Integrated Modified Harmonic Mean Method for Spatial Panel Data Models

Osman Doğan* Ye Yang† Süleyman Taşpınar‡

December 29, 2024

Abstract

In this paper, we propose an integrated modified harmonic mean estimator (IHME) for nested and non-nested model selection problems in spatial panel data models with entity and time fixed effects. We formulate the IHME based on the integrated likelihood functions obtained by analytically integrating out the high-dimensional entity and time fixed effects from the complete likelihood functions. To investigate the finite sample properties of the IHME, we design a comprehensive simulation study that allows for both nested and non-nested model selection exercises in some popular spatial panel data models. Our simulation results show that the IHME has excellent finite sample performance and outperforms some competing estimators in terms of precision. We provide an empirical application on the US house price changes to show the usefulness of the proposed IHME in a model selection exercise.

JEL-Classification: C13, C21, C31.
Keywords: integrated modified harmonic mean estimator, spatial panel data, model selection, Bayes factor.

*Department of Economics, Istanbul Technical University, Istanbul, Türkiye, email: osmandogan@itu.edu.tr.
†School of Accounting, Capital University of Economics and Business, Beijing, China, email: yang.ye@cueb.edu.cn.
‡Department of Economics, Queens College, The City University of New York, New York, U.S., email: staspinar@qc.cuny.edu.

1 Introduction

Spatial panel data models can allow for (i) spatial dependence in the dependent variable, (ii) spatial dependence in the disturbance terms, (iii) temporal effects, (iv) spatio-temporal effects, and (v) heterogeneity across entities and time periods, i.e., the entity and time fixed effects. There is a growing literature on the specification and estimation of spatial panel data models, among others, see Anselin et al. (2008), Elhorst (2014), and Lee and Yu (2010b,c). The high-dimensional entity and time fixed effects in these models pose challenges for estimation and testing approaches.

In the case of static spatial panel data models, the likelihood estimation can be based on two approaches: (i) a direct estimation approach and (ii) a transformation approach (Lee and Yu, 2010a,b). In the direct approach, both entity and time fixed effects are estimated along with the other common parameters. In the transformation approach, the entity and time fixed effects are eliminated from the model by using suitable transformation techniques. The likelihood functions of the transformed models can be interpreted as partial likelihood functions and can yield consistent estimators for all parameters. In the case of dynamic spatial panel data models with both entity and time fixed effects, the (quasi) maximum likelihood estimator (MLE) can have asymptotic bias even when both the numbers of entities and time periods are large (Lee and Yu, 2010a). Besides the likelihood approach, other estimation approaches suggested in the literature are (i) the instrumental variable (IV) and the generalized method of moments (GMM) approaches (Kapoor et al., 2007; Kelejian and Prucha, 2010; Lee and Yu, 2014), (ii) the M-estimation approach (Li and Yang, 2021; Yang, 2022; Yang et al., 2024; Yang, 2018), and (iii) the Bayesian Markov chain Monte Carlo (MCMC) approach (Han and Lee, 2016; Han et al., 2017; LeSage, 2014; Parent and LeSage, 2011, 2012; Yang et al., 2023).

With the advancement of these estimation techniques, researchers developed various hypothesis-testing approaches for the model selection exercises. The classical test statistics such as the Wald statistic, the Lagrange Multiplier (LM) statistic, the GMM gradient statistic, and the $C(\alpha)$ statistic can be used for testing null hypotheses about parameter restrictions in the well-specified spatial models (Anselin, 1988; Anselin et al., 1996; Baltagi and Yang, 2012, 2013; Baltagi et al., 2003, 2014; Bera et al., 2019; Bresson et al., 2007; Jin and Lee, 2018; Taşpınar et al., 2017; Yang, 2021b). The non-nested testing approaches based on the J-statistic, the Cox-type statistic, and the likelihood ratio statistic, focus on model selection exercises for cross-sectional spatial models (Anselin, 1984, 1986; Burridge, 2012; Han and Lee, 2013; Jin and Lee, 2013; Kelejian and Piras, 2016; Liu and Lee, 2019). Similarly, various information criteria have also been considered for non-nested model selection exercises only for cross-sectional spatial models (Yang et al., 2022; Zhang and Yu, 2018). There are only a few studies focusing on non-nested model selection exercises, either based on a testing approach, an information criterion approach, or a Bayesian marginal likelihood approach for spatial panel data models (LeSage, 2014; Yang et al., 2023).

In this paper, we consider the modified harmonic mean method of Gelfand and Dey (1994) for nested and non-nested model selection problems in static and dynamic spatial panel data models that have high-order spatial dependence in the dependent variable and the disturbance terms. The modified harmonic mean method can be easily used to formulate an estimator of the marginal likelihood functions based on the conditional likelihood function obtained by conditioning on the high-dimensional fixed effects. However, this estimator is expected to perform poorly in model selection exercises because the high-dimensional fixed effects cannot be estimated precisely. For example, Chan and Grant (2015) show that the modified harmonic mean estimator based on the conditional likelihood function of an unobserved components model has a substantial finite sample bias and tends to select the wrong model. This problem due to the presence of some high-dimensional parameters in conditional or complete data likelihood functions is also not specific to the modified harmonic mean estimator. Frühwirth-Schnatter and Wagner (2008) show that the marginal likelihood method suggested by Chib (1995) can also be unreliable when it is formulated with the complete data likelihood functions. Similarly, Chan and Grant (2016) show that the deviance information criterion (DIC) based on the conditional likelihood functions of some popular stochastic volatility models also performs poorly in model selection exercises.

We propose an integrated modified harmonic mean estimator (IHME) for estimating the marginal likelihood functions of spatial panel data models. In a Bayesian estimation setting, we assign multivariate normal distribution priors to the entity and time fixed effects. As shown in Yang et al. (2023), these priors allow us to formulate integrated likelihood functions by analytically integrating out the fixed effects from the complete likelihood functions. Our suggested IHME is based on the integrated likelihood functions that are free from the high-dimensional entity and time fixed effects.

To investigate the finite sample properties of the IHME, we design a comprehensive simulation setting that allows for both nested and non-nested model selection exercises involving some popular spatial panel data models. The non-nested model selection problem entails selecting the model with the true spatial weight matrix, while the nested model selection problem involves choosing between static and dynamic specifications. Our simulation results indicate that the IHME has an excellent finite sample performance and outperforms some competing estimators (e.g., the AIC and BIC) in terms of precision. Finally, using a dataset on the US house price changes, we illustrate how to use the proposed IHME in choosing the model that provides relatively "larger" marginal likelihood.

The rest of this paper is organized as follows. In Section 2, we present the models under consideration and briefly discuss some conditions for their stability. In Section 3, we introduce two modified harmonic mean estimators for our models based on two likelihood functions. In Section 4, we consider two extensions of our base specifications and show how our analysis should be adjusted accordingly. In Section 5, we investigate the performance of the proposed modified harmonic mean estimators in an extensive simulation study. In Section 6, we illustrate the methodology with an empirical illustration on the US house price changes. In Section 7, we present our concluding remarks. Some technical results are relegated to an appendix.

2 Model Specifications

In our analysis, we consider both the static and the dynamic spatial panel data specifications. Let $Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$ denote the $n \times 1$ vector of observations on an outcome variable, and let X_t denote the $n \times k$ matrix of observations on k non-stochastic time-varying explanatory variables at time t. Then, a high order static spatial panel data model can be specified as

$$Y_t = \sum_{r_1=1}^{p_1} \lambda_{r_1} W_{r_1} Y_t + X_t \beta + C_n + \alpha_t l_n + U_t, \quad U_t = \sum_{r_2=1}^{p_2} \rho_{r_2} M_{r_2} U_t + V_t, \tag{2.1}$$

for $t = 1, \dots, T$, where β is the $k \times 1$ vector of coefficients for the explanatory variables in X_t, $C_n = (c_1, c_2, \dots, c_n)'$ denotes the $n \times 1$ vector of entity fixed effects, $\alpha_t l_n$ denotes time fixed effect at period t, l_n is the $n \times 1$ vector of ones, $U_t = (u_{1t}, u_{2t}, \dots u_{nt})'$ is the $n \times 1$ vector of regression disturbance term, and $V_t = (v_{1t}, v_{2t}, \dots, v_{nt})'$ is the $n \times 1$ vector of innovation terms. The high order spatial lag terms are $\sum_{r_1=1}^{p_1} W_{r_1} Y_t$ and $\sum_{r_2=1}^{p_2} M_{r_2} U_t$, where W_{r_1} and M_{r_2} are the $n \times n$ non-stochastic spatial weights matrices that have zero diagonal elements for $r_1 = 1, 2, \dots, p_1$ and $r_2 = 1, 2, \dots, p_2$. The scalar spatial parameters are denoted by λ_{r_1} and ρ_{r_2} for $r_1 = 1, 2, \dots, p_1$ and $r_2 = 1, 2, \dots, p_2$. We assume that v_{it}'s are independent and identically distributed normal random variables with mean zero and variance σ^2 across i and t.

The dynamic spatial panel data model additionally includes the temporal and spatio-temporal lags of the outcome variable:

$$\begin{aligned} Y_t &= \sum_{r_1=1}^{p_1} \lambda_{r_1} W_{r_1} Y_t + \gamma Y_{t-1} + \sum_{r_1=1}^{p_1} \eta_{r_1} W_{r_1} Y_{t-1} + X_t \beta + C_n + \alpha_t l_n + U_t, \\ U_t &= \sum_{r_2=1}^{p_2} \rho_{r_2} M_{r_2} U_t + V_t, \end{aligned} \tag{2.2}$$

for $t = 1, \dots, T$, where the scalar parameter γ measures the persistence in the outcome variable, i.e., the temporal effect, and the scalar parameters η_{r_1}'s of spatial time lag terms capture the dynamic diffusion effects, i.e., the spatiotemporal effects. Both in (2.1) and (2.2), we have $c_i + \alpha_t = (c_i + \kappa) + (\alpha_t - \kappa)$ for arbitrary κ. Thus, the entity and time fixed effects are not identified jointly, and a normalization-type constraint must be imposed to achieve identification. For this purpose, we may require that $C_n' l_n = 0$, or we may simply set $\alpha_1 = 0$.

