dynamic Spatial Panel Data Model". In: *Oxford Bulletin of Economics and Statistics* 76.1, pp. 112–138.

Baltagi, Badi H, Peter H Egger, and Michaela Kesina (2015). "Sources of productivity spillovers: panel data evidence from China". In: *Journal of Productivity Analysis* 43.3, pp. 389–402.

Bera, Anil K. et al. (2019). "Robust LM tests for spatial dynamic panel data models". In: *Regional Science and Urban Economics* 76, pp. 47–66.

Berg, Andreas, Renate Meyer, and Jun Yu (2004). "Deviance Information Criterion for Comparing Stochastic Volatility Models". In: *Journal of Business & Economic Statistics* 22.1, pp. 107–120.

Billé, Anna Gloria and Marco Rogna (2022). "The effect of weather conditions on fertilizer applications: A spatial dynamic panel data analysis". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 185.1, pp. 3–36.

Celeux, G. et al. (Dec. 2006). "Deviance information criteria for missing data models". In: *Bayesian Analysis* 1.4, pp. 651–673.

Chan, Joshua C.C. and Angelia L. Grant (2016a). "Fast computation of the deviance information criterion for latent variable models". In: *Computational Statistics & Data Analysis* 100, pp. 847–859.

— (2016b). "On the Observed-Data Deviance Information Criterion for Volatility Modeling". In: *Journal of Financial Econometrics* 14.4, pp. 772–802.

Chen, Weidong, Yufang Peng, and Guanyi Yu (2020). "The influencing factors and spillover effects of interprovincial agricultural carbon emissions in China". In: *PLOS ONE* 15.11, pp. 1–17.

Crespo, Jesús Cuaresma and Martin Feldkircher (2013). "Spatial filtering, model uncertainty and the speed of income convergence in Europe". In: *Journal of Applied Econometrics* 28.4, pp. 720–741.

Debarsy, Nicolas and James P LeSage (2020). "Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity matrices". In: *Journal of Business & Economic Statistics*, pp. 1–12.

Ehlert, Andree and Dirk Oberschachtsiek (2014). "Does managed care reduce health care expenditure? Evidence from spatial panel data". In: *International journal of health care finance and economics* 14.3, pp. 207–227.

Elhorst, J. Paul (2005). "Unconditional Maximum Likelihood Estimation of Linear and Log-Linear Dynamic Models for Spatial Panels". In: *Geographical Analysis* 37.1, pp. 85–106.

Fingleton, Bernard (2008). "A Generalized Method of Moments Estimator for a Spatial Panel Model with an Endogenous Spatial Lag and Spatial Moving Average Errors". In: *Spatial Economic Analysis* 3.1, pp. 27–44.

Gelman, A. et al. (2003). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Boca Raton, FL: Taylor & Francis.

Gori Maia, Alexandre et al. (2021). "The economic impacts of the diffusion of agroforestry in Brazil". In: *Land Use Policy* 108, p. 105489.

Guliyev, Hasraddin (2020). "Determining the spatial effects of COVID-19 using the spatial panel data model". In: *Spatial Statistics* 38, p. 100443.

Haario, Heikki, Eero Saksman, and Johanna Tamminen (2001). "An Adaptive Metropolis Algorithm". In: *Bernoulli* 7.2, pp. 223–242.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.

Han, Xiaoyi and Lung-Fei Lee (2016). "Bayesian Analysis of Spatial Panel Autoregressive Models With Time-Varying Endogenous Spatial Weight Matrices, Common Factors, and Random Coefficients". In: *Journal of Business & Economic Statistics* 34.4, pp. 642–660.

Han, Xiaoyi, Chih-Sheng Hsieh, and Lung fei Lee (2017). "Estimation and model selection of higher-order spatial autoregressive model: An efficient Bayesian approach". In: *Regional Science and Urban Economics* 63, pp. 97 –120.

Horn, Roger A. and Charles R. Johnson (2013). *Matrix Analysis*. Second edition. Cambridge: Cambridge University Press.

Hunneman, Auke, J. Paul Elhorst, and Tammo H. A. Bijmolt (2021). "Store sales evaluation and prediction using spatial panel data models of sales components". In: *Spatial Economic Analysis* 0.0, pp. 1–24. DOI: 10.1080/17421772.2021.1916574.

Jiang, Lei et al. (2021). "Identifying the driving factors of NO 2 pollution of One Belt One Road countries: satellite observation technique and dynamic spatial panel data analysis". In: *Environmental Science and Pollution Research* 28.16, pp. 20393–20407.

Kapoor, Mudit, Harry H. Kelejian, and Ingmar R. Prucha (2007). "Panel data models with spatially correlated error components". In: *Journal of Econometrics* 140.1, pp. 97 –130.

Kelejian, Harry H and Gianfranco Piras (2014). "Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes". In: *Regional Science and Urban Economics* 46, pp. 140–149.

Kelejian, Harry H. and Gianfranco Piras (2016). "A J test for dynamic panel model with fixed effects, and nonparametric spatial and time dependence". In: *Empirical Economics* 51.4, 1581–1605.

Kelejian, Harry H. and Ingmar R. Prucha (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances". In: *Journal of Econometrics* 157, pp. 53–67.

Krisztin, Tamás and Philipp Piribauer (2021). "A Bayesian approach for estimation of weight matrices in spatial autoregressive models". In: *arXiv preprint arXiv:2101.11938*.

Krisztin, Tamás, Philipp Piribauer, and Michael Wögerer (2020). "The spatial econometrics of the coronavirus pandemic". In: *Letters in Spatial and Resource Sciences* 13.3, pp. 209–218.

— (2021). "A spatial multinomial logit model for analysing urban expansion". In: *Spatial Economic Analysis*, pp. 1–22.

