Heterogeneous Spatial Dynamic Panels with an Application to US Housing Data

Yong Bao*
Department of Economics
Purdue University

Xiaoyan Zhou†
Department of Economics
Purdue University

Abstract: This paper proposes two models that incorporate both heterogeneity and multiple sources of spatial correlation for dynamic panels. One uses convex combinations of them to form a single weight matrix. The second one includes explicitly different spatial weight matrices to form a higher-order model. We use a Bayesian scheme for model estimation by deriving the full conditional distributions of heterogeneous parameters. Our Monte Carlo experiments demonstrate their finite-sample performance relative to a baseline model. In our empirical study, we find the importance of including both geographic and non-geographic information in capturing correlations in real house price growth in the US.

Key Words: Heterogeneous spatial dynamic panels; Bayesian; Full conditional distributions; Real house price growth

JEL classification: C13, C23, R21, R31

*Corresponding author. Department of Economics, Purdue University, 403 W State St, West Lafayette, IN 47907, USA. E-mail: ybao@purdue.edu.

†Department of Economics, Purdue University, 403 W State St, West Lafayette, IN 47907, USA. E-mail: zhou1057@purdue.edu.

1. INTRODUCTION

Correlation among economic variables over time and across sections has long been the focal point of empirical analysis. The notion of autocorrelation or autoregression is the traditional approach to capturing correlation in the time domain and it is also extended to the cross-sectional framework by the so-called spatial autoregression.[1] The paradigm spatial autoregressive (SAR) model incorporates the correlations among different units in econometric models through a spatial weight matrix which characterizes the strength of connections among regions, markets, stores, individuals, and so on. A SAR model can be interpreted as a reduced-form regression of the outcome variable on its spatial lag via the weight matrix. For a given unit, its spatial lag can be understood as a weighted average across its possible "neighbors." As such, SAR models can be naturally extended to deal with panel data, when at each point of time there is a cross-sectional SAR. Further, dynamics can be straightforwardly introduced so that spatial dynamic panel models can be used when "lagged" variables, both in time and space, appear on the right-hand side of a panel regression and these lags aim to account for both spatial and temporal correlations. Cross-sectional SAR and spatial dynamic panel models are widely applied to research in various fields of economics including economic growth, public economics, social networks, and real estate economics.[2] The approaches of quasi-maximum likelihood (QML), generalized method of moments (GMM), instrumental variable (IV), indirect inference (II), and Bayesian Markov chain Monte Carlo (MCMC) have been used to estimate these models.[3]

Recently, there has been increasing attention in the literature to the estimation of spatial panels with heterogeneous coefficients. It is argued that on many occasions there are theoretical reasons to believe that the strength of spatial dependence would differ greatly across regions. LeSage and Chih (2018) propose the Bayesian MCMC approach to estimating heterogeneous spatial autoregressive (HSAR) panel models. LeSage et al. (2017) apply the Bayesian MCMC procedure to study duopoly pricing in German gasoline retail mar-

ket. Autant-Bernard and LeSage (2019) estimate the region-specific knowledge production functions based on a Bayesian HSAR model for 94 NUTS-3 regions in France. Aquaro et al. (2015, 2021) study the QML estimation of heterogeneous spatial autoregressive dynamic panel (HSARDP) models.

Despite the great interest in heterogeneous models, most current research assumes that the spatial weight matrix in the model is taken as given. However, sometimes there may be multiple candidate weight matrices that are relevant to the question under study. More importantly, different types of spatial correlation characterized by different weight matrices may have different signs and/or magnitudes. For example, sometimes we may want to incorporate information from both the first-order and second-order contiguity matrices, where the first-order one considers direct geographic neighbors and the second-order one characterizes neighbors of direct neighbors. It is usually expected that the first-order spillovers from direct neighbors are relatively stronger than the second-order spillovers. Also, in some applications, it may make more sense to use both spatial and non-spatial information to generate the spatial weight matrix (e.g., Debarsy and LeSage, 2022). For instance, when we examine the interactions of fiscal policies by local governments, the degree of connection between two regions may depend on not only how close the regions are but also how similar they are in terms of various aspects of socioeconomic conditions. To incorporate the information from multiple sources, Debarsy and LeSage (2022) discuss how to construct a spatial weight matrix using the convex combinations of different types of connectivity matrices. In their approach, the estimated convex combination weights also provide some implications on the relative importance of different information in determining the distance between two units. Lam and Souza (2020) specify the spatial weight matrix by adding a sparse adjustment matrix to the estimated best linear combination of different connectivity matrices based on the adaptive least absolute shrinkage and selection operator (LASSO).

In the heterogeneous framework, with the presence of multiple types of spatial correlation, it is natural and reasonable to allow for heterogeneity in the way that different sources of information are combined. For example, while geographic distance may matter equally for regions A and B, economic distance (or the degree of similarity of economic status) relative to other regions may matter a lot for region A, but not so much for region B, when each region is determining its spending on welfare programs, depending on their industrial structures, labor market characteristics, government compositions, and demographics.

The main contribution of this paper is to take both heterogeneity and multiple sources of spatial correlation into account and model them simultaneously. To achieve this, we propose two approaches. The first approach is to allow for the convex combination weights being heterogeneous across units when a single combined spatial weight matrix is used in HSARDP. We allow for different combination weights for the spatially lagged term and the spatial-temporally lagged term. Our second approach is to use a high-order HSARDP model to incorporate explicitly different information from multiple sources. Note that estimation and inference of the homogeneous version of high-order SAR models have been widely discussed in the literature (e.g., Lee and Liu, 2010; Gupta and Robinson, 2015; Han et al., 2017). However, incorporating heterogeneous spatial parameters allows for a more flexible pattern of spatial correlation, where the magnitudes and signs of different types of spatial spillover can vary across spatial units. Notably, the two approaches give rise to two mathematically equivalent models, but they have totally different implications. We use the Bayesian MCMC scheme for model estimation by deriving the full conditional distributions of various heterogeneous parameters, including the heterogeneous combination weights in the first approach. As an application, we apply our models to the real house price growth rates in 338 metropolitan statistical areas (MSAs) in the US from 1975 to 2019. We find that when both geographic and non-geographic information are used to capture possible spatial correlations, our models find many more MSAs that display significant positive

net spatial parameter estimates and more MSAs where income and population growths have significant positive spill-in impact estimates than otherwise would be found if only geographic information is used.

The plan of this paper is as follows. The next section introduces two models that incorporate both heterogeneity and multiple sources of spatial correlation. We derive the full conditional distributions of various parameters, which are needed for us to implement the Bayesian MCMC cycle for model estimation. The third section presents results from some Monte Carlo experiments by comparing a baseline model and our models. The fourth section contains our empirical study of US housing price growth. The last section concludes.

2. MODEL SPECIFICATION AND ESTIMATION

The standard HSARDP can be written as, for $i = 1, \cdots, N,\ t = 1, , \cdots, T$,

$$y_{it} = \psi_i y_{it}^{\star} + \phi_i y_{i,t-1}^{\star} + \lambda_i y_{i,t-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}, \tag{1}$$

where ε_{it} is the heteroskedastic idiosyncratic error term (with mean zero and variance σ_i^2), $y_{it}^{\star} = \sum_{j=1}^{N} w_{ij} y_{jt}$ defines a composite variable consisting of weighted outcome from neighboring units, ψ_i and ϕ_i represent the spatial dependence parameters for unit i, capturing the contemporaneous and temporal effects on units i's outcome from its neighboring units, respectively, λ_i is the first-order autoregressive parameter that relates unit i's outcome to its lagged value, $\boldsymbol{\beta}_i$ is a $k \times 1$ parameter vector associated with the exogenous regressors $\boldsymbol{x}_{it}$. Heterogeneity is modeled explicitly by indexing the parameters by i. In this set-up, a single spatial weight matrix is used, on the assumption that we are sure about how to define neighbors based on a single measure of spatial contiguity or connectivity. On many occasions, there may exist different ways to quantify the degree of connectivity and even if a single measure is agreed on, we may not be sure about what threshold value of this measure should be used to define neighbors. Thus in this paper we propose two approaches in the

presence of possibly multiple sources or measures of spatial correlation. The first one is to combine possible spatial weight matrices and form a single weight matrix, where the combination weights are also allowed to be heterogeneous across units. The second approach is to use a higher-order HSARDP model, where all possible weight matrices are used. In what follows, $\boldsymbol{\Sigma} = \text{Dg}(\sigma_1^2, \cdots, \sigma_N^2)$, $\text{Dg}(\cdot)$ denotes an operator that stacks diagonally its arguments in order such that it results in a matrix, $\text{dg}(\cdot)$ with a matrix argument stacks in order its argument's diagonal elements to form a column vector, $\mathbb{1}(\cdot)$ is the indicator function that takes on value 1 if its argument is true and zero otherwise, and $\mathbb{L}$ is the lag operator that shifts a time-indexed element/vector one period backward.

2.1 HSARDP with Weight Matrices Combined

Suppose there are in total q number of possible row-normalized weight matrices $\{\boldsymbol{W}_s\}_{s=1}^q$. Correspondingly, the (i,j)-th entries of $\boldsymbol{W}_s$ are denoted by $w_{ij}^{(s)}$. Now define $y_{it}^\star(\boldsymbol{\gamma}_i) = \sum_{s=1}^q \gamma_{is} \sum_{j=1}^N w_{ij}^{(s)} y_{jt}$, where $0 \leq \gamma_{is} \leq 1$ and $\sum_{s=1}^q \gamma_{is} = 1$ are the combination weights. This constitutes for unit i a composite term from neighboring units of q different natures with corresponding weights γ_{is}, $s = 1, \cdots, q$. Note that the vector of weights $\boldsymbol{\gamma}_i = (\gamma_{i1}, \cdots, \gamma_{iq})'$ is heterogeneous across i. This means that a unit may view differently the importance of various neighbors. We also define a composite lagged variable $y_{i,t-1}^\star(\boldsymbol{\delta}_i) = \sum_{s=1}^q \delta_{is} \sum_{j=1}^N w_{ij}^{(s)} y_{j,t-1}$, where $0 \leq \delta_{is} \leq 1$ and $\sum_{s=1}^q \delta_{is} = 1$, to allow for heterogeneous combination of neighbors that lag in both time and space. We treat $\boldsymbol{\gamma}_i$ and $\boldsymbol{\delta}_i$ as two different sets of parameters, recognizing that the importance of various contemporaneous neighbors may be different from that of neighbors from yesterday. Under this specification, we have

$$y_{it} = \psi_i y_{it}^\star(\boldsymbol{\gamma}_i) + \phi_i y_{i,t-1}^\star(\boldsymbol{\delta}_i) + \lambda_i y_{i,t-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}, \tag{2}$$

which has the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\phi}', \boldsymbol{\gamma}', \boldsymbol{\delta}', \boldsymbol{\lambda}', \boldsymbol{\beta}', \boldsymbol{\sigma}^{2\prime})'$, where $\boldsymbol{\psi} = (\psi_1, \cdots, \psi_N)'$, $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_N)'$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \cdots, \boldsymbol{\gamma}_N')'$, $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \cdots, \boldsymbol{\delta}_N')'$, $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_N)'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \cdots, \boldsymbol{\beta}_N')'$, and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \cdots, \sigma_N^2)'$.