Let $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{p_1})'$, $\rho = (\rho_1, \rho_2, \dots, \rho_{p_2})'$, $S(\lambda) = I_n - \sum_{r_1=1}^{p_1} \lambda_{r_1} W_{r_1}$ and $R(\rho) = I_n - \sum_{r_2=1}^{p_2} \rho_{r_2} M_{r_2}$, where I_n is the $n \times n$ identity matrix. Since the spatial autoregressive models represent equilibrium relationships, we require that $S(\lambda)$ and $R(\rho)$ are invertible for all λ and ρ values in their respective parameter space. These matrices will be invertible if the spectral radii of $\sum_{r_1=1}^{p_1} \lambda_{r_1} W_{r_1}$ and $\sum_{r_2=1}^{p_2} \rho_{r_2} M_{r_2}$ are inside the unit circle. Moreover, when all spatial weights matrices are row-normalized, these matrices are invertible under the following sufficient conditions:

$\sum_{r_1=1}^{p_1} |\lambda_{r_1}| < 1$ and $\sum_{r_2=1}^{p_2} |\rho_{r_2}| < 1$.[1] The stability conditions for the dynamic specification can be investigated from its reduced form. Let $A(\lambda, \gamma, \eta) = S^{-1}(\lambda)(\gamma I_n + \sum_{r_1=1}^{p_1} \eta_{r_1} W_{r_1})$, where $\eta = (\eta_1, \eta_2, \ldots, \eta_{p_1})'$. Then, we have

$$Y_t = A(\lambda, \gamma, \eta) Y_{t-1} + S^{-1}(\lambda) \left(X_t \beta + C_n + \alpha_t l_n + R^{-1}(\rho) V_t \right). \tag{2.3}$$

If all eigenvalues of $A(\lambda, \gamma, \eta)$ lie inside the unit circle, then Y_t in (2.3) is stable (Hamilton, 1994, Proposition 10.1). When all spatial weights matrices are row normalized, the spectral radius theorem yields the following sufficient conditions for stability: $\sum_{r_2=1}^{p_2} |\rho_{r_2}| < 1$ and $\sum_{r_1=1}^{p_1} |\lambda_{r_1}| + \sum_{r_1=1}^{p_1} |\eta_{r_1}| + |\gamma| < 1$.

To complete the discussion on model specifications, we assume the following independent prior distributions:

$$\begin{aligned}
&\lambda_{r_1} \sim U(-1, 1), \quad \eta_{r_1} \sim U(-1, 1), \quad r_1 = 1, 2, \ldots, p_1, \quad \rho_{r_2} \sim U(-1, 1), \quad r_2 = 1, 2, \ldots, p_2, \\
&\gamma \sim U(-1, 1), \quad \beta \sim N(\mu_\beta, V_\beta), \quad \sigma^2 \sim IG(a_0, b_0), \quad C_n \sim N(\mu_c, V_c), \\
&\alpha_t \sim N(\mu_\alpha, V_\alpha), \quad t = 1, \ldots, T,
\end{aligned}$$

where $U(-1, 1)$ denotes the uniform distribution over the interval $(-1, 1)$, $N(\mu_d, V_d)$ denotes the normal distribution with mean μ_d and variance V_d for $d \in \{\beta, c, \alpha\}$, and $IG(a_0, b_0)$ denotes the inverse gamma distribution with shape parameter a_0 and scale parameter b_0. The uniform prior distributions are subject to the stability conditions mentioned in the preceding paragraph.

Since we choose to work with the conjugate priors for the common parameters β, σ^2, C_n and α_t, we have closed-forms for the conditional posterior distributions of these parameters, and thus we can directly sample these parameters through a Gibbs sampler. However, in both models, the conditional posterior distributions of the spatial parameters (and the autoregressive parameter) do not take known forms. Therefore, we suggest using the adaptive Metropolis (AM) algorithm to generate draws for these parameters (Roberts and Rosenthal, 2009). We leave the details of the estimation algorithms to Section C in the Appendix.

3 Modified Harmonic Mean Method

Let θ_i be the parameter vector in model M_i, $p(Y|\theta_i, M_i)$ be the likelihood function of model M_i, $p(\theta_i|M_i)$ be the prior distribution of θ_i under M_i, and $p(\theta_i|Y, M_i)$ be the posterior distribution under M_i. Then, the Bayes factor in favor of M_i against model M_j is given by $\mathrm{BF}_{ij} = p(Y|M_i)/p(Y|M_j)$, where $p(Y|M_l) = \int p(y|\theta_l, M_l) p(\theta_l|M_l) \mathrm{d}\theta_l$ is the marginal likelihood function of model M_l for $l \in \{i, j\}$, i.e., the Bayesian evidence for M_l. The result $\mathrm{BF}_{ij} > 1$ will indicate that the observed data are more likely under M_i than M_j.

The marginal likelihood functions do not have closed-forms for our spatial models because the

[1] See Han and Lee (2016) and Yang et al. (2023) on the parameter space of spatial parameters in both static and dynamic panel data models.

spatial parameters cannot be integrated out analytically from the likelihood functions. Therefore, the computation of the marginal likelihood functions will require numerical evaluation of high-dimensional integrals, which can be computationally demanding. Instead, we consider the modified harmonic mean methodology of Gelfand and Dey (1994) to estimate the marginal likelihood functions. Hereafter, we drop the dependency on the model indicator M_l for simplicity.

Let h denote a probability density function whose support is contained in the support of the posterior distribution $p(\theta|Y)$. Then, the modified harmonic mean method follows from the following relationship:

$$\begin{aligned}\mathrm{E}\left(\frac{h(\theta)}{p(Y|\theta)p(\theta)}\big|Y\right) &= \int \frac{h(\theta)}{p(Y|\theta)p(\theta)}p(\theta|Y)\mathrm{d}\theta = \int \frac{h(\theta)}{p(Y|\theta)p(\theta)}\frac{p(Y|\theta)p(\theta)}{p(Y)}\mathrm{d}\theta \\ &= \frac{1}{p(Y)}\int h(\theta)\mathrm{d}\theta = \frac{1}{p(Y)},\end{aligned} \tag{3.1}$$

where the first equality follows from taking expectation with respect to $p(\theta|Y)$, the second equality from $p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}$, and the last equality from the fact that h is a proper density function over the support of the posterior distribution. The result in (3.1) indicates that the marginal likelihood function $p(Y)$ can be expressed as

$$p(Y) = \left(\mathrm{E}\left(\frac{h(\theta)}{p(Y|\theta)p(\theta)}\big|Y\right)\right)^{-1}. \tag{3.2}$$

Let $\{\theta^1, \theta^2, \ldots, \theta^R\}$ be a sequence of posterior draws generated from $p(\theta|Y)$. Using these posterior draws, we can formulate the following estimator of $p(Y)$:

$$\widehat{p}(Y) = \left(\frac{1}{R}\sum_{r=1}^{R}\frac{h(\theta^r)}{p(Y|\theta^r)p(\theta^r)}\right)^{-1}. \tag{3.3}$$

If the ratio $h(\theta)/\left(p(Y|\theta)p(\theta)\right)$ is bounded above over the support of posterior distribution, then this estimator is a simulation consistent estimator when R goes to infinity (Geweke, 1999, p. 46). To guarantee this boundedness condition, following Geweke (1999), we consider a truncated multivariate normal density for h. Define the truncation set Δ as

$$\Delta = \{\theta \in \mathbb{R}^p : (\theta - \widehat{\theta})'\widehat{\Sigma}^{-1}(\theta - \widehat{\theta}) < \chi^2_{\alpha,p}\}, \tag{3.4}$$

where $\widehat{\theta}$ is the posterior mean of θ, $\widehat{\Sigma}$ is the posterior covariance of θ, $\chi^2_{\alpha,p}$ is the $(1-\alpha)$ quantile of the χ^2_p distribution, and χ^2_p denotes the chi-squared distribution with p degrees of freedom.[2] Then, h takes the following form:

$$h(\theta) = (1-\alpha)^{-1}(2\pi)^{-p/2}\left|\widehat{\Sigma}\right|^{-1/2}\exp\left(-\frac{1}{2}(\theta - \widehat{\theta})'\widehat{\Sigma}^{-1}(\theta - \widehat{\theta})\right) \times \mathbf{1}_{\Delta}(\theta), \tag{3.5}$$

[2]To determine Δ, we set $\alpha = 0.05$ in our simulation study.

where $|\cdot|$ denotes the matrix determinant and $\mathbf{1}_\Delta(\theta)$ is the indicator function taking value 1 if $\theta \in \Delta$ and 0 otherwise.

The modified harmonic mean estimator in (3.3) is formulated with the likelihood function $p(Y|\theta)$. However, for our spatial panel data models, there are alternative likelihood functions that can be utilized in formulating the modified harmonic mean estimator. In our specifications, the latent high-dimensional variables are C_n and α, where $\alpha = (\alpha_1, \ldots, \alpha_T)'$. Let $\delta = (\beta', \sigma^2)'$, and $\zeta = (\lambda', \rho')'$ be the $(p_1 + p_2) \times 1$ vector of spatial parameters in the case of the static model in (2.1). Let $\zeta = (\lambda', \rho', \gamma, \eta')'$ be the $(2p_1 + p_2 + 1) \times 1$ vector of parameters in the case of the dynamic model in (2.2).

In formulating the modified harmonic mean estimator, we may first consider the conditional likelihood function $p(Y|\delta, \zeta, C_n, \alpha)$ (conditional on C_n and α). Let $Y = (Y_1', \ldots, Y_T')'$, $X = (X_1', \ldots, X_T')'$ and $Y^s = (I_T \otimes S(\lambda))\, Y - X\beta$. Then, the conditional likelihood function of the static model is given by

$$\begin{aligned} p(Y|\delta, \zeta, C_n, \alpha) &= (2\pi)^{-nT/2} (\sigma^2)^{-nT/2} \left|S(\lambda)\right|^T \left|R(\rho)\right|^T \\ &\quad \times \exp\left(-\frac{1}{2}(Y^s - l_T \otimes C_n - \alpha \otimes l_n)' \left(I_T \otimes \Omega(\omega)\right) (Y^s - l_T \otimes C_n - \alpha \otimes l_n)\right), \end{aligned} \tag{3.6}$$

where $\Omega(\omega) = R'(\rho) R(\rho)/\sigma^2$ and $\omega = (\rho', \sigma^2)'$.

To determine the conditional likelihood function of the dynamic model, we assume that Y_0 is exogenously given. Let $Y_{-1} = (Y_0', \ldots, Y_{T-1}')'$, $D(\gamma, \eta) = \gamma I_n + \sum_{r_1=1}^{p_1} \eta_{r_1} W_{r_1}$ and $Y^d = (I_T \otimes S(\lambda))\, Y - (I_T \otimes D(\gamma, \eta))\, Y_{-1} - X\beta$. Then, the conditional likelihood function of the dynamic model is given by

$$\begin{aligned} p(Y|\delta, \zeta, C_n, \alpha) &= (2\pi)^{-nT/2} (\sigma^2)^{-nT/2} \left|S(\lambda)\right|^T \left|R(\rho)\right|^T \\ &\quad \times \exp\left(-\frac{1}{2}\left(Y^d - l_T \otimes C_n - \alpha \otimes l_n\right)' \left(I_T \otimes \Omega(\omega)\right) \left(Y^d - l_T \otimes C_n - \alpha \otimes l_n\right)\right). \end{aligned} \tag{3.7}$$

Using these conditional likelihood functions, we can formulate the conditional modified harmonic mean estimator (CHME) as

$$\widehat{p}_C(Y) = \left(\frac{1}{R}\sum_{r=1}^{R} \frac{h(\delta^r, \zeta^r, C_n^r, \alpha^r)}{p(Y|\delta^r, \zeta^r, C_n^r, \alpha^r) p(\delta^r, \zeta^r, C_n^r, \alpha^r)}\right)^{-1}. \tag{3.8}$$