Kuschnig, Nikolas et al. (2021). "Spatial spillover effects from agriculture drive deforestation in Mato Grosso, Brazil". In: *Scientific reports* 11.1, pp. 1–9.

Lam, Clifford and Pedro CL Souza (2020). "Estimation and selection of spatial weight matrix in a spatial lag model". In: *Journal of Business & Economic Statistics* 38.3, pp. 693–710.

Lee, Jin and Young Hoon Lee (2021). "Development of a Win Production Function and Evaluation of Cross-Sectional Dependence". In: *Journal of Sports Economics* 22.4, pp. 412–431.

Lee, Lung-fei and Jihai Yu (2010a). "A Spatial Dynamic Panel Data Model with both Time and Individual Fixed Effects". In: *Econometric Theory* 26 (02), pp. 564–597.

— (2010b). "Estimation of spatial autoregressive panel data models with fixed effects". In: *Journal of Econometrics* 154.2, pp. 165 –185.

— (2014). "Efficient GMM estimation of spatial dynamic panel data models with fixed effects". In: *Journal of Econometrics* 180.2, pp. 174 –197.

LeSage, James P. (2014). "Spatial econometric panel data model specification: A Bayesian approach". In: *Spatial Statistics* 9. Revealing Intricacies in Spatial and Spatio-Temporal Data: Papers from the Spatial Statistics 2013 Conference, pp. 122 –145.

LeSage, James P and Yuxue Sheng (2014). "A spatial econometric panel data examination of endogenous versus exogenous interaction in Chinese province-level patenting". In: *Journal of Geographical Systems* 16.3, pp. 233–262.

Li, Fang et al. (2021). "How do non-farm employment and agricultural mechanization impact on large-scale farming? A spatial panel data analysis from Jiangsu Province, China". In: *Land Use Policy* 107, p. 105517.

Li, Liyao and Zhenlin Yang (2020). "Estimation of fixed effects spatial dynamic panel data models with small T and unknown heteroskedasticity". In: *Regional Science and Urban Economics* 81, p. 103520.

Li, Liyao and Zhenlin Yang (2021). "Spatial dynamic panel data models with correlated random effects". In: *Journal of Econometrics* 221.2, pp. 424–454.

Li, Yong, Jun Yu, and Tao Zeng (2017). "Deviance Information Criterion for Model Selection: Justification and Variation". Working Paper, Singapore Management University.

— (2020). "Deviance information criterion for latent variable models and misspecified models". In: *Journal of Econometrics* 216.2, pp. 450–493.

Millar, Russell B. (2009). "Comparison of Hierarchical Bayesian Models for Overdispersed Count Data using DIC and Bayes' Factors". In: *Biometrics* 65.3, pp. 962–969.

Mundlak, Yair (1978). "On the pooling of time series and cross section data". In: *Econometrica: journal of the Econometric Society*, pp. 69–85.

Parent, Olivier and James P. LeSage (2010). "A spatial dynamic panel model with random effects applied to commuting times". In: *Transportation Research Part B: Methodological* 44.5, pp. 633–645.

— (2011). "A space-time filter for panel data models containing random effects". In: *Computational Statistics & Data Analysis* 55.1, pp. 475 –490.

— (2012). "Spatial dynamic panel data models with random effects". In: *Regional Science and Urban Economics* 42.4, pp. 727–738.

Piribauer, Philipp and Jesús Cuaresma Crespo (2016). "Bayesian variable selection in spatial autoregressive models". In: *Spatial Economic Analysis* 11.4, pp. 457–479.

Qu, Xi and Lung-fei Lee (2015). "Estimating a spatial autoregressive model with an endogenous spatial weight matrix". In: *Journal of Econometrics* 184.2, pp. 209–232.

Qu, Xi, Lung fei Lee, and Jihai Yu (2017). "QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices". In: *Journal of Econometrics* 197.2, pp. 173 –201.

Roberts, Gareth O. and Jeffrey S. Rosenthal (2009). "Examples of Adaptive MCMC". In: *Journal of Computational and Graphical Statistics* 18.2, pp. 349–367.

Robinson, P.M. (2008). "Correlation testing in time series, spatial and cross-sectional data". In: *Journal of Econometrics* 147.1, pp. 5 –16.

Song, Yang, Dayu Liu, and Qiaoru Wang (2021). "Identifying characteristic changes in club convergence of China's urban pollution emission: A spatial-temporal feature analysis". In: *Energy Economics* 98, p. 105243.

Spiegelhalter, David J. et al. (2002). "Bayesian measures of model complexity and fit". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.4, pp. 583–639.

Taşpınar, Süleyman, Osman Doğan, and Anil K. Bera Bera (2017). "GMM Gradient Tests for Spatial Dynamic Panel Data Models". In: *Regional Science and Urban Economics* 65, pp. 65–88.

Yang, Cynthia Fan (2021a). "Common factors and spatial dependence: an application to US house prices". In: *Econometric Reviews* 40.1, pp. 14–50.

Yang, Zhenlin (2018). "Unified M-estimation of fixed-effects spatial dynamic models with short panels". In: *Journal of Econometrics* 205.2, pp. 423–447.

— (2021b). "Joint tests for dynamic and spatial effects in short panels with fixed effects and heteroskedasticity". In: *Empirical Economics* 60, pp. 51–92.

Yu, Jihai, Robert de Jong, and Lung fei Lee (2008). "Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large". In: *Journal of Econometrics* 146.1, pp. 118 –134.

Zhang, Xinyu and Jihai Yu (2018). "Spatial weights matrix selection and model averaging for spatial autoregressive models". In: *Journal of Econometrics* 203.1, pp. 1–18.