For the homogeneous SAR with a single convexly combined weight matrix from row-normalized weight matrices, Debarsy and LeSage (2022) point out that the weight matrices cannot be highly correlated and that the spatial parameter cannot be equal to zero for identification purpose. We expect that similar conditions on the weight matrices need to be imposed in our heterogeneous framework and that $\Pr(\psi_i = 0) = 0$ and $\Pr(\phi_i = 0) = 0$ for each i. It is beyond the scope of this paper to derive a full set of conditions for identification. When there are only two candidate weight matrices, however, it is relatively straightforward to show sufficient conditions for identification of the spatial dependence parameters and combination weights. Since $\boldsymbol{\gamma}_i$ affects only the i-th row of $\boldsymbol{W}(\boldsymbol{\gamma})$, denoted by $\mathbf{w}_i(\boldsymbol{\gamma}_i)$, and ψ_i is the coefficient on $\mathbf{w}_i(\boldsymbol{\gamma}_i)\boldsymbol{y}_{\circ t}$, where $\boldsymbol{y}_{\circ t} = (y_{1t}, \cdots, y_{Nt})'$, we can look at the identification issue for $(\psi_i, \boldsymbol{\gamma}_i)'$ separately for each i. When $q = 2$, $\boldsymbol{\gamma}_i = (\gamma_{i1}, 1 - \gamma_{i1})'$, and therefore we only need to identify $(\psi_i, \gamma_{i1})'$. Suppose $(\psi_i, \gamma_{i1})'$ is not identifiable. In other words, there exist two different sets of $(\psi_i, \gamma_{i1})'$, which we respectively denote as $(\psi_i^{(a)}, \gamma_{i1}^{(a)})'$ and $(\psi_i^{(b)}, \gamma_{i1}^{(b)})'$, that result in the same $\psi_i \mathbf{w}_i(\boldsymbol{\gamma}_i)$:

$$\psi_i^{(a)}\gamma_{i1}^{(a)}\mathbf{w}_i^{(1)} + \psi_i^{(a)}(1 - \gamma_{i1}^{(a)})\mathbf{w}_i^{(2)} = \psi_i^{(b)}\gamma_{i1}^{(b)}\mathbf{w}_i^{(1)} + \psi_i^{(b)}(1 - \gamma_{i1}^{(b)})\mathbf{w}_i^{(2)}, \tag{3}$$

where $\mathbf{w}_i^{(s)}$ is the i-th row of $\boldsymbol{W}_s$. Since both weight matrices are row-sum normalized, by multiplying a column vector of ones to both sides of (3), we obtain

$$\psi_i^{(a)}\gamma_{i1}^{(a)} + \psi_i^{(a)}(1 - \gamma_{i1}^{(a)}) = \psi_i^{(b)}\gamma_{i1}^{(b)} + \psi_i^{(b)}(1 - \gamma_{i1}^{(b)}). \tag{4}$$

Therefore, we have $\psi_i^{(a)} = \psi_i^{(b)}$. Furthermore, if $\psi_i^{(a)} = \psi_i^{(b)} \neq 0$, we have

$$\gamma_{i1}^{(a)}\mathbf{w}_i^{(1)} + (1 - \gamma_{i1}^{(a)})\mathbf{w}_i^{(2)} = \gamma_{i1}^{(b)}\mathbf{w}_i^{(1)} + (1 - \gamma_{i1}^{(b)})\mathbf{w}_i^{(2)}. \tag{5}$$

Then we can get $\gamma_{i1}^{(a)}w_{ij}^{(1)} + (1-\gamma_{i1}^{(a)})w_{ij}^{(2)} = \gamma_{i1}^{(b)}w_{ij}^{(1)} + (1-\gamma_{i1}^{(b)})w_{ij}^{(2)}$, and therefore $(\gamma_{i1}^{(a)} - \gamma_{i1}^{(b)})w_{ij}^{(1)} = (\gamma_{i1}^{(a)} - \gamma_{i1}^{(b)})w_{ij}^{(2)}$, for each $j \in \{1, \cdots, N\}$. Recall that $w_{ij}^{(s)}$ denote the (i,j)-th element of $\boldsymbol{W}_s$. Hence, it is obvious that $\gamma_{i1}^{(a)} \neq \gamma_{i1}^{(b)}$ only if $\mathbf{w}_i^{(1)} = \mathbf{w}_i^{(2)}$. In other words, as long as the i-th row of $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are not identical, $(\psi_i, \gamma_{i1})'$ is identifiable, assuming that both weight matrices are row-sum normalized and that the spatial coefficients are nonzero. Following the same procedure for $(\phi_i, \delta_{i1})'$, we see that similar results hold.[4]

Let $\boldsymbol{\varepsilon}_{\circ t} = (\varepsilon_{1t}, \cdots, \varepsilon_{Nt})'$, $\boldsymbol{x}_{\circ t} = (\boldsymbol{x}_{1t}', \cdots, \boldsymbol{x}_{Nt}')'$, $\boldsymbol{W}(\boldsymbol{\gamma})$ consist of $\sum_{s=1}^{q}\gamma_{is}\mathbf{w}_i^{(s)}$ as its i-th row, $\boldsymbol{W}(\boldsymbol{\delta})$ consist of $\sum_{s=1}^{q}\delta_{is}\mathbf{w}_i^{(s)}$ as its i-th row, $\boldsymbol{\Psi} = \mathrm{Dg}(\psi_1, \cdots, \psi_N)$, $\boldsymbol{\Phi} = \mathrm{Dg}(\phi_1, \cdots, \phi_N)$, $\boldsymbol{\Lambda} = \mathrm{Dg}(\lambda_1, \cdots, \lambda_N)'$, and $\boldsymbol{B} = \mathrm{Dg}(\boldsymbol{\beta}_1', \cdots, \boldsymbol{\beta}_N')$. Then we can write (2) compactly as

$$\boldsymbol{A}\boldsymbol{y}_{\circ t} = \boldsymbol{C}\boldsymbol{y}_{\circ t-1} + \boldsymbol{B}\boldsymbol{x}_{\circ t} + \boldsymbol{\varepsilon}_{\circ t}, \tag{6}$$

where $\boldsymbol{A} = \boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}(\boldsymbol{\gamma})$ and $\boldsymbol{C} = \boldsymbol{\Phi}\boldsymbol{W}(\boldsymbol{\delta}) + \boldsymbol{\Lambda}$. In (6), the matrix $\boldsymbol{W}(\boldsymbol{\gamma})$ captures contemporaneous spatial effects from different sources, whereas spatial-temporal effects are absorbed in $\boldsymbol{W}(\boldsymbol{\delta})$.

Given the exogenous $\boldsymbol{x}_{\circ t}$, if we assume stationarity (so that $\boldsymbol{y}_{\circ t}$ and $\boldsymbol{y}_{\circ t-1}$ have the same distribution), then $\mathrm{E}(\boldsymbol{y}_{\circ t}) = (\boldsymbol{A}-\boldsymbol{C})^{-1}\boldsymbol{B}\boldsymbol{x}_{\circ t}$, $\mathrm{Var}(\boldsymbol{y}_{\circ t}) = \mathrm{vec}^{-1}((\boldsymbol{A}\otimes\boldsymbol{A} - \boldsymbol{C}\otimes\boldsymbol{C})^{-1}\mathrm{vec}(\boldsymbol{\Sigma})) \equiv \boldsymbol{\Omega}$, and $\mathrm{Cov}(\boldsymbol{y}_{\circ t}, \boldsymbol{y}_{\circ t-1}) = \boldsymbol{C}\boldsymbol{\Omega}$. Note that the condition for stationarity is that the characteristic roots of the matrix $\boldsymbol{A}^{-1}\boldsymbol{C}$ are within the unit circle. For a homogeneous spatial dynamic panel with one spatial weight matrix ($q=1$), Elhorst (2012) discusses the implied parameter constraints. Let $\mathbf{H} \in \mathbb{R}^{3N+2N(q-1)}$ be the parameter space of $(\boldsymbol{\eta}', \boldsymbol{\gamma}', \boldsymbol{\delta}')$, where $\boldsymbol{\eta} = (\boldsymbol{\eta}_1', \cdots, \boldsymbol{\eta}_N')'$, $\boldsymbol{\eta}_i = (\psi_i, \phi_i, \lambda_i)'$, such that the characteristic roots of the matrix $\boldsymbol{A}^{-1}\boldsymbol{C}$ are inside the unit circle.[5] (When we put $\boldsymbol{\eta}_i \in \mathbf{H}$, it means that $\boldsymbol{\eta}_i$ is in the relevant subspace of $\mathbf{H}$. The same applies to $\boldsymbol{\gamma}_i \in \mathbf{H}$ and $\boldsymbol{\delta}_i \in \mathbf{H}$.) Thus under the stationarity condition, we can put down the joint sample log likelihood function of $\boldsymbol{y} = (\boldsymbol{y}_{\circ 1}', \cdots, \boldsymbol{y}_{\circ T}')'$ as

$$\mathcal{L}(\boldsymbol{y}|\boldsymbol{\theta}) = -\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\ln\det(\boldsymbol{\Upsilon}) - \frac{1}{2}[\boldsymbol{y} - \mathrm{E}(\boldsymbol{y})]'\boldsymbol{\Upsilon}^{-1}[\boldsymbol{y} - \mathrm{E}(\boldsymbol{y})], \tag{7}$$

where $\mathrm{E}(\boldsymbol{y})$ consists of $(\boldsymbol{A}-\boldsymbol{C})^{-1}\boldsymbol{B}\boldsymbol{x}_{\circ t}$, $t=1,\cdots,T$, and $\boldsymbol{\Upsilon}$ is an $NT\times NT$ tri-block-diagonal matrix with $\boldsymbol{\Omega}$, $\boldsymbol{C\Omega}$, and $\boldsymbol{\Omega C}'$ spanning its main diagonal, super-diagonal, and sub-diagonal blocks, respectively.

On the other hand, conditional on $\boldsymbol{y}_{\circ t-1}$, we have $\mathrm{E}(\boldsymbol{y}_{\circ t}|\boldsymbol{y}_{\circ t-1})=\boldsymbol{A}^{-1}\boldsymbol{C}\boldsymbol{y}_{\circ t-1}+\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{x}_{\circ t}\equiv\boldsymbol{\mu}_t$ and $\mathrm{Var}(\boldsymbol{y}_{\circ t}|\boldsymbol{y}_{\circ t-1})=\boldsymbol{A}^{-1}\boldsymbol{\Sigma}\boldsymbol{A}^{-1'}$. Then joint sample log likelihood function of $(\boldsymbol{y}_{\circ 0}',\boldsymbol{y}')'$ can be put as

$$\begin{aligned}\mathscr{L}(\boldsymbol{y}_{\circ 0},\boldsymbol{y}|\boldsymbol{\theta}) &= \ln(\mathrm{pdf}(\boldsymbol{y}_{\circ 0}))-\frac{NT}{2}\ln(2\pi)-\frac{T}{2}\ln\det(\boldsymbol{A}^{-1}\boldsymbol{\Sigma}\boldsymbol{A}^{-1'})\\ &\quad -\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{y}_{\circ t}-\boldsymbol{\mu}_t)'\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{y}_{\circ t}-\boldsymbol{\mu}_t)\\ &= \ln(\mathrm{pdf}(\boldsymbol{y}_{\circ 0}))-\frac{NT}{2}\ln(2\pi)-\frac{T}{2}\ln\det(\boldsymbol{\Sigma})+T\ln\det(\boldsymbol{A})\\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_{\circ t}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_{\circ t}. \qquad (8)\end{aligned}$$

We could establish asymptotic results of the QML estimator that maximizes (7) or (8) following the lines of Aquaro et al. (2021) and a necessary condition for consistency of the estimator is that $T\to\infty$, namely, the so-called large panels.[6] In the following, we focus on instead the Bayesian approach of estimation and inference. Given the priors and likelihood function, we can derive the full conditional distribution for each set of parameters and start the MCMC cycle. Instead of using the likelihood function (7) that is based on the unconditional distribution, which involves the inverse of the $NT\times NT$ matrix $\boldsymbol{\Upsilon}$, we treat $\boldsymbol{y}_{\circ 0}$ as a vector of additional parameters and use the likelihood function (8).