Alternatively, in formulating the modified harmonic mean estimator, we may consider the integrated likelihood function $p(Y|\delta, \zeta)$ by analytically integrating out the high-dimensional C_n and α from the complete likelihood function $p(Y, C_n, \alpha|\delta, \zeta) = p(Y|\delta, \zeta, C_n, \alpha) p(C_n) p(\alpha)$:

$$p(Y|\delta, \zeta) = \int p(Y, C_n, \alpha|\delta, \zeta)\, \mathrm{d}C_n \mathrm{d}\alpha = \int p(Y|\delta, \zeta, C_n, \alpha) p(C_n) p(\alpha)\, \mathrm{d}C_n \mathrm{d}\alpha, \tag{3.9}$$

where $p(C_n)$ and $p(\alpha)$ are the prior distributions of C_n and α, respectively. Since the prior distributions $p(C_n)$ and $p(\alpha)$ are multivariate normal distributions, it is analytically possible to derive $p(Y|\delta,\zeta)$ from (3.9) (Yang et al., 2023). To that end, let $B_n = (I_n,\, l_n)$, $B = I_T \otimes B_n$, $\mu_f = (\mu_c^{'}, \mu_\alpha)^{'}$, $\mu_F = l_T \otimes \mu_f$, and $V_F = I_T \otimes V_f$, where $V_f = \text{blkdiag}(V_c, V_\alpha)$ is the block diagonal matrix formed by V_c and V_α. Then, the integrated likelihood function of the static specification in (2.1) can be derived from (3.9) as

$$\begin{aligned} p(Y|\delta,\zeta) = {} & (2\pi)^{-nT/2}(\sigma^2)^{-nT/2}\,|S(\lambda)|^T\,|R(\rho)|^T\,|V_F|^{-1/2}\,|K_F|^{-1/2} \\ & \times \exp\left(-\frac{1}{2}\left(Y^{s'}\left(I_T\otimes\Omega(\omega)\right)Y^s + \mu_F^{'}V_F^{-1}\mu_F - k_F^{'}K_F^{-1}k_F\right)\right), \end{aligned} \tag{3.10}$$

where $K_F = B^{'}\left(I_T\otimes\Omega(\omega)\right)B + V_F^{-1}$ and $k_F = B^{'}\left(I_T\otimes\Omega(\omega)\right)Y^s + V_F^{-1}\mu_F$. Similarly, the integrated likelihood function of the dynamic specification in (2.2) can be determined as

$$\begin{aligned} p(Y|\delta,\zeta) = {} & (2\pi)^{-nT/2}(\sigma^2)^{-nT/2}\,|S(\lambda)|^T\,|R(\rho)|^T\,|V_F|^{-1/2}\,|K_F|^{-1/2} \\ & \times \exp\left(-\frac{1}{2}\left(Y^{d'}\left(I_T\otimes\Omega(\omega)\right)Y^d + \mu_F^{'}V_F^{-1}\mu_F - k_F^{'}K_F^{-1}k_F\right)\right), \end{aligned} \tag{3.11}$$

where $K_F = B^{'}(I_T\otimes\Omega(\omega))B + V_F^{-1}$ and $k_F = B^{'}(I_T\otimes\Omega(\omega))Y^d + V_F^{-1}\mu_F$.

Using the integrated likelihood functions in (3.10) and (3.11), we suggest the following integrated modified harmonic mean estimator (IHME):

$$\widehat{p}_I(Y) = \left(\frac{1}{R}\sum_{r=1}^{R}\frac{h(\delta^r,\zeta^r)}{p(Y|\delta^r,\zeta^r)p(\delta^r,\zeta^r)}\right)^{-1}. \tag{3.12}$$

Comparing the IHME with the CHME, we see that the computation of the CHME requires the posterior draws of the high-dimensional fixed effects C_n and α, while that of the IHME does not. Because these high dimensional parameters may not be estimated precisely, we expect that the CHME will be imprecise relative to the IHME (Chan and Grant, 2015, 2016; Frühwirth-Schnatter and Wagner, 2008).

Remark 1. *The marginal likelihood method suggested by Chib (1995) can also be used in our setting for nested and non-nested model selection exercises. This method is based on the following identity:*

$$\ln\widehat{p}(Y) = \ln p(Y|\widehat{\delta},\widehat{\zeta},\widehat{C}_n,\widehat{\alpha}) + \ln p(\widehat{\delta},\widehat{\zeta},\widehat{C}_n,\widehat{\alpha}) - \ln p(\widehat{\delta},\widehat{\zeta},\widehat{C}_n,\widehat{\alpha}|Y), \tag{3.13}$$

where $\widehat{a}$ can be the posterior mean of a for $a \in \{\delta,\zeta,C_n,\alpha\}$. If all conditional posterior distributions take known forms, then we can estimate the last term $\ln p(\widehat{\delta},\widehat{\zeta},\widehat{C}_n,\widehat{\alpha}|Y)$ by following Chib (1995). However, since the conditional posterior distribution of ζ takes an unknown form in our case, we need to resort to Chib and Jeliazkov (2001) instead of Chib (1995). As shown in Frühwirth-Schnatter and Wagner (2008), the marginal likelihood estimator based on (3.13) can be unreliable because the high-dimensional fixed effects C_n and α can not be estimated precisely. Since the integrated log-likelihood function $\ln p(Y|\widehat{\delta},\widehat{\zeta})$ is analytically available in our setting, we should use the following *identity to formulate an estimator:*

$$\ln \widehat{p}(Y) = \ln p(Y|\widehat{\delta}, \widehat{\zeta}) + \ln p(\widehat{\delta}, \widehat{\zeta}) - \ln p(\widehat{\delta}, \widehat{\zeta}|Y), \tag{3.14}$$

The first two terms can be computed with the posterior draws. The last term is unknown and can be expressed as:

$$\ln p(\widehat{\delta}, \widehat{\zeta}|Y) = \ln p(\widehat{\zeta}|Y) + \log(\widehat{\beta}|\widehat{\zeta}, Y) + \ln p(\widehat{\sigma}^2|\widehat{\zeta}, \widehat{\beta}, Y). \tag{3.15}$$

To estimate $\ln p(\widehat{\delta}, \widehat{\zeta}|Y)$*, we first need to adjust the estimation algorithms in Section C of Appendix by using the integrated likelihood functions. The last term* $\ln p(\widehat{\sigma}^2|\widehat{\zeta}, \widehat{\beta}, Y)$ *take a known form and can be estimated from the output of the adjusted algorithms. However, the first two terms,* $\ln p(\widehat{\zeta}|Y)$ *and* $\ln(\widehat{\beta}|\widehat{\zeta}, Y)$*, can be estimated by following Chib and Jeliazkov (2001) and will require additional appropriate reduced Gibbs sampler runs.*

In our simulation, we also consider BIC and AIC formulated from the integrated likelihood function $p(Y|\theta)$, where $\theta = (\delta', \zeta')'$. The BIC is derived from a large sample approximation to the log-marginal likelihood function of the candidate model. The marginal likelihood function of model M_k can be expressed as $p(Y|M_k) = \int_{\Theta_k} p(Y|\theta_k, M_k)p(\theta_k|M_k)d\theta_k$, where θ_k is the $P_k \times 1$ parameter vector in the model M_k. The Laplace approximation to $p(Y|M_k)$ gives (Schwarz, 1978):

$$\ln p(Y|M_k) = \ln p(Y|\widehat{\theta}_k, M_k) + \ln p(\widehat{\theta}_k|M_k) + \frac{P_k \pi}{2} - \frac{P_k \ln N}{2} - \frac{1}{2}|J_k(\widehat{\theta}_k)| + O_p(1/N), \tag{3.16}$$

where $\widehat{\theta}_k$ is the MLE of θ_k, $J_k(\widehat{\theta}_k) = -\frac{1}{n}\frac{\partial^2 \ln p(Y|\widehat{\theta}_k, M_k)}{\partial\theta_k \partial\theta_k'}$, and $N = nT$. Under a non-informative prior distribution and ignoring all $O_p(1)$ terms in (3.16), Schwarz (1978) define the BIC for M_k as

$$\text{BIC}_k = -2 \ln p(Y|\widehat{\theta}_k, M_k) + P_k \ln N. \tag{3.17}$$

The Laplace approximation in (3.16) can also be used to show that the difference between the BIC's of two models is asymptotically equivalent to the log Bayes factor (Kass and Raftery, 1995). That is, for any $\epsilon > 0$, we have

$$\lim_{N\to\infty} P\left(\left|\frac{\text{BIC}_k - \text{BIC}_l}{\ln \text{BF}_{kl}} - 1\right| > \epsilon\right) = 0, \tag{3.18}$$

where $\text{BF}_{kl} = p(Y|M_k)/p(Y|M_l)$ is the Bayes factor of M_k against M_l. Thus, the BIC is also a consistent model selection criterion like the Bayes factor, i.e., both BIC and BF select the true model with probability approaching one when $N \to \infty$. Since the IHME is a simulation consistent estimator of the marginal likelihood function, this result indicates that the IHME and BIC will perform similarly when the sample size is large.

On the other hand, using a decision-theoretic approach, we can show that the AIC chooses the model whose predictive distribution is close to the true DGP. Let $g(Y)$ be the DGP, and

$Y_{\text{rep}} = (y_{1,\text{rep}}, \ldots, y_{N,\text{rep}})'$ be the $N \times 1$ vector of replicate data generated from $g(Y)$ independently from the observed data Y. Consider the Kullback-Leibler (KL) divergence between $g(Y_{\text{rep}})$ and the generic predictive distribution $p(Y_{\text{rep}}|Y)$:

$$KL\left(g(Y_{\text{rep}}), p(Y_{\text{rep}}|Y)\right) = \mathrm{E}_{Y_{\text{rep}}}\left(\ln \frac{g(Y_{\text{rep}})}{p(Y_{\text{rep}}|Y)}\right) == c - \int \ln p(Y_{\text{rep}}|Y) g(Y_{\text{rep}}) dY_{\text{rep}}, \tag{3.19}$$

where the expectation $\mathrm{E}_{Y_{\text{rep}}}$ is with respect to $g(Y_{\text{rep}})$, and $c = \int \ln g(Y_{\text{rep}}) g(Y_{\text{rep}}) dY_{\text{rep}}$, which is constant across candidate models. In (3.19), if we replace the generic predictive distribution $p(Y_{\text{rep}}|Y)$ with the plug-in predictive distribution $p(Y_{\text{rep}}|\widehat{\theta})$, then it can be shown that (Burnham and Anderson, 2002; Li et al., 2020)

$$\begin{aligned}
\mathrm{E}_Y\left(2 \times KL\left(g(Y_{\text{rep}}), p(Y_{\text{rep}}|\widehat{\theta})\right)\right) &= 2c + \mathrm{E}_Y\left(\int -2\ln p(Y_{\text{rep}}|\widehat{\theta}) g(Y_{\text{rep}}) dY_{\text{rep}}\right) \\
&= 2c + \mathrm{E}_Y\left(-2\ln p(Y_{\text{rep}}|\widehat{\theta}) + 2P\right) + o(1) \\
&= 2c + \mathrm{E}_Y\left(\text{AIC}\right) + o(1),
\end{aligned} \tag{3.20}$$

where the expectation E_Y is with respect to $g(Y)$. Thus, the AIC measure is an asymptotically unbiased estimator of $\mathrm{E}_Y\left(2 \times KL\left(g(Y_{\text{rep}}), p(Y_{\text{rep}}|\widehat{\theta})\right)\right) - 2c$. This theoretical result indicates that a model with a smaller AIC value will perform better in terms of predictive performance.