For parameters indexed by i, we can assign the following prior distributions and they are independent across i. We assign a normal prior $\boldsymbol{\beta}_i\sim\mathrm{N}(\boldsymbol{\xi}_i,\boldsymbol{V}_i)$, where $\boldsymbol{V}_i=v_i^2\boldsymbol{I}_k$, and an inverse Gamma prior $\sigma_i^2\sim\mathrm{IG}(a_i,b_i)$. For the autoregressive parameters ψ_i, ϕ_i, and λ_i, we assign uniform (on the interval $(-1,1)$) priors (subject to the stationarity parameter constraint). For the combination parameters, $\boldsymbol{\gamma}_i\sim\mathrm{DIR}(g_1,\cdots,g_q)$ and $\boldsymbol{\delta}_i\sim\mathrm{DIR}(d_1,\cdots,d_q)$, namely, Dirichlet distributions with corresponding parameters. If no useful prior informa-

tion regarding the combination weights is available, we can set uniform priors for $\boldsymbol{\gamma}_i$ and $\boldsymbol{\delta}_i$, namely, $g_1 = \cdots = g_q = 1$ and $d_1 = \cdots = d_q = 1$. Lastly, the prior distribution of $\boldsymbol{y}_{\circ 0}$ is specified as $\mathrm{N}(\boldsymbol{\Xi}\boldsymbol{x}_{\circ 0}, \boldsymbol{I}_N)$, where $\boldsymbol{\Xi} = \mathrm{Dg}(\boldsymbol{\xi}_1', \cdots, \boldsymbol{\xi}_N')$. The prior mean and variance of $\boldsymbol{y}_{\circ 0}$ are based on the derived unconditional mean and variance of $\boldsymbol{y}_{\circ t}$ by imposing $\psi_i = \phi_i = \lambda_i = 0$, $\sigma_i^2 = 1$, and $\boldsymbol{\beta}_i = \boldsymbol{\xi}_i$ for all i.

To facilitate presentation of the full conditional distributions for each set of parameters in the MCMC cycle, we use the following notation to denote cross-sectional terms: $\boldsymbol{y}_{i\circ} = (y_{i1}, \cdots, y_{iT})'$, $\mathbb{L}\boldsymbol{y}_{i\circ} = (y_{i0}, \cdots, y_{i,T-1})'$, $\boldsymbol{\varepsilon}_{i\circ} = (\varepsilon_{i1}, \cdots, \varepsilon_{iT})'$, $\boldsymbol{X}_{i\circ} = (\boldsymbol{x}_{i1}', \cdots, \boldsymbol{x}_{iT}')'$, $\boldsymbol{y}_{i\circ}^\star(\boldsymbol{\gamma}_i) = (y_{i1}^\star(\boldsymbol{\gamma}_i), \cdots, y_{iT}^\star(\boldsymbol{\gamma}_i))'$, $\mathbb{L}\boldsymbol{y}_{i\circ}^\star(\boldsymbol{\delta}_i) = (y_{i,0}^\star(\boldsymbol{\delta}_i), \cdots, y_{i,T-1}^\star(\boldsymbol{\delta}_i))'$. Obviously, $\boldsymbol{\varepsilon}_{i\circ} = \boldsymbol{y}_{i\circ} - \psi_i \boldsymbol{y}_{i\circ}^\star(\boldsymbol{\gamma}_i) - \phi_i \mathbb{L}\boldsymbol{y}_{i\circ}^\star(\boldsymbol{\delta}_i) - \lambda_i \mathbb{L}\boldsymbol{y}_{i\circ} - \boldsymbol{X}_{i\circ}\boldsymbol{\beta}_i$ and $\sum_{t=1}^T \boldsymbol{\varepsilon}_{\circ t}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_{\circ t} = \sum_{i=1}^N (\boldsymbol{\varepsilon}_{i\circ}' \boldsymbol{\varepsilon}_{i\circ} / \sigma_i^2)$. Further, let $\boldsymbol{\beta}_{-i} = (\boldsymbol{\beta}_1', \cdots, \boldsymbol{\beta}_{i-1}', \boldsymbol{\beta}_{i+1}', \cdots, \boldsymbol{\beta}_N')'$ and $\boldsymbol{\sigma}_{-i}^2$, $\boldsymbol{\eta}_{-i}$, $\boldsymbol{\gamma}_{-i}$, etc. are defined similarly. Note that $P(\boldsymbol{y}|\boldsymbol{y}_{\circ 0}, \boldsymbol{\theta}) = \exp[\mathcal{L}(\boldsymbol{y}_{\circ 0}, \boldsymbol{y}|\boldsymbol{\theta}) - \ln(\mathrm{pdf}(\boldsymbol{y}_{\circ 0}))]$.

First, the full conditional distribution of each $\boldsymbol{\beta}_i$ is

$$
\begin{aligned}
&P(\boldsymbol{\beta}_i|\boldsymbol{y}, \boldsymbol{\beta}_{-i}, \boldsymbol{\sigma}^2, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{y}_{\circ 0}) \propto P(\boldsymbol{y}_{\circ 0}) P(\boldsymbol{y}|\boldsymbol{y}_{\circ 0}, \boldsymbol{\theta}) P(\boldsymbol{\beta}_i) \\
&\propto \exp\left(-\frac{1}{2\sigma_i^2}\boldsymbol{\varepsilon}_{i\circ}'\boldsymbol{\varepsilon}_{i\circ}\right) \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\xi}_i)' \boldsymbol{V}_i^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\xi}_i)\right] \\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}_i' \left(\frac{\boldsymbol{X}_{i\circ}'\boldsymbol{X}_{i\circ}}{\sigma_i^2} + \boldsymbol{V}_i^{-1}\right) \boldsymbol{\beta}_i\right\} \\
&\quad \times \exp\left\{\boldsymbol{\beta}_i' \left[\frac{(\boldsymbol{X}_{i\circ}'(\boldsymbol{y}_{i\circ} - \psi_i \boldsymbol{y}_{i\circ}^\star(\boldsymbol{\gamma}_i) - \phi_i \mathbb{L}\boldsymbol{y}_{i\circ}^\star(\boldsymbol{\delta}_i) - \lambda_i \mathbb{L}\boldsymbol{y}_{i\circ})}{\sigma_i^2} + \boldsymbol{V}_i^{-1}\boldsymbol{\xi}_i\right]\right\} \\
&= \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_i' \boldsymbol{D}_i^{-1} \boldsymbol{\beta}_i - 2\boldsymbol{\beta}_i' \boldsymbol{D}_i^{-1} \hat{\boldsymbol{\beta}}_i)\right] \\
&\propto \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)' \boldsymbol{D}_i^{-1} (\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)\right], \qquad (9)
\end{aligned}
$$

where $\boldsymbol{D}_i = (\boldsymbol{X}_{i\circ}'\boldsymbol{X}_{i\circ}/\sigma_i^2 + \boldsymbol{V}_i^{-1})^{-1}$ and $\hat{\boldsymbol{\beta}}_i = \boldsymbol{D}_i[(\boldsymbol{X}_{i\circ}'(\boldsymbol{y}_{i\circ} - \psi_i \boldsymbol{y}_{i\circ}^\star(\boldsymbol{\gamma}_i) - \phi_i \mathbb{L}\boldsymbol{y}_{i\circ}^\star(\boldsymbol{\delta}_i) - \lambda_i \mathbb{L}\boldsymbol{y}_{i\circ})/\sigma_i^2 + \boldsymbol{V}_i^{-1}\boldsymbol{\xi}_i]$. Hence, $(\boldsymbol{\beta}_i|\boldsymbol{y}, \boldsymbol{\beta}_{-i}, \boldsymbol{\sigma}^2, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{y}_{\circ 0}) \sim \mathrm{N}(\hat{\boldsymbol{\beta}}_i, \boldsymbol{D}_i)$.

Next, the full conditional distribution of each σ_i^2 is

$$
P(\sigma_i^2|\boldsymbol{y}, \boldsymbol{\sigma}_{-i}^2, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{y}_{\circ 0}) \propto P(\boldsymbol{y}_{\circ 0}) P(\boldsymbol{y}|\boldsymbol{y}_{\circ 0}, \boldsymbol{\theta}) P(\sigma_i^2)
$$

$$\propto (\sigma_i^2)^{-(\frac{T}{2}+a_i+1)} \exp\left[-\frac{1}{\sigma_i^2}\left(\frac{1}{2}\boldsymbol{\varepsilon}_{i\circ}'\boldsymbol{\varepsilon}_{i\circ}+b_i\right)\right]. \tag{10}$$

It follows that $(\sigma_i^2|\boldsymbol{y},\sigma_{-i}^2,\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{y}_{\circ 0}) \sim \text{IG}(\hat{a}_i,\hat{b}_i)$, where $\text{IG}(\hat{a}_i,\hat{b}_i)$ denotes the inverse Gamma distribution with the shape parameter $\hat{a}_i = T/2 + a_i$ and scale parameter $\hat{b}_i = \boldsymbol{\varepsilon}_{i\circ}'\boldsymbol{\varepsilon}_{i\circ}/2 + b_i$.

The full conditional distribution of $\boldsymbol{\eta}_i$ is

$$\begin{aligned} &P(\boldsymbol{\eta}_i|\boldsymbol{y},\boldsymbol{\eta}_{-i},\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{y}_{\circ 0}) \propto P(\boldsymbol{y}_{\circ 0})P(\boldsymbol{y}|\boldsymbol{y}_{\circ 0},\boldsymbol{\theta})P(\boldsymbol{\eta}_i) \\ &\propto \left|\boldsymbol{A}\right|^T \exp\left[-\frac{1}{2\sigma_i^2}\boldsymbol{\varepsilon}_{i\circ}'\boldsymbol{\varepsilon}_{i\circ}\right]\mathbb{1}(\boldsymbol{\eta}_i \in \mathbf{H}). \end{aligned} \tag{11}$$

This is not a standard distribution and we need to use the Metropolis-Hastings algorithm to sample. However, the standard Metropolis-Hastings algorithm used by LeSage and Chih (2018) turns out to be computationally inefficient for sampling $\boldsymbol{\eta}_i$. Hence, we are going to instead apply the adaptive Metropolis (AM) algorithm adopted by Han and Lee (2016), which specifies the covariance matrix of the proposal distribution based on the historical MCMC draws of $\boldsymbol{\eta}_i$. To be more specific, the proposal distribution under the AM algorithm of Roberts and Rosenthal (2009) for the r^{th} iteration during the burn-in period is given as below (recall that $\boldsymbol{\eta}_i = (\psi_i,\phi_i,\lambda_i)'$ is of dimension 3):

$$P(\boldsymbol{\eta}_i^c|\boldsymbol{\eta}_i^{[1:r-1]}) = \begin{cases} \text{N}(\boldsymbol{\eta}_i^{(r-1)},(0.1^2/3)\boldsymbol{I}_3), & r \leq 6 \\ (1-\zeta)\text{N}(\boldsymbol{\eta}_i^{(r-1)},(2.38^2/3)\boldsymbol{\Delta}) + \zeta\text{N}(\boldsymbol{\eta}_i^{(r-1)},(0.1^2/3)\boldsymbol{I}_3), & r > 6 \end{cases} \tag{12}$$

where $\boldsymbol{\eta}_i^{[1:r-1]} = (\boldsymbol{\eta}_i^{(1)},\cdots,\boldsymbol{\eta}_i^{(r-1)})$ denote the historical draws from the previous $r-1$ iterations, and $\boldsymbol{\Delta}$ is the sample covariance matrix of the historical draws $\boldsymbol{\eta}_i^{[1:r-1]}$. Therefore, after the first 6 iterations, the proposal distribution is a combination of two normal distributions, where the first component is based on the historical MCMC draws and the second component is aimed at preventing the procedure from producing a singular proposal covariance matrix. The combination weight may be set as $\zeta = 0.05$ following Roberts and Rosenthal (2009). By making use of the information from historical draws, the AM algo-

rithm can lead to more efficient mixing and faster convergence of the MCMC draws. After the burn-in period, the proposal distribution will no longer be adjusted and will be fixed as the proposal distribution used in the last burn-in iteration. Furthermore, the acceptance probability of a candidate $\boldsymbol{\eta}_i^c$ is calculated as

$$\nu(\boldsymbol{\eta}_i^\dagger, \boldsymbol{\eta}_i^c) = \min\left\{\frac{P(\boldsymbol{\eta}_i^c|\boldsymbol{y}, \boldsymbol{\eta}_{-i}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{y}_{\circ 0})}{P(\boldsymbol{\eta}_i^\dagger|\boldsymbol{y}, \boldsymbol{\eta}_{-i}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{y}_{\circ 0})}, 1\right\}, \tag{13}$$

where $\boldsymbol{\eta}_i^\dagger$ is the current draw of $\boldsymbol{\eta}_i$.

The full conditional distribution of each $\boldsymbol{\gamma}_i$ is

$$\begin{aligned} &P(\boldsymbol{\gamma}_i|\boldsymbol{y}, \boldsymbol{\gamma}_{-i}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{y}_{\circ 0}) \propto P(\boldsymbol{y}_{\circ 0})P(\boldsymbol{y}|\boldsymbol{y}_{\circ 0}, \boldsymbol{\theta})P(\boldsymbol{\gamma}_i) \\ &\propto \left|\boldsymbol{A}\right|^T \exp\left[-\frac{1}{2\sigma_i^2}\boldsymbol{\varepsilon}_{i\circ}'\boldsymbol{\varepsilon}_{i\circ}\right] \prod_{s=1}^{q}(\gamma_{is})^{g_s-1}\mathbb{1}\left(\sum_{s=1}^{q}\gamma_{is} = 1, 0 \leq \gamma_{is} \leq 1, \boldsymbol{\gamma}_i \in \mathbf{H}\right). \end{aligned} \tag{14}$$

The conditional distribution of each $\boldsymbol{\delta}_i$ can be derived similarly. Again, since the full conditional distributions for each $\boldsymbol{\gamma}_i$ and $\boldsymbol{\delta}_i$ are both nonstandard, we resort to the AM algorithm.

Lastly, the full conditional distribution of $\boldsymbol{y}_{\circ 0}$ is

$$\begin{aligned} &P(\boldsymbol{y}_{\circ 0}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) \propto P(\boldsymbol{y}_{\circ 0})P(\boldsymbol{y}|\boldsymbol{y}_{\circ 0}, \boldsymbol{\theta}) \\ &\propto \exp\left[-\frac{1}{2}(\boldsymbol{y}_{\circ 0}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}_{\circ 0} - 2\boldsymbol{y}_{\circ 0}'\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\mu}})\right], \end{aligned} \tag{15}$$

where $\hat{\boldsymbol{\Omega}} = (\boldsymbol{C}'\boldsymbol{\Sigma}^{-1}\boldsymbol{C} + \boldsymbol{I}_N)^{-1}$, and $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\Omega}}[\boldsymbol{C}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}\boldsymbol{y}_{\circ 1} - \boldsymbol{B}\boldsymbol{x}_{\circ 0}) + \boldsymbol{\Xi}\boldsymbol{x}_{\circ 0}]$. Hence, $(\boldsymbol{y}_{\circ 0}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) \sim \mathrm{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Omega}})$.

To summarize, starting from some initial values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\sigma}^{2(0)}$, $\boldsymbol{\eta}^{(0)}$, $\boldsymbol{\gamma}^{(0)}$, $\boldsymbol{\delta}^{(0)}$, and $\boldsymbol{y}_{\circ 0}^{(0)}$, we can implement the Gibbs sampler by sampling through the full conditional distributions of $\boldsymbol{\beta}_i$, σ_i^2, $\boldsymbol{\eta}_i$, $\boldsymbol{\gamma}_i$, $\boldsymbol{\delta}_i$, and $\boldsymbol{y}_{\circ 0}$ via (9)–(15) for each $i \in \{1, 2, ..., N\}$ in each iteration $r \in \{1, 2, \cdots, R\}$, where R denotes the total number of simulations.[7] Point estimates for parameters are from the posterior means and "significance" may be signified by the highest posterior density Bayesian credible intervals.

2.2 High-order HSARDP

Instead of using the combination of multiple connectivity matrices to specify a single spatial weight matrix, high-order spatial autoregressive models treat these connectivity matrices as different spatial weight matrices. A high-order HSARDP can be written as below.

$$y_{it} = \sum_{s=1}^{q} \psi_i^{(s)} y_{it}^{(s)\star} + \sum_{s=1}^{q} \phi_i^{(s)} y_{i,t-1}^{(s)\star} + \lambda_i y_{i,t-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}, \tag{16}$$

where $y_{it}^{(s)\star} = \sum_{j=1}^{N} w_{ij}^{(s)} y_{jt}$. Now define $\boldsymbol{A} = \boldsymbol{I}_N - \sum_{s=1}^{q} \boldsymbol{\Psi}_s \boldsymbol{W}_s$ and $\boldsymbol{C} = \sum_{s=1}^{q} \boldsymbol{\Phi}_s \boldsymbol{W}_s + \boldsymbol{\Lambda}$, where $\boldsymbol{\Psi}_s = \mathrm{Dg}(\psi_1^{(s)}, \cdots, \psi_N^{(s)})$, and $\boldsymbol{\Phi}_s = \mathrm{Dg}(\phi_1^{(s)}, \cdots, \phi_N^{(s)})$. We can derive similarly the sample log likelihood function, either the unconditional or conditional one as in the previous subsection. The cross-section $T \times 1$ vector of errors is defined accordingly as $\boldsymbol{\varepsilon}_{i\circ} = \boldsymbol{y}_{i\circ} - \sum_{s=1}^{q} \psi_i^{(s)} \boldsymbol{y}_{i\circ}^{(s)\star} - \sum_{s=1}^{q} \phi_i^{(s)} \mathbb{L}\boldsymbol{y}_{i\circ}^{(s)\star} - \lambda_i \mathbb{L}\boldsymbol{y}_{i\circ} - \boldsymbol{X}_{i\circ}\boldsymbol{\beta}_i$. Let $\boldsymbol{\kappa}_i = (\psi_i^{(1)}, \cdots, \psi_i^{(q)}, \phi_i^{(1)}, \cdots, \phi_i^{(q)}, \lambda_i)'$ and $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_1', \cdots, \boldsymbol{\kappa}_N')'$. The parameter space of $\boldsymbol{\kappa}$ is $\mathbf{K} \in \mathbb{R}^{N(2q+1)}$ such that the characteristic roots of the matrix $\boldsymbol{A}^{-1}\boldsymbol{C}$ are inside the unit circle. (When we put $\boldsymbol{\kappa}_i \in \mathbf{K}$, it means that $\boldsymbol{\kappa}_i$ is in the relevant subspace of $\mathbf{K}$.) The vector of parameters in (16) can be collected as $\boldsymbol{\theta} = (\boldsymbol{\kappa}', \boldsymbol{\beta}', \boldsymbol{\sigma}^{2\prime})'$.

The cycle of the MCMC algorithm outlined in the previous section carries over to $\boldsymbol{\beta}_i$, σ_i^2, and $\boldsymbol{\kappa}_i$ for each i. When sampling $\boldsymbol{\kappa}_i$ conditional on $\boldsymbol{y}, \boldsymbol{\kappa}_{-i}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{y}_{\circ 0}$, we need to maintain the stationarity condition $\boldsymbol{\kappa}_i \in \mathbf{K}$. Mathematically, one can think of (6) as a restricted version of (16). In (6), the first term on the right-hand side is $\sum_{s=1}^{q} \psi_i \gamma_{is} \sum_{j=1}^{N} w_{ij}^{(s)} y_{jt}$ and in (16) it is $\sum_{s=1}^{q} \psi_i^{(s)} \sum_{j=1}^{N} w_{ij}^{(s)} y_{jt}$, so $\psi_i^{(s)}$ can be interpreted as $\psi_i \gamma_{is}$. (Similar analysis applies to the second terms in (6) and (16).) With $\gamma_{is} \in [0, 1]$, spatial connections from all sources may matter and they have the same direction (determined by the sign of ψ_i) for each unit under (6). Yet, for the higher-order HSARDP, the spatial parameters on different weight matrices for the same unit can have different signs.

2.3 Marginal Effects

In addition to the point estimates obtained using the MCMC procedure, we may also be interested in estimating the marginal effects of exogenous explanatory variables on the outcome variable. Let $\boldsymbol{B}_l = \mathrm{Dg}(\beta_{1l}, \cdots, \beta_{Nl})$, and $\boldsymbol{x}_{\circ tl} = (x_{1tl}, \cdots, x_{Ntl})'$, for each $l \in \{1, \cdots, k\}$, where β_{il} and x_{itl} denote the l^{th} elements of $\boldsymbol{\beta}_i$ and $\boldsymbol{x}_{it}$, respectively, $i \in \{1, \cdots, N\}$ and $t \in \{1, \cdots, T\}$. As discussed previously, under stationarity and given $\boldsymbol{x}_{\circ t}$, we can obtain $\mathrm{E}(\boldsymbol{y}_{\circ t}) = (\boldsymbol{A} - \boldsymbol{C})^{-1}\boldsymbol{B}\boldsymbol{x}_{\circ t}$. Then the partial derivative matrix of $\mathrm{E}(\boldsymbol{y}_{\circ t})$ with respect to the values of l^{th} explanatory variable is

$$\frac{\partial \mathrm{E}(\boldsymbol{y}_{\circ t})}{\partial \boldsymbol{x}_{\circ tl}'} = (\boldsymbol{A} - \boldsymbol{C})^{-1}\boldsymbol{B}_l \equiv \boldsymbol{M}_l. \tag{17}$$

Again, note that the definitions of matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ under HSARDP with combined matrices are different from those under the high-order one.