4 Some Extensions

In this section, we explore two extensions of our model specifications and demonstrate how the IHME should be adjusted accordingly. In the first extension, we consider the following spatial Durbin versions:

$$Y_t = \sum_{r_1=1}^{p_1} \lambda_{r_1} W_{r_1} Y_t + X_t\beta + \sum_{r_1=1}^{p_1} W_{r_1} X_t \phi_{r_1} + C_n + \alpha_t l_n + U_t, \tag{4.1}$$

$$Y_t = \sum_{r_1=1}^{p_1} \lambda_{r_1} W_{r_1} Y_t + \gamma Y_{t-1} + \sum_{r_1=1}^{p_1} \eta_{r_1} W_{r_1} Y_{t-1} + X_t\beta + \sum_{r_1=1}^{p_1} W_{r_1} X_t \phi_{r_1} + C_n + \alpha_t l_n + U_t, \tag{4.2}$$

where $\sum_{r_1=1}^{p_1} W_{r_1} X_t$ denotes the spatial lag terms of X_t and $\{\phi_{r_1}\}_{r_1=1}^{p_1}$ is the sequence of associated parameter vectors. These specifications suggest that the outcome variable of a spatial entity may also depend on the characteristics (exogenous variables) of other spatial entities. This does not present any challenges for our estimation approach in Section 2 or the modified harmonic mean method in Section 3, as we can always define a new set of regressors by combining X_t and $\sum_{r_1=1}^{p_1} W_{r_1} X_t$, and reformulate the models in terms of this new set.

In the second extension, we consider the time-varying spatial weights matrices versions of our specifications:

$$Y_t = \sum_{r_1=1}^{p_1} \lambda_{r_1} W_{r_1 t} Y_t + X_t \beta + C_n + \alpha_t l_n + U_t, \tag{4.3}$$

$$Y_t = \sum_{r_1=1}^{p_1} \lambda_{r_1} W_{r_1 t} Y_t + \gamma Y_{t-1} + \sum_{r_1=1}^{p_1} \eta_{r_1} W_{r_1, t-1} Y_{t-1} + X_t \beta + C_n + \alpha_t l_n + U_t, \tag{4.4}$$

where $U_t = \sum_{r_2=1}^{p_2} \rho_{r_2} M_{r_2 t} U_t + V_t$. In these specifications, $\{W_{r_1 t}\}_{r_1=1}^{p_1}$ and $\{M_{r_2 t}\}_{r_2=1}^{p_2}$ are sequences of time-varying spatial weights matrices for $t = 1, 2, \ldots, T$. Let $S_t(\lambda) = I_n - \sum_{r_1=1}^{p_1} \lambda_{r_1} W_{r_1 t}$, $R_t(\rho) = I_n - \sum_{r_2=1}^{p_2} \rho_{r_2} M_{r_2 t}$, $S(\lambda) = \text{Diag}\left(S_1(\lambda), \ldots, S_T(\lambda)\right)$, and $R(\rho) = \text{Diag}\left(R_1(\rho), \ldots, R_T(\rho)\right)$. In terms of these new notations, the conditional likelihood function of the static model can be stated as

$$\begin{aligned} p(Y|\delta, \zeta, C_n, \alpha) &= (2\pi)^{-nT/2} (\sigma^2)^{-nT/2} \left|S(\lambda)\right| \left|R(\rho)\right| \\ &\times \exp\left(-\frac{1}{2} (Y^s - l_T \otimes C_n - \alpha \otimes l_n)^{'} \Omega(\omega) (Y^s - l_T \otimes C_n - \alpha \otimes l_n) \right), \end{aligned} \tag{4.5}$$

where $Y^s = S(\lambda)Y - X\beta$ and $\Omega(\omega) = R^{'}(\rho)R(\rho)/\sigma^2$. The corresponding integrated likelihood function can be determined as

$$\begin{aligned} p(Y|\delta, \zeta) &= (2\pi)^{-nT/2} (\sigma^2)^{-nT/2} \left|S(\lambda)\right| \left|R(\rho)\right| \left|V_F\right|^{-1/2} \left|K_F\right|^{-1/2} \\ &\times \exp\left(-\frac{1}{2} \left(Y^{s'} \Omega(\omega) Y^s + \mu_F^{'} V_F^{-1} \mu_F - k_F^{'} K_F^{-1} k_F \right) \right), \end{aligned} \tag{4.6}$$

where $K_F = B^{'}\Omega(\omega)B + V_F^{-1}$ and $k_F = B^{'}\Omega(\omega)Y^s + V_F^{-1}\mu_F$.

To determine the conditional likelihood function of the dynamic model, let $D_t(\gamma, \eta) = \gamma I_n + \sum_{r_1=1}^{p_1} \eta_{r_1} W_{r_1 t}$, $D(\gamma, \eta) = \text{Diag}\left(D_1(\gamma, \eta), \ldots, D_T(\gamma, \eta)\right)$ and $Y^d = S(\lambda)Y - D(\gamma, \eta)Y_{-1} - X\beta$. Then, the conditional likelihood function can be derived as

$$\begin{aligned} p(Y|\delta, \zeta, C_n, \alpha) &= (2\pi)^{-nT/2} (\sigma^2)^{-nT/2} \left|S(\lambda)\right| \left|R(\rho)\right| \\ &\times \exp\left(-\frac{1}{2} \left(Y^d - l_T \otimes C_n - \alpha \otimes l_n \right)^{'} \Omega(\omega) \left(Y^d - l_T \otimes C_n - \alpha \otimes l_n \right) \right). \end{aligned} \tag{4.7}$$

We can use this conditional function to determine the integrated likelihood function of the dynamic specification as

$$\begin{aligned} p(Y|\delta, \zeta) &= (2\pi)^{-nT/2} (\sigma^2)^{-nT/2} \left|S(\lambda)\right| \left|R(\rho)\right| \left|V_F\right|^{-1/2} \left|K_F\right|^{-1/2} \\ &\times \exp\left(-\frac{1}{2} \left(Y^{d'} \Omega(\omega) Y^d + \mu_F^{'} V_F^{-1} \mu_F - k_F^{'} K_F^{-1} k_F \right) \right), \end{aligned} \tag{4.8}$$

where $K_F = B^{'}\Omega(\omega)B + V_F^{-1}$ and $k_F = B^{'}\Omega(\omega)Y^d + V_F^{-1}\mu_F$. The conditional and integrated likelihood functions presented in this section then can be used to easily compute the CHME and the IHME derived in Section 3.

5 Simulations

In this section, we examine the performance of the IHME and the CHME in a simulation study. To this end, we are interested in the performance of these estimators in two scenarios: (i) a spatial weights matrix selection problem, and (ii) a model selection problem between the static specification versus the dynamic specification. The first scenario provides examples of the non-nested model selection while the second provides examples of the nested model selection. We consider the following specifications for the data generating processes (DGPs):

$$\begin{aligned}
\text{M1}: \quad & Y_t = \lambda_1 W_1 Y_t + X_t\beta + C_n + \alpha_t l_n + U_t, \quad U_t = \rho_1 M_1 U_t + V_t, \\
\text{M2}: \quad & Y_t = \lambda_1 W_1 Y_t + \lambda_2 W_2 Y_t + X_t\beta + C_n + \alpha_t l_n + U_t, \quad U_t = \rho_1 M_1 U_t + V_t, \\
\text{M3}: \quad & Y_t = \lambda_1 W_1 Y_t + \gamma Y_{t-1} + \eta_1 W_1 Y_{t-1} + X_t\beta + C_n + \alpha_t l_n + U_t, \quad U_t = \rho_1 M_1 U_t + V_t, \\
\text{M4}: \quad & Y_t = \lambda_1 W_1 Y_t + \lambda_2 W_2 Y_t + \gamma Y_{t-1} + \eta_1 W_1 Y_{t-1} + \eta_2 W_2 Y_{t-1} + X_t\beta + C_n + \alpha_t l_n + U_t, \\
& U_t = \rho_1 M_1 U_t + V_t,
\end{aligned}$$

for $t = 1, 2, \ldots, T$. We consider two explanatory variables $X_t = (X_{1t}, X_{2t})$, whose elements X_{1it}'s and X_{2it}'s are drawn independently from $N(0, 2)$. For the parameters, we consider the following true values: $\lambda_1 = 0.3$, $\lambda_2 = 0.1$, $\gamma = 0.3$, $\eta_1 = -0.1$, $\eta_2 = -0.1$, $\rho_1 = 0.3$, $\beta_1 = \beta_2 = 1$, and $\sigma^2 = 1$. The time fixed effects are drawn independently from the standard normal distribution. The entity fixed effects are generated as follows: $C_n = 0.3\overline{X}_1 + 0.3\overline{X}_2 + \boldsymbol{\epsilon}$, where $\overline{X}_1$ and $\overline{X}_2$ are the time-averages of X_{1t} and X_{2t} 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, and set the sample size to $(n, T) = (40, 20)$. We set the length of the MCMC chain to 5000 draws and discard the first 2000 draws as burn-ins.[3]

Besides the IHME and the CHME, we also consider the performance of the AIC and BIC in our simulation. The AIC is defined by $\text{AIC} = -2\ln p(Y|\tilde{\theta}) + 2P$, and the BIC by $\text{BIC} = -2\ln p(Y|\tilde{\theta}) + P\ln(NT)$, where $\tilde{\theta}$ is the maximum likelihood estimator (MLE) of θ. In computing these measures, we approximate $p(Y|\tilde{\theta})$ by taking the maximum of likelihood function over the posterior draws, i.e., $p(Y|\tilde{\theta}) \approx \max\{p(Y|\theta^1), p(Y|\theta^2), \ldots, p(Y|\theta^R)\}$. These measures are also formulated with the integrated likelihood functions given in (3.10) and (3.11).

For the weights matrix selection problem (the non-nested model selection problem), we consider the following experiment. For a given DGP (out of M1 through M4), we generate 200 samples using the queen contiguity-based weights matrices. Thus, in this experiment, the correct model is the one with the queen contiguity-based weights matrices while the misspecified model is the one with the rook contiguity-based weights matrices. We then use each sample data to estimate the correct

[3]See Appendix D for the details on the spatial weights matrices and Appendix B on the elicitation of the hyper-parameters.

and the misspecified models. Finally, we use the estimation output of each model to estimate the selection criteria, namely, the IHME, the CHME, the AIC and the BIC.

For the static versus dynamic specification problem (the nested model selection problem), we consider four experiments. The goal of each experiment is to see whether the selection criteria can distinguish the true DGP from the misspecified model. In the first experiment, we investigate the performance of the selection criteria in distinguishing M1 from M3. We generate 200 samples according to M1 by using the rook contiguity-based weights matrices. We then use each sample to estimate M1 and M3 and all selection measures. In the second experiment, we repeat the first experiment for M2 and M4, where the true DGP is M2 with the spatial weights matrices generated according to the queen contiguity case. In the third experiment, the true model is M3 with the spatial weights matrices generated according to the queen contiguity case, and the misspecified model is M1. Finally, in the fourth experiment, the true model is M4 with the spatial weights matrices generated according to the rook contiguity case, and the misspecified model is M2.