The i-th diagonal element of $\boldsymbol{M}_l$, denoted by $m_{ii}^{(l)}$, estimates the direct effect of the l-th covariate on the i-th unit's outcome; the sum of all the elements of the i-th row, excluding $m_{ii}^{(l)}$, of $\boldsymbol{M}_l$ estimates the total marginal effects on unit i's outcome of the l-th covariate from all the other units; the sum of all the elements of the i-th column of $\boldsymbol{M}_l$, excluding $m_{ii}^{(l)}$, estimates the total marginal effects of unit i's l-th covariate on the outcomes of all the other units. Following LeSage et al. (2017) and LeSage and Chih (2018), we can interpret them as the direct impact (DI), indirect spill-in impact (SII), and indirect spill-out impact (SOI), respectively. Numerically, we can stack these measures of impact of unit i's, $i = 1, \cdots, N$, l-th covariate as $N \times 1$ vectors, namely, $\mathrm{dg}(\boldsymbol{M}_l)$, $\boldsymbol{M}_l\mathbf{1} - \mathrm{dg}(\boldsymbol{M}_l)$, and $(\mathbf{1}'\boldsymbol{M}_l)' - \mathrm{dg}(\boldsymbol{M}_l)$, respectively, where $\mathbf{1}$ is an $N \times 1$ vector of ones. We can calculate them from each simulated draw and use the posterior means as point estimates of them. Similarly as before, we signify their levels of significance using the highest posterior density credible intervals.

3. MONTE CARLO EXPERIMENTS

In this section, we present the results from a Monte Carlo study to investigate the small-sample properties of the Bayesian estimator under three different model specifications, named as Models A, B, and C, respectively. Model A refers to a standard HSARDP as specified by (1); Model B is an HSARDP with heterogeneous convex combinations of two weight matrices, see (2); Model C is a high-order HSARDP, see (16). In the experiments, we include a constant term and one exogenous regressor in all three model specifications.[8]

Two weight matrices are considered for Model B when they are combined as a single weight matrix and the same matrices are used for Model C (which is now a second-order HSARDP): $\boldsymbol{W}_1$ is a first-order contiguity matrix, where $w_{ij}^{(1)} = 1$ if $j = i-1$ or $j = i+1$, and zero otherwise; $\boldsymbol{W}_2$ is a second-order contiguity matrix where $w_{ij}^{(2)} = 1$ if $j = i-2$ or $j = i+2$, and zero otherwise. Model A uses $\boldsymbol{W}_1$ as its weight matrix. To make it comparable with Models A and B, we define pseudo parameters under Model C, namely, $\psi_i = \psi_i^{(1)} + \psi_i^{(2)}$ and $\phi_i = \phi_i^{(1)} + \phi_i^{(2)}$, and report the average biases and root mean squared errors (RMSEs) of the estimated ψ_i and ϕ_i across the three models. We set the number of replications to be $1,000$, and for each replication, we set the number of simulations for the MCMC sampler as $5,000$ with the first $2,000$ as the burn-in draws. We include results when $T = 25, 50, 100$ and $N = 25, 50, 100$ in this section and provide additional results when $N = 200$ in the Supplementary Appendix. The true parameter values, values of the exogenous regressor, and the error terms are redrawn in each replication of data simulation. We denote, for each region i, the constant as α_i and the coefficient of the single exogenous regressor as β_i. Moreover, we choose the values of prior hyperparameters such that the priors are relatively uninformative.[9] Table 1 lists parameter configurations we use across the three models.

Insert Table 1 here.

Since in practice we do not know the data generating process (DGP), we first discuss

the issue of model selection before presenting the bias and RMSE results. We use the observed-data Deviance Information Criterion (DIC, see Chan and Grant (2016)) for this purpose. In our model specifications, DIC is defined as DIC $= \overline{D} + p_D$, where $\overline{D} = -2\mathrm{E}_{\boldsymbol{y}_{\circ 0},\boldsymbol{\theta}}[\ln P(\boldsymbol{y}|\boldsymbol{y}_{\circ 0},\boldsymbol{\theta})|\boldsymbol{y}]$, $p_D = \overline{D} - \hat{D}$, and $\hat{D} = -2\ln P(\boldsymbol{y}|\hat{\boldsymbol{y}}_{\circ 0},\hat{\boldsymbol{\theta}})$. $\hat{\boldsymbol{y}}_{\circ 0}$ and $\hat{\boldsymbol{\theta}}$ denote the corresponding point estimates based on posterior means. $\overline{D}$ can be used to measure model fit, and p_D is called the effective number of parameters, which can be used to measure model complexity. Moreover, we approximate the expectation term in $\overline{D}$ using the average of $\ln P(\boldsymbol{y}|\boldsymbol{y}_{\circ 0},\boldsymbol{\theta})$ over all the posterior draws of $\boldsymbol{y}_{\circ 0}$ and $\boldsymbol{\theta}$.

Table 2 presents the model selection results for each DGP–N–T combination. When the DGP is model A, we see that Model A is strongly favored over Model C. Although DIC may not strongly favor Model A over Model B when T is small, it would perform significantly better as T goes up, and the true model is strongly favored when $T = 100$. When the DGP is generated from Model B or Model C, DIC strongly favors the true model under all N-T combinations. Therefore, our results indicate that in practice when one is uncertain about the choice of weight matrices, whether being a single one of single source, or a single one of multiple sources combined, or multiple ones of multiple sources, the DIC seems to be a reliable model selection criterion.

Insert Table 2 here.

Table 3 contains the average bias and RMSE results when the DGP is Model A. Under this specification, $\boldsymbol{\gamma}_i = \boldsymbol{\delta}_i = (1,0)'$ in Model B and $\psi_i^{(2)} = \phi_i^{(2)} = 0$ in Model C. This experimental design corresponds to the scenario when a researcher may be using more than one measure of connectivity when in fact there is only one. Model C has $2N(q-1) = 2N$ additional parameters relative to Model A, pertaining to the second-order spatial and spatial-temporal coefficients, but has the same number of parameters as Model B ($N(2q+1) + N(k+1) = 3N + 2N(q-1) + N(k+1)$). Thus we would expect that for a given N-T combination, the RMSEs associated with the estimated ψ_i and ϕ_i from Models

B and C are higher than those from Model A. This is evidenced by Table 3. We also see increased RMSEs in the estimated λ_i, β_i, and σ_i^2, but to a much less extent. Lastly, we observe that, for a fixed N, when T goes up, the biases and RMSEs across the three models all go down.

Insert Table 3 here.

Table 4 reports the bias and RMSE when the DGP is generated according to Model B. Immediately, we see that ignoring multiple sources of spatial correlation and using the baseline HSARDP leads to much higher average RMSEs, especially in terms of estimating ψ_i, ϕ_i, α_i, and σ_i^2 and when T is relatively large. Comparing results from Model B and those from Model C, we see that the former estimates ψ_i and ϕ_i with less bias and lower RMSE and also fares in general better in estimating other parameters.

Insert Table 4 here.

We now turn to the situation when Model C is the true one, as reported in Table 5. As before, by assuming only one source of spatial connectivity, Model A estimates all the parameters with much higher RMSEs and gives significant bias in estimating σ_i^2. Also, Model C performs typically better than Model B in estimating all the parameters, which is expected because Model C is more flexible than Model B, given that Model C would allow the spatial correlations characterized by various weight matrices to have different signs.

Insert Table 5 here.

4. AN EMPIRICAL STUDY

In this section, we investigate the spatial-temporal dynamics in real house price changes among different MSAs in the US. The data used are similar to those in Aquaro et al. (2021)

and Yang (2021), but we have updated their quarterly data and the extended sample period runs from 1975:Q2 to 2019:Q4. The Freddie Mac House Price Index (FMHPI), Consumer Price Index (CPI), population, and nominal per capita income for 377 MSAs are used to construct the growth rates. They are de-factored and de-seasonalized using the same method as in Aquaro et al. (2021) with the aim of filtering out the effects of seasonal trends and national factors so that the estimated parameters only reflect the local influences and spillovers. For comparison purpose, the same baseline model as in Aquaro et al. (2021) is used, namely, model (1), where an adjacency matrix is used as the spatial weight matrix such that two MSAs are considered as neighbors if the distance between their centers is within 75 miles. With such a geographic distance threshold, 39 MSAs in the sample don't have any neighbors. Excluding these 39 MSAs, we have a panel of $N = 338$ and $T = 179$. The outcome variable y_{it} is the quarterly growth rate in real house prices and the explanatory variables $\boldsymbol{x}_{it}$ include quarterly growth rates of population and real per capita disposable income. Accordingly, we put $\boldsymbol{\beta}_i = (\alpha_i, \beta_i^{pop}, \beta_i^{inc})'$, where α_i is for the constant term. This spatial weight matrix with a geographic distance threshold of 75 miles is denoted by $\boldsymbol{W}_{g1}$.

In addition to $\boldsymbol{W}_{g1}$, we consider two different candidate weight matrices, recognizing that spatial correlation may travel well beyond the 75-mile ring and it may also be of non-geographic nature. $\boldsymbol{W}_{g2}$ characterizes neighbors such that the corresponding entries are non-zero when their distances are greater than 75 miles but less than or equal to 150 miles. So this matrix represents a second distance ring of relatively more remote neighboring areas. $\boldsymbol{W}_n$ is constructed such that two MSAs are regarded as "neighbors" if the correlation coefficient of their estimated residuals under the baseline model is significantly positive. In other words, $\boldsymbol{W}_n$ would treat two MSAs as neighbors if there is some positive correlation in their local housing markets that is not captured by the geographic weight matrix $\boldsymbol{W}_{g1}$. The significance of correlation coefficients is based on the testing procedure in Yang (2021).

We use Models A, B, and C to designate again the baseline model (1), HSARDP with combined matrices (2), and higher-order HSARDP (16) (with $q = 2$), respectively. Under Models B and C, we can choose two different sets of weight matrices: $\{\boldsymbol{W}_{g1}, \boldsymbol{W}_{g2}\}$ and $\{\boldsymbol{W}_{g1}, \boldsymbol{W}_{n}\}$.[10] For example, under Model B, the estimated combination weights based on $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_{n}$ would represent the relative importance of geographic and non-geographic information in specifying the spatial weight matrix. It can be expected that the spatial correlations characterized by different candidate weight matrices may have different magnitudes and/or signs, and that there may be heterogeneity in the relative importance of different types of spatial correlation.

In estimation, we continue to use the relatively uninformative priors as in the previous section. We use $12,000$ iterations in the MCMC cycle with the first 2,000 discarded. Figure 1 presents the posterior means of estimated $\psi_i + \phi_i$ under the baseline model, where (and also in Figures 2 and 3) MSAs in deep gray have positive estimates and those in light gray have negative estimates.[11] If the same color scheme is used, this figure will be very similar to Figure 1.(a) in Aquaro et al. (2021), which represents their QML estimates for the net spatial parameters based on the same spatial weight matrix.

Insert Figure 1 here.

Table 6 summarizes the estimation results under the baseline model and Models B and C with two different sets of weight matrices. The baseline model shows that 153 of the 338 MSAs have significantly positive net spatial parameter estimates while 33 have significantly negative estimates, based on the 95% highest posterior density credible intervals. When it comes to the estimated coefficient on the temporal lag of outcome variable, all but one MSAs have positive $\hat{\lambda}_i$, and 335 of them are statistically significant, suggesting that for nearly all of the MSAs, real house price changes are positively correlated over time within each MSA. Moreover, 90 MSAs have significantly positive $\hat{\beta}_i^{pop}$ and 79 MSAs have significantly positive $\hat{\beta}_i^{inc}$. Similar to the results in Aquaro et al. (2021), the estimates $\hat{\beta}_i^{pop}$

in general have larger magnitudes than $\hat{\beta}_i^{inc}$ and tend to be more significant.