We resort to histogram plots to concisely summarize the simulation results (Chan and Grant, 2016). For the IHME and CHME, we subtract the logarithm of the estimated marginal likelihood function under the misspecified model from the logarithm of the estimated marginal likelihood function under the true model. For the AIC and BIC, we subtract the estimates of these measures under the true model from the estimates obtained under the misspecified model. The differences for 200 samples are then presented in histograms for each experiment. If a selection criterion performs well, then the majority of its differences should be positive over 200 samples. We also provide two summary statistics in the legend of histograms: (i) "Pos" shows the percentage of positive differences, and (ii) "Sd" shows the standard deviation of differences.

Figures 1 through 4 present the results of the weights matrix selection problem. In general, the IHME presents an excellent performance in choosing the correct model. For example, in Figure 1, the percentages of the positive differences (Pos) are 100% for all measures. However, the standard deviation of differences (Sd) for the IHME is 0.962, which is much smaller than that of the CHME (10.09), AIC (1.946), and BIC (1.946), indicating that the IHME is the most precise among the four competing selection measures. Similar remarks apply to the results for M2 and M4 in Figures 2 and 4. For the case M3 in Figure 3, the IHME, AIC, and BIC still perform satisfactorily. However, the percentage of positive differences for the CHME is zero, indicating that the CHME selects the true weights matrix in none of the 200 samples.

Figures 5 through 8 present the results of the nested model selection problem. In Figure 5, all four measures select the true model and the IHME has the smallest standard deviation. In Figure 6, similar to the first experiment, all measures have positive differences and the standard deviation for the IHME is the smallest. In Figure 7, the CHME selects the true model in none of the samples, but the IHME, AIC and BIC select the true model in all samples. The IHME is again the most precise estimator. Finally, in Figure 8, all four selection measures perform well in selecting the true model. The IHME again reports a smaller standard deviation (10.083) than that of the CHME (15.987).

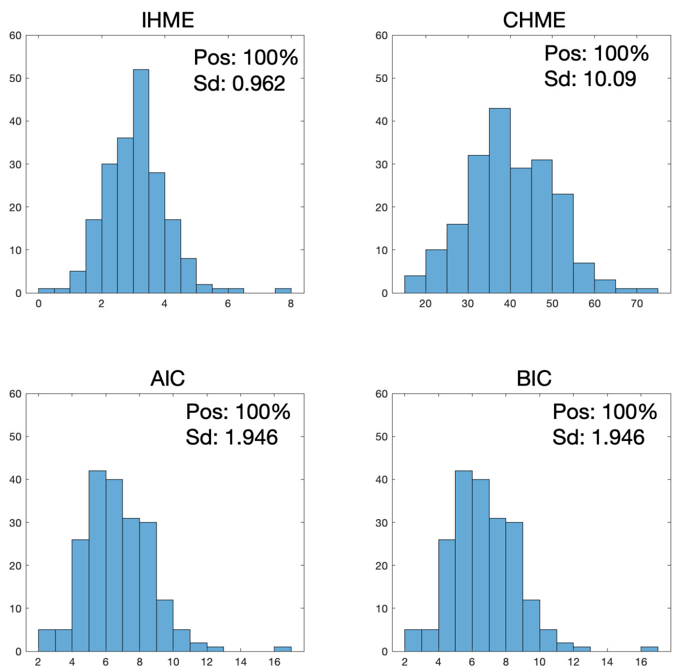

Figure 1: Selecting the true weight matrix in M1.

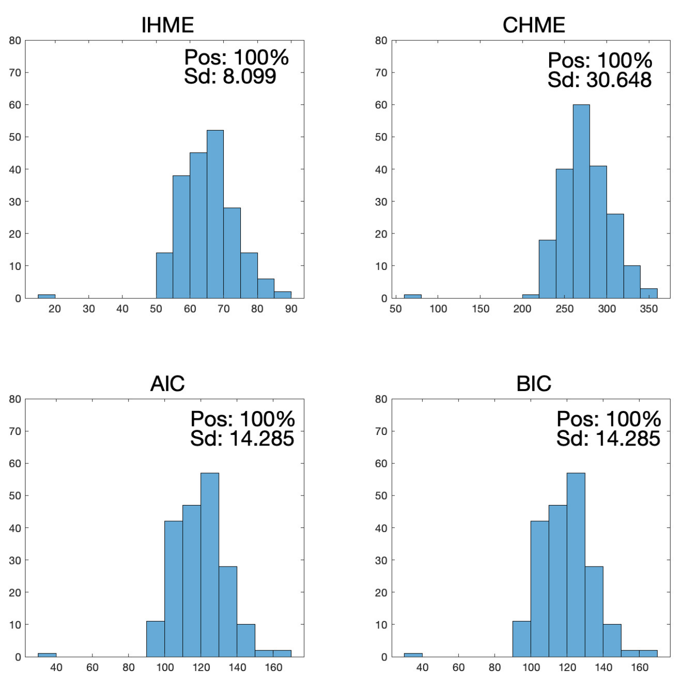

Figure 2: Selecting the true weight matrix in M2.

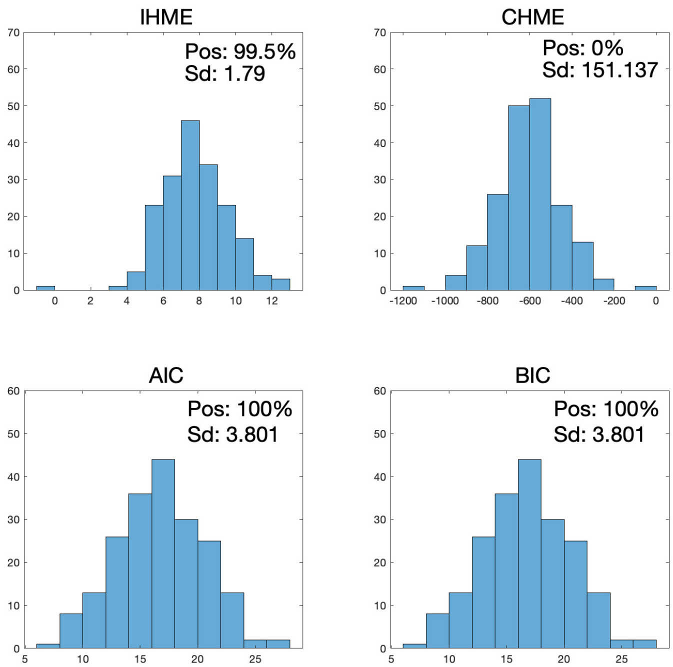

Figure 3: Selecting the true weight matrix in M3.

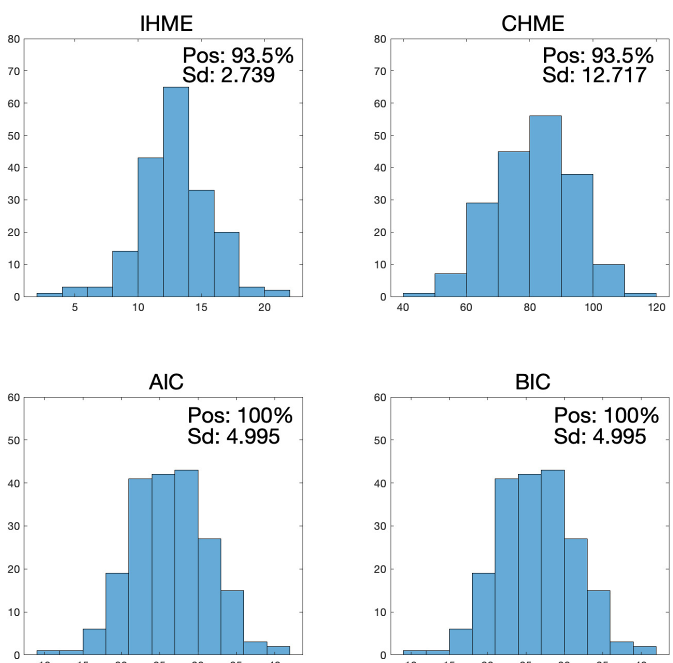

Figure 4: Selecting the true weight matrix in M4.

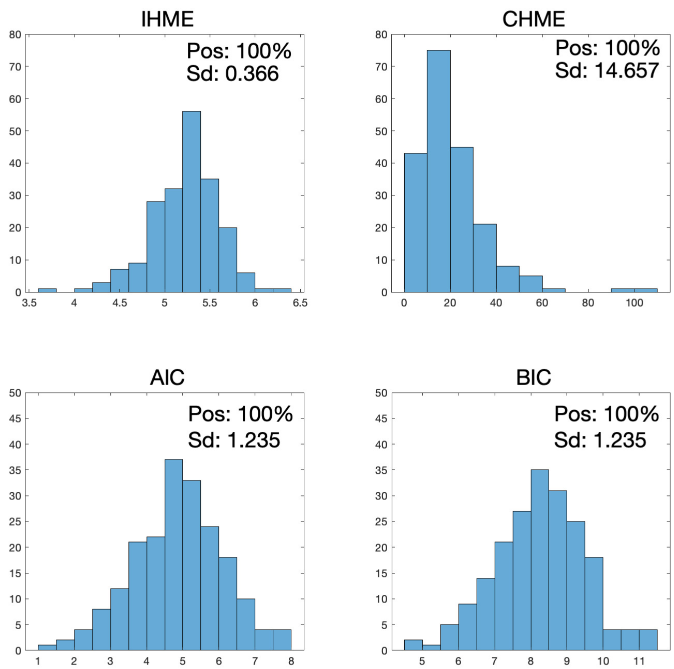

Figure 5: Comparing M1 with M3 when the true model is M1.

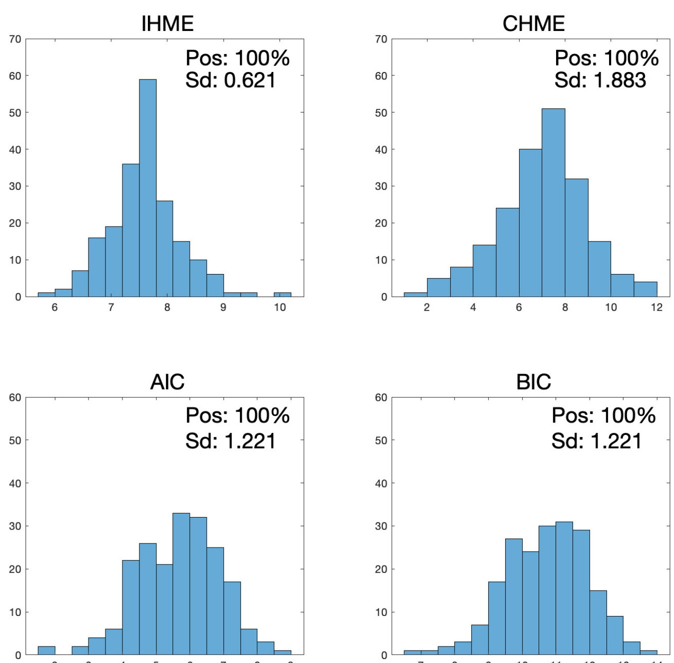

Figure 6: Comparing M2 with M4 when the true model is M2.