Insert Table 6 here.

When using $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_{g2}$ as the candidate weight matrices, Models B and C provide comparable results and they are similar to those from Model A. Although incorporating multiple geographic weight matrices based on different distance thresholds do affect the signs and/or magnitudes of net spatial parameter estimates for some MSAs, it seems that it does not change much the pattern of spatial correlation. The story is quite different, however, when we instead use $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$ that incorporate both geographic and non-geographic information. Now there are many more MSAs with significantly positive $\hat{\psi}_i + \hat{\phi}_i$ under Model B or C compared to the baseline model. This makes sense because when we incorporate different types of spatial correlation based on both geographic and non-geographic information, we may be able to capture more interactions among local housing markets. It is highly possible that there are some unobservable economic, social or migration factors leading to spillover effects among local housing markets even though these markets are not close to each other in the geographic sense. In particular, we find that there are 54 MSAs whose net spatial parameter estimates are insignificant under the baseline model but become significantly positive under Model C with $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$. When comparing Model B with $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$ against the baseline model, we find that 58 MSAs, including, for example, San Francisco-Oakland-Hayward, have significantly positive net spatial parameter estimates even though they are insignificant under Model A. Intuitively, consider Los Angeles-Long Beach-Anaheim that is more than 150 miles away from San Francisco-Oakland-Hayward. Given that they are of similar economic, social, and demographic characteristics, one would naturally expect that the housing market of Los Angeles-Long Beach-Anaheim has some impact on that of San Francisco-Oakland-Hayward, even though they are not geographic neighbors. If we only use the spatial weight matrices based on geographic distance, we fail to capture the spatial correlation between the two markets.

Figures 2 and 3 respectively present the estimates of net spatial parameters under Models B and C when we use $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$ as the candidate weight matrices. In comparison with Figure 1, which corresponds to results under the baseline model, we see that there is less heterogeneity across MSAs in the signs of $\hat{\psi}_i+\hat{\phi}_i$ once we incorporate the non-geographic information.

Insert Figures 2 and 3 here.

Based on the DIC values from Table 6, we see that the baseline model is the least favored and using both geographic and non-geographic information is strongly preferred to using only geographic information. Also, Model C is slightly favored over Model B. This is expected because as shown in the previous section, Model C tends to be more flexible than Model B.[12]

Table 7 reports the estimated marginal effects of income growth and population growth on house price growth. Notably, there are many more MSAs with significantly positive indirect spill-in impact estimates under Models B and C with $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$ compared to cases when only geographic information is used to capture spatial correlation, while the direct and spill-out impact estimates are quite stable across different specifications. Recall that in this setup a positive spill-in impact implies that increases in population and income in MSAs close-by and/or of similar characteristics can create upward pressure on an MSA's real house price. When we use only geographic distance to define neighbors, we would underestimate the number of neighbors that can exert positive pressure on an MSA's housing market.

Insert Table 7 here.

It would be interesting to investigate the degree of heterogeneity in the combination weights if we can settle down with using Model B. Figures 4 and 5 present the estimates of combination weights γ_i and δ_i under Model B with $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$ as the candidate weight

matrices.[13] In these two figures, for MSAs that are in deep gray, the γ_i (δ_i) estimates are greater than 0.5, which means that geographic distance is more relevant for specifying the spatial weight matrix for the spatial lag (spatial temporal lag). In contrast, for the MSAs that are in light gray, the non-geographic distance is more relevant. We can see from the two figures that there is strong evidence of heterogeneity in the estimated combination weights across different MSAs. Notably, it seems that higher importance of non-geographic distance is usually the case in economically highly active areas, such as New York-Newark-Jersey City, Boston-Cambridge-Newton, Chicago-Naperville-Elgin, Dallas-Fort Worth-Arlington, Seattle-Tacoma-Bellevue, San Francisco-Oakland-Hayward, and so on. Table 8 presents the group average estimates of combination weights by region under the same model specification. In particular, when specifying the spatial weight matrix for the spatial lag, non-geographic distance is relatively more important for MSAs in Mideast, New England, and Southwest, where there are a number of economically highly active and densely populated MSAs. In contrast, geographic distance is relatively more important for MSAs in the Rocky Mountain region.

Insert Figures 4 and 5 here.

Insert Table 8 here.

5. CONCLUDING REMARKS

In this paper, we introduce heterogeneity in the specification of spatial weight matrix into the heterogeneous spatial autoregressive dynamic panel models by incorporating heterogeneous convex combinations of different connectivity matrices. A high-order model is also proposed as an alternative approach to incorporating multiple types of spatial correlation characterized by different weight matrices. We propose a Bayesian MCMC estimation procedure and conduct Monte Carlo experiments to compare the finite-sample performance of these two models relative to the standard model that incorporates only a single source of

spatial correlation. It is found that the DIC performs very well in picking up the true model when one is uncertain about the choice of spatial weight matrices. When there are multiple sources of correlation in the true model, the baseline model can lead to substantially biased estimation results. We apply our extended models to study real house price changes among 338 MSAs in the US, where two geographic and one non-geographic connectivity matrices are considered. Based on the model selection criterion DIC, models taking into account both geographic and non-geographic correlation are strongly preferred and they find many more MSAs that have significant positive net spatial parameter estimates and more MSAs whose income and population growths have significant positive spill-in impact estimates.

For future research, it may be interesting to impose Bayesian shrinkage on the heterogeneous coefficients or the combination weights on different spatial weight matrices. This would allow for heterogeneous selection of the most relevant spatial weight matrices for different spatial units. Most Bayesian LASSO papers focus on imposing shrinkage on the coefficients of explanatory variables ($\boldsymbol{\beta}$), while only a small number of papers apply shrinkage to the spatial dependence parameters. Lam and Souza (2020)'s approach allows for selection of connectivity matrices in the homogeneous coefficient setting. Under the heterogeneous framework, a potential challenge is how to impose stationary condition on the spatial dependence parameters. We need to specify the shrinkage priors which can accommodate this condition. Most of the frequently used shrinkage priors are based on normal or beta distributions and there does not seem to be a straightforward way to take care of the stationarity condition.

ACKNOWLEDGEMENTS

The authors are grateful to two anonymous referees and the editor-in-chief (Paul Elhorst) for their helpful comments. The second author thanks Josh Chan for fruitful discussion. The authors are responsible for all remaining errors.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

NOTES

[1]Another prominent approach to modeling cross-sectional dependence is to assume a multifactor structure, where cross-sectional units are simultaneously affected by a limited number of common factors.

[2]See, *inter alia*, Fingleton (2001), Ertur and Koch (2007), and Parent and LeSage (2012) in economic growth; Baicker (2005) and Han and Lee (2016) in public economics; Lin (2010) and Lin and Weinberg (2014) in social networks; Liu et al. (2018) in real estate economics.

[3]For QML, see Lee (2004), Lee and Yu (2010), and Qu et al. (2017); for GMM, see Lee (2007), Lee and Yu (2014), and Jin and Lee (2019); for IV, see Kelejian and Piras (2014) and Qu et al. (2016); for II, see Kyriacou et al. (2017) and Bao et al. (2020); for Bayesian MCMC, see Han et al. (2017), Han and Lee (2016), and Parent and LeSage (2012).

[4]In our simulations in the third section, the highest (theoretical) degree of similarity, as defined in LeSage and Pace (2014), among the weight matrices is 0, and in our empirical study in the fourth section, it is 0.0533. Additional simulation results in the Supplementary Appendix suggest that model (2) may be subject to identification concerns when we have weight matrices that are highly similar. We thank a referee for bringing the issue of identification to our attention and the editor-in-chief for suggesting the additional simulations.

[5]Although each $\boldsymbol{\gamma}_i$ contains q elements, there are only $q-1$ parameters that need to be estimated given the constraint $\sum_{s=1}^{q}\gamma_{is}=1$. Similarly, there are $q-1$ parameters in each $\boldsymbol{\delta}_i$.

[6]The large-T asymptotic result may be restrictive for many empirical researchers when dealing with micro-level data. Intuitively, large-T asymptotics is needed since there are so many parameters arising from unit-level heterogeneity. We thank a referee for pointing this out. Simulation results in the third section and additional results in the Supplementary Appendix suggest that the performance of the Bayesian estimator is not that sensitive to N.

[7]When implementing the Gibbs sampler, we actually sample $\boldsymbol{\eta}_i$, $\boldsymbol{\gamma}_i$, and $\boldsymbol{\delta}_i$ together in a single block to increase computational efficiency. This would allow us to evaluate the stationarity condition only once instead of three times for each i within each simulation, although it may take larger number of simulations for the sampler to converge. Moreover, note that unlike the other sets of parameters, we don't need to update $\boldsymbol{y}_{\circ 0}$ for each $i \in \{1,2,...,N\}$ in each simulation because it doesn't vary across different units.

[8]When simulating data, the values of exogenous regressor for each spatial unit i and time t are independently sampled from a standard normal distribution.

[9]We set $\boldsymbol{\xi}_i=\mathbf{0}$, $v_i^2=100$, and $a_i=b_i=0$, for each $i\in\{1,\cdots,N\}$.

[10]It would be less intuitive if we choose a third set $\{\boldsymbol{W}_{g2},\boldsymbol{W}_n\}$ since $\boldsymbol{W}_n$ is based on the residuals from the baseline model using $\boldsymbol{W}_{g1}$. There are 11 MSAs that have no neighbors

as defined in the second geographic weight matrix $\boldsymbol{W}_{g2}$. For these 11 MSAs, we set the corresponding weights $\gamma_{i2} = \delta_{i2} = 0$ in Model B and $\psi_i^{(2)} = \phi_i^{(2)} = 0$ in Model C. Similarly, we did the same for the 15 MSAs that have no neighbors as defined in $\boldsymbol{W}_n$ when the second set $\{\boldsymbol{W}_{g1}, \boldsymbol{W}_n\}$ is used for Model B and Model C.

[11]Figures of the heterogeneous estimates in this section are all generated from the R code of Aquaro et al. (2021). The sum $\psi_i + \phi_i$ is of particular interest, as it represents the net spatial effect from both contemporaneous and temporal components. Furthermore, the category "No-Neigh" in the legends of figures consists of the 39 MSAs without any neighbor as defined in $\boldsymbol{W}_{g1}$.

[12]Note that when Model B is a restricted version of Model C, it implies that for each i, $\psi_i^{(s)}$, $s = 1, \cdots, q$, are of the same sign, and that for each i, $\phi_i^{(s)}$, $s = 1, \cdots, q$, also have the same sign. We may use the post-convergence posterior draws under Model C to see what proportion of the draws having $\psi_i^{(s)}$ and $\phi_i^{(s)}$ satisfying these constraints for all i. For our empirical data, regardless of which of the two different sets of weight matrices is used, we get a p-value of 0 (in the sense that 0% of the draws meet these constraints). This is also consistent with the DIC result, which favors Model C. As a referee points out, the DIC approach is more general and useful for comparing many different aspects of model specifications.

[13]Given $q = 2$ and the requirement that weights sum up to 1, γ_i and $1 - \gamma_i$ are the weights attached to $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$, respectively, for the i-th MSA's composite term $y_{it}^{\star}$ on the right-hand side of (2). Similarly, δ_i and $1 - \delta_i$ are those for $y_{it-1}^{\star}$.

REFERENCES

M. Aquaro, N. Bailey, and M. H. Pesaran. Quasi maximum likelihood estimation of spatial models with heterogeneous coefficients. CESifo Working Paper Series 5428, CESifo, 2015.

M. Aquaro, N. Bailey, and M. H. Pesaran. Estimation and inference for spatial models with heterogeneous coefficients: An application to US house prices. *Journal of Applied Econometrics*, 36(1):18–44, 2021. doi:10.1002/jae.2792.

C. Autant-Bernard and J. P. LeSage. A heterogeneous coefficient approach to the knowledge production function. *Spatial Economic Analysis*, 14(2):196–218, 2019. doi:10.1080/17421772.2019.1562201.

K. Baicker. The spillover effects of state spending. *Journal of Public Economics*, 89(2): 529–544, 2005. ISSN 0047-2727. doi:10.1016/j.jpubeco.2003.11.003.

Y. Bao, X. Liu, and L. Yang. Indirect inference estimation of spatial autoregressions. *Econometrics*, 8(3), 2020. doi:10.3390/econometrics8030034.

J. C. C. Chan and A. L. Grant. On the observed-data deviance information criterion for volatility modeling. *Journal of Financial Econometrics*, 14(4):772–802, 2016. doi:10.1093/jjfinec/nbw002.

N. Debarsy and J. P. LeSage. Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity matrices. *Journal of Business & Economic Statistics*, 40(2):547–558, 2022. doi:10.1080/07350015.2020.1840993.

J. Elhorst. Dynamic spatial panels: Models, methods, and inferences. *Journal of Geographical Systems*, 14(1):5–28, 2012. doi:10.1007/s10109-011-0158-4.

C. Ertur and W. Koch. Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics*, 22(6):1033–1062, 2007. doi:10.1002/jae.963.

B. Fingleton. Equilibrium and economic growth: Spatial econometric models and simulations. *Journal of Regional Science*, 41(1):117–147, 2001. doi:10.1111/0022-4146.00210.

A. Gupta and P. M. Robinson. Inference on higher-order spatial autoregressive models with increasingly many parameters. *Journal of Econometrics*, 186(1):19–31, 2015. doi:10.1016/j.jeconom.2014.12.008.

X. Han and L.-F. Lee. Bayesian analysis of spatial panel autoregressive models with time-varying endogenous spatial weight matrices, common factors, and random coefficients. *Journal of Business & Economic Statistics*, 34(4):642–660, 2016. doi:10.1080/07350015.2016.1167058.

X. Han, C.-S. Hsieh, and L.-F. Lee. Estimation and model selection of higher-order spatial autoregressive model: An efficient Bayesian approach. *Regional Science and Urban Economics*, 63:97–120, 2017. doi:10.1016/j.regsciurbeco.2016.12.003.

F. Jin and L.-F. Lee. GEL estimation and tests of spatial autoregressive models. *Journal of Econometrics*, 208(2):585–612, 2019. doi:10.1016/j.jeconom.2018.07.007.

H. H. Kelejian and G. Piras. Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics*, 46:140–149, 2014. ISSN 0166-0462. doi:10.1016/j.regsciurbeco.2014.03.001.

M. Kyriacou, P. C. B. Phillips, and F. Rossi. Indirect inference in spatial autoregression. *The Econometrics Journal*, 20(2):168–189, 2017. doi:10.1111/ectj.12084.

C. Lam and P. C. L. Souza. Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business & Economic Statistics*, 38(3):693–710, 2020. doi:10.1080/07350015.2019.1569526.

L.-F. Lee. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6):1899–1925, 2004. doi:10.1111/j.1468-0262.2004.00558.x.

L.-F. Lee. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137(2):489–514, 2007. doi:10.1016/j.jeconom.2005.10.004.

L.-F. Lee and X. Liu. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory*, 26(1):187–230, 2010. doi:10.1017/S0266466609090653.

L.-F. Lee and J. Yu. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154(2):165–185, 2010. doi:10.1016/j.jeconom.2009.08.001.

L.-F. Lee and J. Yu. Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics*, 180(2):174–197, 2014. doi:10.1016/j.jeconom.2014.03.003.

J. P. LeSage and Y.-Y. Chih. A Bayesian spatial panel model with heterogeneous coefficients. *Regional Science and Urban Economics*, 72:58–73, 2018. doi:10.1016/j.regsciurbeco.2017.02.007.

J. P. LeSage and R. K. Pace. The biggest myth in spatial econometrics. *Econometrics*, 2 (4):217–249, 2014. doi:10.3390/econometrics2040217.

J. P. LeSage, C. Vance, and Y.-Y. Chih. A Bayesian heterogeneous coefficients spatial autoregressive panel data model of retail fuel duopoly pricing. *Regional Science and Urban Economics*, 62:46–55, 2017. doi:10.1016/j.regsciurbeco.2016.11.003.

X. Lin. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4):825–860, 2010. doi:10.1086/653506.

X. Lin and B. A. Weinberg. Unrequited friendship? How reciprocity mediates adolescent peer effects. *Regional Science and Urban Economics*, 48:144–153, 2014. doi:10.1016/j.regsciurbeco.2014.06.001.

X. Liu, J. Chen, and S. Cheng. A penalized quasi-maximum likelihood method for variable selection in the spatial autoregressive model. *Spatial Statistics*, 25:86–104, 2018. ISSN 2211-6753. doi:10.1016/j.spasta.2018.05.001.

O. Parent and J. P. LeSage. Spatial dynamic panel data models with random effects. *Regional Science and Urban Economics*, 42(4):727–738, 2012. doi:10.1016/j.regsciurbeco.2012.04.008.

X. Qu, X. Wang, and L.-F. Lee. Instrumental variable estimation of a spatial dynamic panel model with endogenous spatial weights when T is small. *The Econometrics Journal*, 19 (3):261–290, 2016. doi:10.1111/ectj.12069.

X. Qu, L.-F. Lee, and J. Yu. QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. *Journal of Econometrics*, 197(2): 173–201, 2017. doi:10.1016/j.jeconom.2016.11.004.

G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009. doi:10.1198/jcgs.2009.06134.

C. F. Yang. Common factors and spatial dependence: An application to US house prices. *Econometric Reviews*, 40(1):14–50, 2021. doi:10.1080/07474938.2020.1741785.

Table 1: Parameter Configurations in Monte Carlo Experiments

	Model A	Model B	Model C
ψ_i	iidU$[-0.4, 0.4]$	iidU$[-0.4, 0.4]$	
$\psi_i^{(1)}$			iidU$[-0.4, 0.4]$
$\psi_i^{(2)}$			$\psi_i - \psi_i^{(1)}$, where $\psi_i \sim$ iidU$[-0.4, 0.4]$
ϕ_i	iidU$[-0.4, 0.4]$	iidU$[-0.4, 0.4]$	
$\phi_i^{(1)}$			iidU$[-0.4, 0.4]$
$\phi_i^{(2)}$			$\phi_i - \phi_i^{(1)}$, where $\phi_i \sim$ iidU$[-0.4, 0.4]$
λ_i	iidU$[\lvert\psi_i + \phi_i\rvert - 1, 1 - \lvert\psi_i + \phi_i\rvert]$	iidU$[\lvert\psi_i + \phi_i\rvert - 1, 1 - \lvert\psi_i + \phi_i\rvert]$	iidU$[\lvert\psi_i + \phi_i\rvert - 1, 1 - \lvert\psi_i + \phi_i\rvert]$
α_i	iidU$[0, 1]$	iidU$[0, 1]$	iidU$[0, 1]$
β_i	iidU$[0, 1]$	iidU$[0, 1]$	iidU$[0, 1]$
σ_i^2	iidU$(0, 1]$	iidU$(0, 1]$	iidU$(0, 1]$
γ_i		iidU$[0.1, 0.8]$	
δ_i		iidU$[0.1, 0.8]$	

Note: iidU$[-0.4, 0.4]$, for example, means iid uniform on the interval $[-0.4, 0.4]$.

Table 2: DIC results under all DGP-N-T combinations

	$N = 25$			$N = 50$			$N = 100$		
DGP: Model A	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
% A $\succ$ B	34.2	71.2	92.9	26.8	77.4	96.9	20.9	85.3	98.5
% A $\succ$ C	96.4	99.8	99.6	99.5	100	99.9	100	100	99.8
DGP: Model B	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
% B $\succ$ A	99.9	99.9	99.7	100	100	100	100	100	100
% B $\succ$ C	100	99.4	95.8	100	99.9	97.2	100	100	99.0
DGP: Model C	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
% C $\succ$ A	100	100	100	100	100	100	100	100	100
% C $\succ$ B	92.0	99.4	99.7	97.0	99.6	99.9	99.4	100	100

Note: % A $\succ$ B, for instance, denotes the percentage (out of $1,000$ simulations) when DIC favors Model A over Model B.

Table 3: Average bias and RMSE for region-specific parameters when DGP is Model A

	Model A			Model B			Model C		
$N = 25$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0024	0.0001	0.0001	0.0069	0.0053	0.0022	0.0061	0.0062	0.0022
ϕ_i	0.0003	-0.0005	0.0008	0.0001	-0.0012	0.0004	-0.0016	-0.0024	-0.0001
λ_i	-0.0242	-0.0116	-0.0067	-0.0214	-0.0101	-0.0060	-0.0237	-0.0115	-0.0069
α_i	0.0240	0.0144	0.0105	0.0196	0.0112	0.0092	0.0220	0.0124	0.0103
β_i	-0.0015	0.0006	0.0002	-0.0036	-0.0009	-0.0007	0.0021	0.0030	0.0015
σ_i^2	0.0667	0.0313	0.0167	0.0549	0.0283	0.0174	0.0828	0.0425	0.0226
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2294	0.1775	0.1316	0.2557	0.2127	0.1737	0.2741	0.2328	0.1872
ϕ_i	0.2010	0.1545	0.1134	0.2402	0.1973	0.1541	0.2456	0.2042	0.1599
λ_i	0.1618	0.1077	0.0750	0.1626	0.1087	0.0773	0.1704	0.1110	0.0768
α_i	0.2864	0.2116	0.1548	0.3070	0.2394	0.1862	0.3375	0.2531	0.1931
β_i	0.1619	0.1075	0.0739	0.1619	0.1080	0.0744	0.1707	0.1110	0.0752
σ_i^2	0.2224	0.1372	0.0930	0.2127	0.1351	0.0936	0.2424	0.1489	0.0983
$N = 50$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0003	-0.0005	-0.0008	0.0061	0.0031	0.0016	0.0053	0.0033	0.0024
ϕ_i	0.0020	0.0000	0.0003	-0.0002	-0.0016	0.0001	-0.0011	-0.0026	-0.0003
λ_i	-0.0225	-0.0113	-0.0061	-0.0202	-0.0104	-0.0055	-0.0221	-0.0112	-0.0062
α_i	0.0247	0.0158	0.0109	0.0207	0.0137	0.0104	0.0201	0.0137	0.0102
β_i	-0.0022	0.0004	-0.0002	-0.0047	-0.0010	-0.0010	0.0007	0.0029	0.0010
σ_i^2	0.0647	0.0330	0.0156	0.0538	0.0300	0.0169	0.0792	0.0430	0.0215
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2287	0.1795	0.1316	0.2561	0.2161	0.1739	0.2714	0.2354	0.1858
ϕ_i	0.2005	0.1548	0.1137	0.2415	0.1987	0.1564	0.2462	0.2055	0.1606
λ_i	0.1619	0.1085	0.0747	0.1624	0.1100	0.0772	0.1691	0.1119	0.0765
α_i	0.2880	0.2130	0.1556	0.3061	0.2398	0.1915	0.3349	0.2568	0.1964
β_i	0.1614	0.1077	0.0737	0.1614	0.1080	0.0742	0.1694	0.1111	0.0750
σ_i^2	0.2190	0.1397	0.0943	0.2102	0.1367	0.0980	0.2385	0.1511	0.0990
$N = 100$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	-0.0009	0.0001	-0.0002	0.0039	0.0040	0.0027	0.0033	0.0045	0.0029
ϕ_i	0.0017	0.0000	-0.0002	0.0007	-0.0007	0.0000	-0.0005	-0.0012	-0.0005
λ_i	-0.0230	-0.0118	-0.0057	-0.0209	-0.0108	-0.0052	-0.0227	-0.0117	-0.0057
α_i	0.0256	0.0155	0.0100	0.0213	0.0131	0.0088	0.0219	0.0126	0.0085
β_i	-0.0020	-0.0011	0.0000	-0.0043	-0.0025	-0.0008	0.0009	0.0008	0.0012
σ_i^2	0.0650	0.0312	0.0155	0.0543	0.0282	0.0171	0.0784	0.0407	0.0212
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2293	0.1788	0.1302	0.2562	0.2149	0.1728	0.2718	0.2341	0.1841
ϕ_i	0.2019	0.1552	0.1137	0.2417	0.1985	0.1567	0.2472	0.2053	0.1598
λ_i	0.1602	0.1087	0.0746	0.1612	0.1102	0.0772	0.1680	0.1119	0.0765
α_i	0.2873	0.2125	0.1561	0.3063	0.2400	0.1898	0.3368	0.2556	0.1957
β_i	0.1619	0.1077	0.0736	0.1617	0.1079	0.0740	0.1699	0.1110	0.0748
σ_i^2	0.2196	0.1374	0.0916	0.2106	0.1354	0.1002	0.2378	0.1493	0.1024