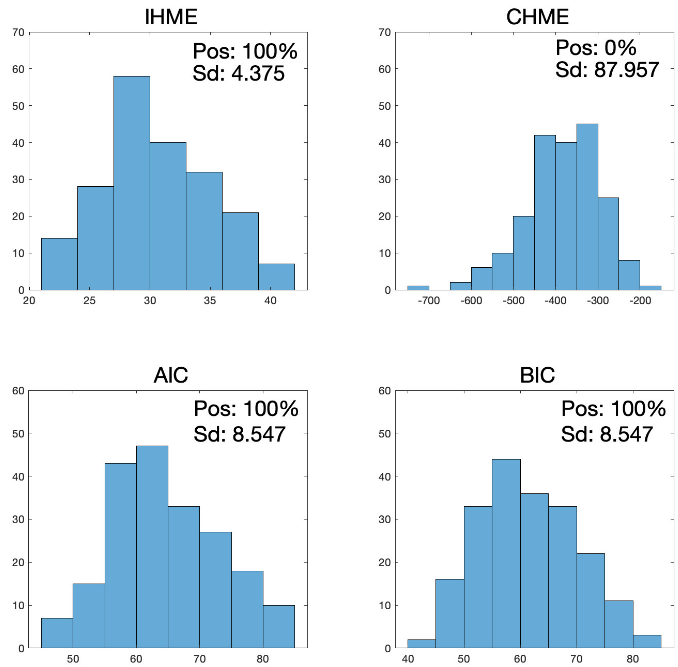

Figure 7: Comparing M3 with M1 when the true model is M3.

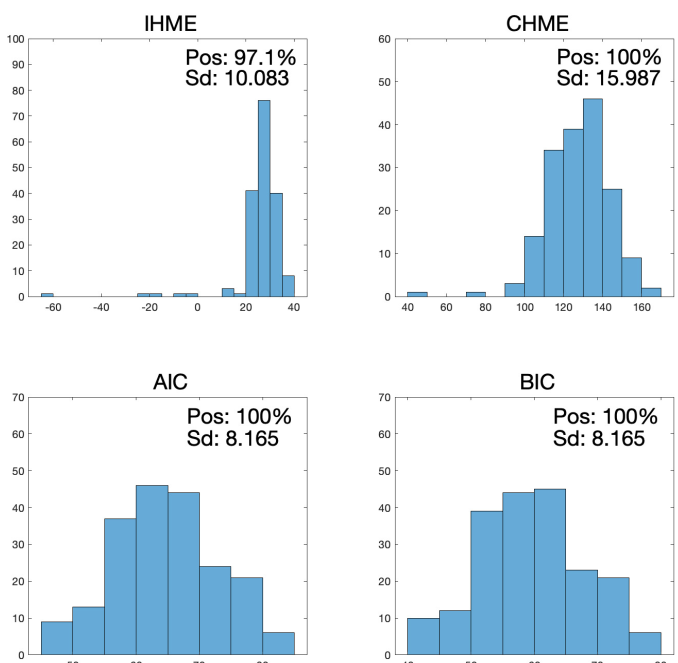

Figure 8: Comparing M4 with M2 when the true model is M4.

6 An Empirical Illustration

In this section, we provide an empirical illustration using a dataset on house price changes in the US to flesh out the usefulness of the proposed IHME. The dataset includes observations on average house prices, population, and income, for 377 Metropolitan Statistical Areas (MSAs) over the period 1975Q1-2014Q4. This dataset was introduced by Bailey et al. (2016) and later extended by Yang (2021a), Aquaro et al. (2021) and Bao and Zhou (2023).

Bailey et al. (2016) introduce a two-step approach for the analysis of spatiotemporal data, where data are defactored in the first step to account for strong cross-sectional dependence and the defactored data are modeled as a spatial dynamic model in the second step. Their specification allows for cross-sectional heterogeneity in spatial, temporal and spatio-temporal spillover dynamics. Their findings indicate evidence for significant positive and negative spillover effects in house price changes in the US, with larger positive spillover effects in magnitude than the negative spillover effects. Aquaro et al. (2021) explores cross-sectional heterogeneity in spatial spillover dynamics in the house price changes using a heterogeneous SAR model for the house price changes in the US. They find evidence for significant heterogeneity in the spatiotemporal dynamics of the house price changes. Yang (2021a) introduces a one-step estimation approach for a static panel data model that allows for strong cross-sectional dependence through common factors as well as weak cross-sectional dependence through spatial lag terms. Her findings indicate that regional common shocks contribute to the strong spatial dependence in house price changes in the US and that the strength of weak spatial dependence might be overestimated if strong dependence is not accounted for.

Bao and Zhou (2023) consider a spatial dynamic panel data model with cross-sectional heterogeneity in spatial and temporal spillover dynamics. For the spatial weights matrix specification, they consider two cases, a single convexly combined weights matrix from row-normalized weights matrices, and a high-order specification that treats weights matrices as different channels for spatial spillovers. They introduce Bayesian estimation methods for these models, and show that the specifications with both geographical and non-geographical spatial weights matrices are preferred over the specifications with geographical spatial weights matrices only. Yang et al. (2023) propose the observed-data deviance information criterion (DIC) based on the integrated likelihood functions for the static and dynamic spatial panel data models. They find that the dynamic specification for the house price changes is preferred over the static specification, and the spatial weights matrix with a radius threshold of 125 miles is preferred over the ones with 75 and 100 miles.

Following this strand of literature on the US house price changes, we use our proposed IHME in model selection problem between the static and the dynamic spatial panel data specifications, denoted by E-M1 and E-M3, respectively. We also consider the Durbin counterparts of these specifications, denoted by E-M1D and E-M3D, respectively. Thus, our model selection exercise is based on the following four specifications:

$$\begin{aligned}
&\text{E-M1}: \quad Y_t = \lambda W Y_t + X_t\beta + C_n + \alpha_t l_n + U_t,\\
&\text{E-M1D}: \quad Y_t = \lambda W Y_t + X_t\beta_1 + W X_t\beta_2 + C_n + \alpha_t l_n + U_t,\\
&\text{E-M3}: \quad Y_t = \lambda W Y_t + \gamma Y_{t-1} + \eta W Y_{t-1} + X_t\beta + C_n + \alpha_t l_n + U_t,\\
&\text{E-M3D}: \quad Y_t = \lambda W Y_t + \gamma Y_{t-1} + \eta W Y_{t-1} + X_t\beta_1 + W X_t\beta_2 + C_n + \alpha_t l_n + U_t,
\end{aligned}$$

where Y_t is the percent quarterly rate of change of real house price at period t, and X_t contains the percent quarterly rate of change of population (*gpop*) and the percent quarterly rate of change in real capita income (*ginc*) at period t. All specifications include entity and time fixed effects. To keep the discussion concise, we consider a spatial weights matrix that treats two MSAs as neighbors if these MSAs are located within a threshold distance of 125 miles (Yang et al., 2023).[4]

We use Algorithms 1 and 2 in Appendix C to estimate these models. In the estimation, the number of draws is set to 10,000, with the initial 5,000 draws discarded as burn-ins. The trace plots shown in Appendix E indicate that our suggested Gibbs sampler does not have any convergence issues. In Table 1, we provide estimated posterior means and standard deviations. We make the following two important observations. First, from the values of the IHME, the CHME, the AIC and the BIC in the last panel, the dynamic models E-M3 and E-M3D are preferred over the static models E-M1 and E-M1D, respectively. This is consistent with the observation that the estimate of γ capturing the dynamic effect is 0.664, and the estimate of η capturing the spatio-temporal effect is -0.567. They are large in magnitude and statistically significant. As for the selection between E-M3 and E-M3D, comparing the values of the four criteria, we observe that the IHME prefers the Durbin version E-M3D, while the other three criteria prefer E-M3. Note that the coefficient estimate for the Durbin term $W \times$ *gpop* is statistically significant at 5% level. Thus, we conclude that there is statistical evidence in favor of the dynamic Durbin specification (E-M3D) over other specifications.

Second, note that the estimate of λ is 0.748 in E-M1 and 0.796 in E-M3D, and they are statistically significant. Hence, there is evidence of strong contemporaneous spatial dependence in house price changes. The coefficient estimates of the exogenous variables *gpop* and *ginc* are 0.350 and 0.111 in E-M1, and 0.349 and 0.117 in E-M1D, respectively. The corresponding estimates are nearly the same for these two models. This is consistent with the observation that the coefficient estimates of the Durbin terms, $W \times$ *gpop* and $W \times$ *ginc*, although significant, are relatively small, 0.043 and -0.021, respectively. However, when the temporal and spatio-temporal lags are incorporated into the model, the magnitude of the coefficient estimates for the exogenous variables *gpop* and *ginc* decline significantly. For example, in the case of E-M3, they become 0.142 and 0.051, respectively. This is consistent with the fact that the estimates for γ capturing the dynamic effect and η capturing the spatio-temporal effect are both large in magnitude and statistically significant. Overall, our coefficient estimates are consistent with those provided in Yang (2021a) and Yang et al. (2023).

[4]More specifically, the distance is the geodesic distance using the Haversine formula.

Table 1: Estimation results for the house price changes in the US

	E-M1	E-M1D	E-M3	E-M3D
gpop	0.350***	0.349***	0.142***	0.140***
	(.009)	(.009)	(.007)	(.007)
ginc	0.111***	0.117***	0.051***	0.053***
	(.004)	(.004)	(.003)	(.003)
$W \times gpop$		0.043**		0.030**
		(.018)		(.013)
$W \times ginc$		-0.021***		-0.006
		(.006)		(.005)
λ	0.748***	0.748***	0.796***	0.796***
	(.002)	(.003)	(.002)	(.002)
γ			0.664***	0.664***
			(.003)	(.003)
η			-0.567***	-0.567***
			(.004)	(.004)
σ^2	0.891***	0.891***	0.473***	0.473***
	(.005)	(.005)	(.003)	(.003)
IHME	-133399.87	-133421.21	-131113.01	-131104.50
CHME	-94105.47	-94685.20	-74656.74	-76149.53
AIC	268366.93	268365.66	263685.73	263697.66
BIC	270490.35	270496.94	265813.08	265832.87

Significance levels: *: 10%, **: 5%, and ***: 1%.

7 Conclusion

In this paper, we proposed the IHME for the nested and non-nested model selection exercises for spatial panel data models that have high-order spatial dependence in the dependent variable and the disturbance terms. This estimator is based on the integrated likelihood functions obtained by analytically integrating out the high-dimensional fixed effects from the complete likelihood functions. We showed that the CHME formulated with the conditional likelihood function can perform poorly because the high-dimensional fixed effects cannot be estimated precisely. The results from our simulation exercises showed that the IHME had an excellent performance and outperformed the competing estimators in terms of precision. In an empirical illustration on the US house price changes, we showed how the IHME can be utilized in model selection exercises for choosing the model that yields relatively larger marginal likelihood. In future studies, our approach can be considered for static and dynamic spatial panel data models that have interactive fixed effects.

Appendix

A Abbreviations

In this section, we provide frequently used abbreviations in the main text.

MLE	Maximum Likelihood Estimator
LM	Lagrange Multiplier
IV	Instrumental Variable
GMM	Generalized Method of Moments
AM	Adaptive Metropolis
IHME	Integrated Modified Harmonic Mean Estimator
CHME	Conditional Modified Harmonic Mean Estimator
DGP	Data-Generating Process
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
DIC	Deviance Information Criterion
BF_{ij}	Bayes Factor of M_i against M_j

B Prior Distributions

In this section, we specify the prior distributions (Han and Lee, 2016; Han et al., 2017; Yang et al., 2023). Recall that:

$$\begin{aligned}
&\lambda_{r_1} \sim U(-1,1), \quad \eta_{r_1} \sim U(-1,1), \quad r_1 = 1,2,\ldots,p_1, \quad \rho_{r_2} \sim U(-1,1), \quad r_2 = 1,2,\ldots,p_2,\\
&\gamma \sim U(-1,1), \quad \beta \sim N(\mu_\beta, V_\beta), \quad \sigma^2 \sim IG(a_0, b_0),\\
&C_n \sim N(\mu_c, V_c), \quad \alpha_t \sim N(\mu_\alpha, V_\alpha), \quad t = 1,\ldots,T.
\end{aligned}$$

In our simulation study, we choose the following values for the hyper-parameters: $\mu_\beta = 0_{k\times 1}$, $V_\beta = 10I_k$, $a_0 = 0.01$ and $b_0 = 0.01$, $\mu_c = 0_{n\times 1}$, $V_c = 10I_n$, $\mu_\alpha = 0$ and $V_\alpha = 10$.

C Estimation Algorithms

In this section, we provide estimation algorithms for our main specifications. In the case of the static model, let $\delta = (\beta', \sigma^2)'$, and $\zeta = (\lambda', \rho')'$ be the $(p_1 + p_2) \times 1$ vector of spatial parameters. In the case of the dynamic model, let $\zeta = (\lambda', \rho', \gamma, \eta')'$ be the $(2p_1 + p_2 + 1) \times 1$ vector of parameters.

Algorithm 1 (Estimation Algorithm for the static model in (2.1))**.**

1. Sampling step for C_n: Let $Y_t^c = S(\lambda)Y_t - X_t\beta - \alpha_t l_n$ for $t = 1,2,\ldots,T$. Then,

$$C_n | Y, \delta, \zeta, \alpha \sim N(\widehat{\mu}_c, \widehat{V}_c), \tag{C.1}$$

where $\widehat{V}_c = \left(V_c^{-1} + T\Omega(\omega)\right)^{-1}$, $\widehat{\mu}_c = \widehat{V}_c\left(V_c^{-1}\mu_c + \Omega(\omega)\sum_{t=1}^T Y_t^c\right)$, and $\Omega(\omega) = R^{'}(\rho)R(\rho)/\sigma^2$.

2. Sampling step for α: Let $Y_t^\alpha = S(\lambda)Y_t - X_t\beta - C_n$ for $t = 1, 2, \ldots, T$. Then,

$$\alpha_t|Y, \delta, \zeta, C_n, \alpha_{-t} \sim N(\widehat{\mu}_{\alpha_t}, \widehat{V}_{\alpha_t}), \quad t = 2, \ldots, T,$$

where $\widehat{V}_{\alpha_t} = \left(V_\alpha^{-1} + l_n^{'}\Omega(\omega)l_n\right)^{-1}$, $\widehat{\mu}_{\alpha_t} = \widehat{V}_{\alpha_t}\left(V_\alpha^{-1}\mu_\alpha + l_n^{'}\Omega(\omega)Y_t^\alpha\right)$, and $\Omega(\omega) = R^{'}(\rho)R(\rho)/\sigma^2$.

3. Sampling step for β: Let $Y_t^\beta = S(\lambda)Y_t - C_n - \alpha_t l_n$ for $t = 1, 2, \ldots, T$. Then,

$$\beta|Y, \varrho, \sigma^2, \zeta, C_n, \alpha \sim N(\widehat{\mu}_\beta, \widehat{V}_\beta),$$

where $\widehat{V}_\beta = \left(V_\beta^{-1} + \sum_{t=1}^T X_t^{'}\Omega(\omega)X_t\right)^{-1}$, $\widehat{\mu}_\beta = \widehat{V}_\beta\left(V_\beta^{-1}\mu_\beta + \sum_{t=1}^T X_t^{'}\Omega(\omega)Y_t^\beta\right)$, and $\Omega(\omega) = R^{'}(\rho)R(\rho)/\sigma^2$.

4. Sampling step for σ^2:

$$\sigma^2|Y, \beta, \varrho, \zeta, C_n, \alpha \sim \text{IG}\,(a, b)\,, \tag{C.2}$$

where $a = a_0 + nT/2$, $b = b_0 + \frac{1}{2}\sum_{t=1}^T\left(Y_t^\beta - X_t\beta\right)^{'}\left(\sigma^2\Omega(\omega)\right)\left(Y_t^\beta - X_t\beta\right)$, and $\Omega(\omega) = R^{'}(\rho)R(\rho)/\sigma^2$.

5. Sampling step for ζ: Use Algorithm 3 to sample ζ.

Algorithm 2 (Estimation Algorithm for the dynamic model in (2.2))**.**

1. Sampling step for C_n: Let $Y_t^c = S(\lambda)Y_t - D(\gamma, \eta)Y_{t-1} - X_t\beta - \alpha_t l_n$ for $t = 1, 2, \ldots, T$. Then,

$$C_n|Y, \delta, \zeta, \alpha \sim N(\widehat{\mu}_c, \widehat{V}_c), \tag{C.3}$$

where $\widehat{V}_c = \left(V_c^{-1} + T\Omega(\omega)\right)^{-1}$, $\widehat{\mu}_c = \widehat{V}_c\left(V_c^{-1}\mu_c + \Omega(\omega)\sum_{t=1}^T Y_t^c\right)$, and $\Omega(\omega) = R^{'}(\rho)R(\rho)/\sigma^2$.

2. Sampling step for α: Let $Y_t^\alpha = S(\lambda)Y_t - D(\gamma, \eta)Y_{t-1} - X_t\beta - C_n$ for $t = 1, 2, \ldots, T$. Then,

$$\alpha_t|Y, \delta, \zeta, C_n, \alpha_{-t} \sim N(\widehat{\mu}_\alpha, \widehat{V}_\alpha), \quad t = 2, \ldots, T,$$

where $\widehat{V}_\alpha = \left(V_\alpha^{-1} + l_n^{'}\Omega(\omega)l_n\right)^{-1}$, $\widehat{\mu}_\alpha = \widehat{V}_\alpha\left(V_\alpha^{-1}\mu_\alpha + l_n^{'}\Omega(\omega)Y_t^\alpha\right)$, and $\Omega(\omega) = R^{'}(\rho)R(\rho)/\sigma^2$.

3. Sampling step for β: Let $Y_t^{\beta} = S(\lambda)Y_t - D(\gamma,\eta)Y_{t-1} - C_n - \alpha_t l_n$ for $t = 1, 2, \ldots, T$. Then,

$$\beta|Y, \sigma^2, \zeta, C_n, \alpha \sim N(\widehat{\mu}_{\beta}, \widehat{V}_{\beta}),$$

where $\widehat{V}_{\beta} = \left(V_{\beta}^{-1} + \sum_{t=1}^{T} X_t' \Omega(\omega) X_t\right)^{-1}$, $\widehat{\mu}_{\beta} = \widehat{V}_{\beta}\left(V_{\beta}^{-1}\mu_{\beta} + \sum_{t=1}^{T} X_t' \Omega(\omega) Y_t^{\beta}\right)$, and $\Omega(\omega) = R'(\rho)R(\rho)/\sigma^2$.

4. Sampling step for σ^2:

$$\sigma^2|Y, \beta, \zeta, C_n, \alpha \sim \mathrm{IG}\,(a, b)\,, \tag{C.4}$$

where $a = a_0 + nT/2$, $b = b_0 + \frac{1}{2}\sum_{t=1}^{T}\left(Y_t^{\beta} - X_t\beta\right)'\left(\sigma^2\Omega(\omega)\right)\left(Y_t^{\beta} - X_t\beta\right)$, and $\Omega(\omega) = R'(\rho)R(\rho)/\sigma^2$.

5. Sampling step for ζ: Use Algorithm 3 to sample ζ.

Algorithm 3 (The AM Algorithm for ζ)**.** *Let $a^{(g)}$ be the posterior draw generated at the gth iteration, where $a \in \{\zeta, \delta, C_n, \alpha\}$. Then, the AM algorithm consists of the following steps.*

1. Draw a candidate $\tilde{\zeta}$ as follows: At the iteration g for $g = 1, 2, \ldots, G$,
   (a) if $g \leq 2\bar{p}$, propose $\tilde{\zeta} \sim N\left(\zeta^{(g-1)}, \frac{(0.1)^2}{\bar{p}} \times I_{\bar{p}}\right)$, 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 $\tilde{\zeta} \sim 0.95 \times N\left(\zeta^{(g-1)}, \kappa\frac{(2.38)^2}{\bar{p}} \times \mathrm{Cov}\left(\zeta^{(0)}, \ldots, \zeta^{(g-1)}\right)\right) + 0.05 \times N\left(\zeta^{(g-1)}, \frac{(0.1)^2}{\bar{p}} \times I_{\bar{p}}\right)$.

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

3. Set the acceptance probability to

$$\Pr(\zeta^{(g-1)}, \tilde{\zeta}) = \min\left(\frac{p(Y|\tilde{\zeta}, \delta^{(g-1)}, C_n^{(g-1)}, \alpha^{(g-1)})}{p(Y|\zeta^{(g-1)}, \delta^{(g-1)}, C_n^{(g-1)}, \alpha^{(g-1)})}, 1\right).$$

   Then, return $\tilde{\zeta}$ with probability $\Pr(\zeta^{(g-1)}, \tilde{\zeta})$; otherwise return $\zeta^{(g-1)}$.

D Spatial Weights Matrices

In this section, we describe how we create the spatial weight matrices based on the rook and queen contiguity used in the Monte Carlo simulation. 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.

E Trace Plots for the Empirical Illustration

In this section, we provide the trace plots of the four models in the empirical illustration to investigate the convergence properties of the our samplers. Figure E.1 contains the trace plots for E-M1, Figure E.2 for E-M1D, Figure E.3 for E-M3 and Figure E.4 for E-M3D. These trace plots indicate that there are no converge issues.

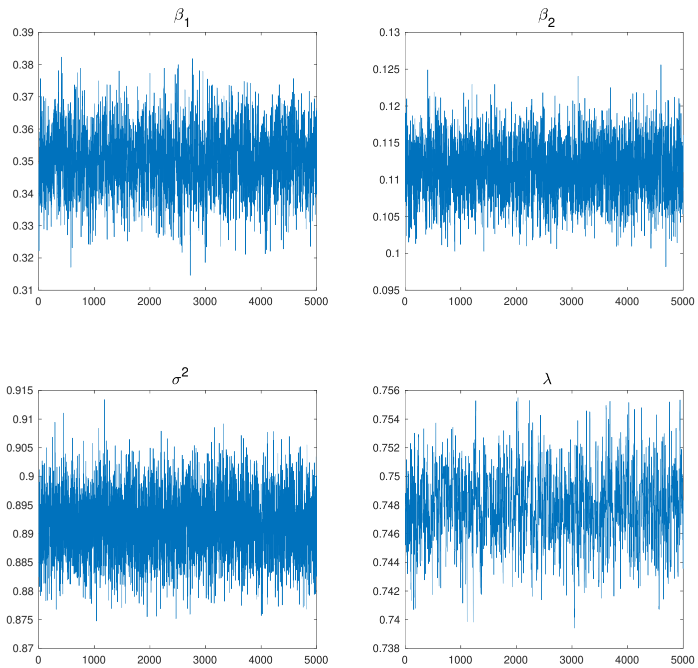

Figure E.1: Trace plots for E-M1 in the empirical illustration

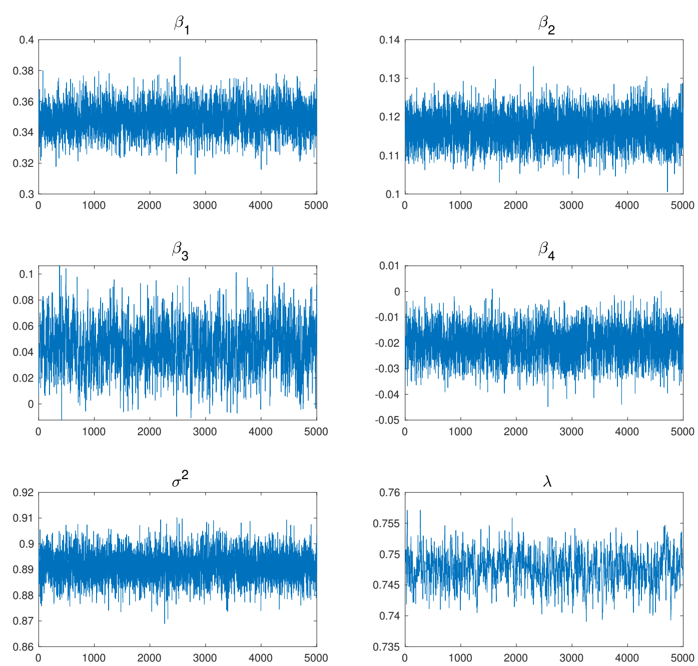

Figure E.2: Trace plots for E-M1D in the empirical illustration

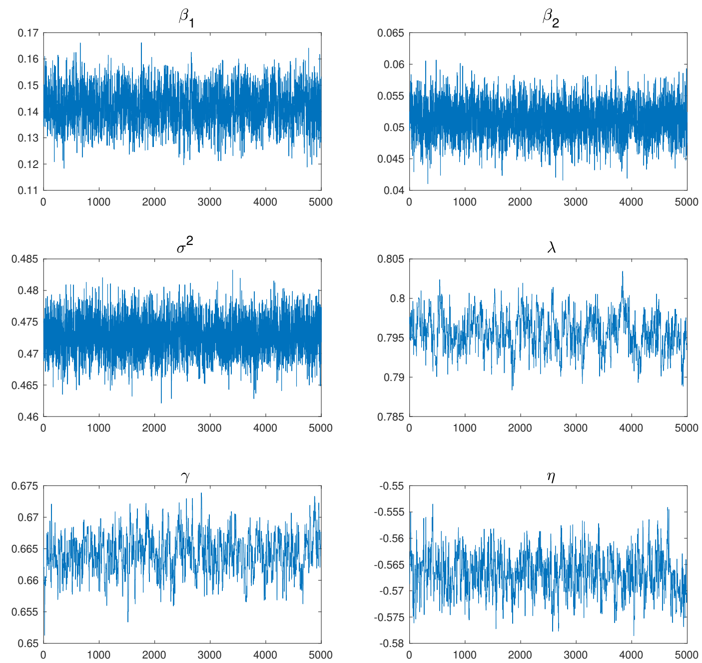

Figure E.3: Trace plots for E-M3 in the empirical illustration

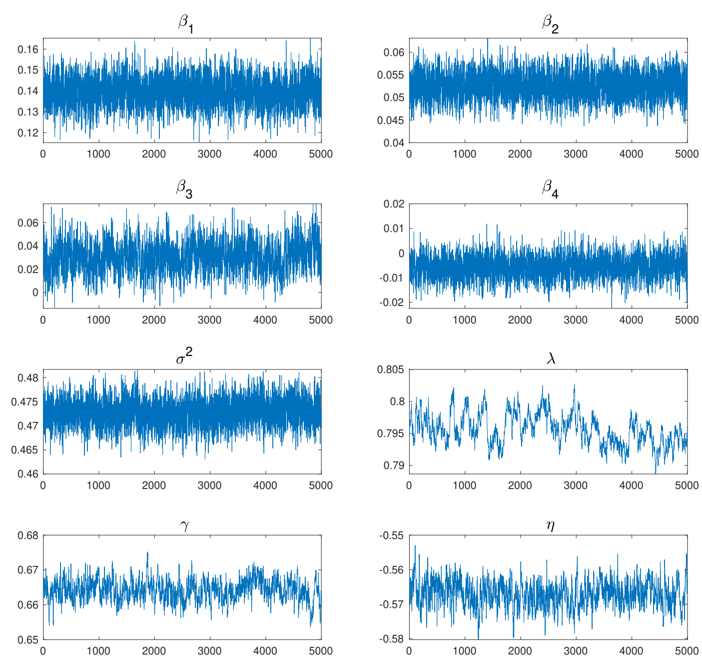

Figure E.4: Trace plots for E-M3D in the empirical illustration

References

Anselin, Luc (1984). "Specification tests on the structure of interaction in spatial econometric models". In: *Papers of the Regional Science Association* 54, pp. 165–182.

— (1986). "Non-nested tests of the weight structure in spatial autoregressive models: some Monte Carlo results". In: *Journal of Regional Science* 26.2, pp. 267–284.

— (1988). *Spatial econometrics: Methods and Models*. New York: Springer.

Anselin, Luc, Anil K. Bera, Raymond Florax, and Mann J. Yoon (1996). "Simple diagnostic tests for spatial dependence". In: *Regional Science and Urban Economics* 26.1, pp. 77 –104.

Anselin, Luc, Julie Le Gallo, and Hubert Jayet (2008). "Spatial Panel Econometrics". In: *The Econometrics of Panel Data*. Ed. by Laszlo Matyas et al. Vol. 46. Advanced Studies in Theoretical and Applied Econometrics. Springer Berlin Heidelberg.

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.

Bailey, Natalia, Sean Holly, and M. Hashem Pesaran (2016). "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence". In: *Journal of Applied Econometrics* 31.1, pp. 249–280.

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., 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.

Bao, Yong and Xiaoyan Zhou (2023). "Heterogeneous spatial dynamic panels with an application to US housing data". In: *Spatial Economic Analysis* 18.2, pp. 259–285.

Bera, Anil K., Osman Doğan, Süleyman Taşpınar, and Yufan Leiluo (2019). "Robust LM tests for spatial dynamic panel data models". In: *Regional Science and Urban Economics* 76, pp. 47–66.

Bresson, Georges, Badi H. Baltagi, and Alain Pirotte (2007). "Panel unit root tests and spatial dependence". In: *Journal of Applied Econometrics* 22.2, pp. 339–360.

Burnham, K.P. and D.R. Anderson (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Second Edition. New York: Springer.

Burridge, Peter (2012). "Improving the J Test in the SARAR Model by Likelihood-based Estimation". In: *Spatial Economic Analysis* 7.1, pp. 75–107.

Chan, Joshua C.C. and Angelia L. Grant (2015). "Pitfalls of estimating the marginal likelihood using the modified harmonic mean". In: *Economics Letters* 131, pp. 29–33.

Chan, Joshua C.C. and Angelia L. Grant (2016). “On the Observed-Data Deviance Information Criterion for Volatility Modeling”. In: *Journal of Financial Econometrics* 14.4, pp. 772–802.

Chib, Siddhartha (1995). “Marginal Likelihood from the Gibbs Output”. In: *Journal of the American Statistical Association* 90.432, pp. 1313–1321.

Chib, Siddhartha and Ivan Jeliazkov (2001). “Marginal Likelihood From the Metropolis–Hastings Output”. In: *Journal of the American Statistical Association* 96.453, pp. 270–281.

Elhorst, J. Paul (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels.* Springer Briefs in Regional Science. New York: Springer Berlin Heidelberg.

Frühwirth-Schnatter, Sylvia and Helga Wagner (2008). “Marginal likelihoods for non-Gaussian models using auxiliary mixture sampling”. In: *Computational Statistics & Data Analysis* 52.10, pp. 4608–4624.

Gelfand, A. E. and D. K. Dey (1994). “Bayesian Model Choice: Asymptotics and Exact Calculations”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 56.3, pp. 501–514. (Visited on 08/22/2022).

Geweke, John (1999). “Using simulation methods for bayesian econometric models: inference, development,and communication”. In: *Econometric Reviews* 18.1, pp. 1–73.

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

Han, Xiaoyi and Lung fei Lee (2013). “Model selection using J-test for the spatial autoregressive model vs. the matrix exponential spatial model”. In: *Regional Science and Urban Economics* 43.2, pp. 250–271.

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.

Jin, Fei and Lung fei Lee (2013). “Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances”. In: *Regional Science and Urban Economics* 43.4, pp. 590–616.

— (2018). “Outer-product-of-gradients tests for spatial autoregressive models”. In: *Regional Science and Urban Economics* 72, pp. 35–57.

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.

Kass, Robert E. and Adrian E. Raftery (1995). “Bayes Factors”. In: *Journal of the American Statistical Association* 90.430, pp. 773–795.

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.

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.

— (2010c). "Some recent developments in spatial panel data models". In: *Regional Science and Urban Economics* 40.5, pp. 255–271.

— (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.

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

Liu, Tuo and Lung fei Lee (2019). "A likelihood ratio test for spatial model selection". In: *Journal of Econometrics* 213.2, pp. 434–458.

Parent, Olivier and James P. LeSage (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.

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

Schwarz, Gideon (1978). "Estimating the Dimension of a Model". In: *The Annals of Statistics* 6.2, pp. 461 –464.

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, Ye (2022). "Unified M-estimation of matrix exponential spatial dynamic panel specification". In: *Econometric Reviews* 41.7, pp. 729–748.

Yang, Ye, Osman Doğan, and Süleyman Taşpınar (2022). "Model selection and model averaging for matrix exponential spatial models". In: *Econometric Reviews* 41.8, pp. 827–858.

— (2023). "Observed-data DIC for spatial panel data models". In: *Empirical Economics* 64, pp. 1281–1314.

Yang, Ye, Osman Doğan, and Süleyman Taşpınar (2024). "Estimation of Matrix Exponential Unbalanced Panel Data Models with Fixed Effects: An Application to US Outward FDI Stock". In: *Journal of Business & Economic Statistics* 42.2, pp. 469–484.

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.

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.