Notes: True parameter values are generated by sampling from $\psi_i \sim \text{iidU}[-0.4, 0.4]$, $\phi_i \sim \text{iidU}[-0.4, 0.4]$, $\lambda_i \sim \text{iidU}[|\psi_i + \phi_i| - 1, 1 - |\psi_i + \phi_i|]$, $\alpha_i \sim \text{iidU}[0, 1]$, $\beta_i \sim \text{iidU}[0, 1]$, $\sigma_i^2 \sim \text{iidU}(0, 1]$. Average bias and RMSE are computed as $N^{-1} \sum_{i=1}^{N} R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)$ and $N^{-1} \sum_{i=1}^{N} \sqrt{R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)^2}$, respectively, where $R = 1,000$ is the number of simulations.

Table 4: Average bias and RMSE for region-specific parameters when DGP is Model B

	Model A			Model B			Model C		
$N = 25$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0033	0.0007	0.0010	0.0064	0.0034	0.0002	0.0071	0.0058	0.0019
ϕ_i	-0.0015	-0.0004	-0.0011	-0.0004	0.0000	0.0003	-0.0025	-0.0013	0.0006
λ_i	-0.0232	-0.0098	-0.0048	-0.0220	-0.0110	-0.0062	-0.0234	-0.0115	-0.0062
α_i	0.0261	0.0141	0.0096	0.0215	0.0123	0.0102	0.0220	0.0114	0.0090
β_i	-0.0011	0.0000	0.0000	-0.0032	-0.0013	-0.0006	0.0018	0.0019	0.0013
σ_i^2	0.0908	0.0558	0.0421	0.0453	0.0207	0.0131	0.0815	0.0404	0.0222
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2816	0.2407	0.2023	0.2429	0.1979	0.1551	0.2735	0.2345	0.1886
ϕ_i	0.2504	0.2133	0.1821	0.2321	0.1872	0.1438	0.2461	0.2061	0.1595
λ_i	0.1704	0.1180	0.0874	0.1622	0.1078	0.0750	0.1710	0.1121	0.0764
α_i	0.3357	0.2736	0.2329	0.2996	0.2319	0.1768	0.3379	0.2556	0.1935
β_i	0.1662	0.1099	0.0755	0.1602	0.1072	0.0739	0.1715	0.1107	0.0750
σ_i^2	0.2378	0.1515	0.1091	0.2057	0.1295	0.0897	0.2402	0.1462	0.0960
$N = 50$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0010	0.0004	0.0006	0.0049	0.0025	0.0017	0.0051	0.0042	0.0030
ϕ_i	0.0016	0.0002	0.0003	0.0001	-0.0008	0.0005	-0.0009	-0.0024	0.0000
λ_i	-0.0208	-0.0099	-0.0041	-0.0201	-0.0109	-0.0057	-0.0215	-0.0116	-0.0058
α_i	0.0222	0.0123	0.0074	0.0202	0.0118	0.0081	0.0200	0.0119	0.0084
β_i	-0.0013	0.0004	-0.0002	-0.0041	-0.0012	-0.0009	0.0008	0.0020	0.0008
σ_i^2	0.0881	0.0565	0.0388	0.0443	0.0227	0.0119	0.0780	0.0413	0.0205
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2794	0.2399	0.2015	0.2445	0.2002	0.1556	0.2730	0.2342	0.1863
ϕ_i	0.2511	0.2145	0.1847	0.2331	0.1884	0.1450	0.2471	0.2077	0.1618
λ_i	0.1672	0.1174	0.0864	0.1601	0.1085	0.0745	0.1693	0.1124	0.0759
α_i	0.3292	0.2726	0.2321	0.2983	0.2330	0.1808	0.3339	0.2553	0.1959
β_i	0.1645	0.1100	0.0755	0.1596	0.1071	0.0737	0.1695	0.1109	0.0749
σ_i^2	0.2331	0.1533	0.1057	0.2042	0.1318	0.1018	0.2363	0.1486	0.1022
$N = 100$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0006	0.0005	0.0001	0.0031	0.0029	0.0019	0.0040	0.0043	0.0032
ϕ_i	-0.0005	0.0000	-0.0001	0.0001	-0.0003	0.0001	-0.0012	-0.0010	-0.0008
λ_i	-0.0217	-0.0102	-0.0042	-0.0207	-0.0109	-0.0057	-0.0214	-0.0112	-0.0058
α_i	0.0233	0.0140	0.0085	0.0204	0.0131	0.0092	0.0210	0.0123	0.0092
β_i	-0.0019	-0.0009	0.0000	-0.0044	-0.0023	-0.0007	0.0004	0.0008	0.0009
σ_i^2	0.0864	0.0548	0.0384	0.0443	0.0220	0.0123	0.0769	0.0398	0.0203
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2809	0.2386	0.2006	0.2454	0.2003	0.1548	0.2729	0.2346	0.1848
ϕ_i	0.2512	0.2138	0.1831	0.2336	0.1877	0.1452	0.2471	0.2062	0.1612
λ_i	0.1672	0.1172	0.0870	0.1601	0.1085	0.0750	0.1687	0.1121	0.0761
α_i	0.3293	0.2713	0.2302	0.2975	0.2305	0.1814	0.3325	0.2529	0.1966
β_i	0.1650	0.1097	0.0752	0.1602	0.1070	0.0736	0.1700	0.1106	0.0748
σ_i^2	0.2314	0.1514	0.1044	0.2038	0.1329	0.1030	0.2347	0.1461	0.1040

Notes: True parameter values are generated by sampling from $\psi_i \sim \text{iidU}[-0.4, 0.4]$, $\phi_i \sim \text{iidU}[-0.4, 0.4]$, $\lambda_i \sim \text{iidU}[|\psi_i + \phi_i| - 1, 1 - |\psi_i + \phi_i|]$, $\alpha_i \sim \text{iidU}[0, 1]$, $\beta_i \sim \text{iidU}[0, 1]$, $\sigma_i^2 \sim \text{iidU}(0, 1]$, $\gamma_i \sim \text{iidU}[0.1, 0.8]$, $\delta_i \sim \text{iidU}[0.1, 0.8]$. Average bias and RMSE are computed as $N^{-1} \sum_{i=1}^{N} R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)$ and $N^{-1} \sum_{i=1}^{N} \sqrt{R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)^2}$, respectively, where $R = 1,000$ is the number of simulations.

Table 5: Average bias and RMSE for region-specific parameters when DGP is Model C

	Model A			Model B			Model C		
$N = 25$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0008	-0.0019	-0.0001	0.0098	0.0095	0.0080	0.0020	0.0028	0.0009
ϕ_i	0.0017	0.0012	0.0025	-0.0034	-0.0028	-0.0003	0.0001	0.0006	0.0015
λ_i	-0.0168	-0.0048	0.0014	-0.0176	-0.0081	-0.0030	-0.0199	-0.0101	-0.0057
α_i	0.0195	0.0131	0.0053	0.0177	0.0108	0.0053	0.0214	0.0119	0.0071
β_i	-0.0004	0.0004	0.0007	-0.0057	-0.0038	-0.0038	0.0026	0.0031	0.0021
σ_i^2	0.2281	0.1901	0.1761	0.1184	0.0812	0.0731	0.0878	0.0458	0.0264
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.3903	0.3883	0.3812	0.3173	0.3150	0.3062	0.2621	0.2229	0.1767
ϕ_i	0.3538	0.3463	0.3386	0.2866	0.2742	0.2637	0.2392	0.1969	0.1543
λ_i	0.2078	0.1751	0.1585	0.1757	0.1288	0.1047	0.1658	0.1107	0.0791
α_i	0.4455	0.4192	0.4098	0.3523	0.3036	0.3023	0.3363	0.2526	0.1970
β_i	0.1847	0.1244	0.0879	0.1724	0.1150	0.0816	0.1707	0.1112	0.0757
σ_i^2	0.4015	0.3210	0.2883	0.2719	0.1924	0.3206	0.2566	0.1604	0.1392
$N = 50$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0011	0.0004	0.0013	0.0088	0.0079	0.0075	0.0019	0.0010	0.0008
ϕ_i	0.0017	-0.0002	0.0003	-0.0034	-0.0021	0.0015	-0.0008	-0.0015	0.0006
λ_i	-0.0173	-0.0059	-0.0007	-0.0163	-0.0086	-0.0043	-0.0190	-0.0097	-0.0051
α_i	0.0235	0.0142	0.0081	0.0206	0.0126	0.0080	0.0206	0.0126	0.0078
β_i	-0.0005	0.0001	0.0003	-0.0060	-0.0045	-0.0040	0.0010	0.0020	0.0010
σ_i^2	0.2110	0.1785	0.1611	0.1127	0.0800	0.0641	0.0816	0.0442	0.0243
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.3859	0.3850	0.3748	0.3142	0.3094	0.3009	0.2603	0.2220	0.1754
ϕ_i	0.3517	0.3445	0.3361	0.2854	0.2736	0.2638	0.2386	0.1970	0.1552
λ_i	0.2040	0.1717	0.1547	0.1747	0.1288	0.1023	0.1642	0.1110	0.0774
α_i	0.4415	0.4162	0.4050	0.3494	0.2977	0.2785	0.3337	0.2530	0.2017
β_i	0.1822	0.1234	0.0870	0.1714	0.1143	0.0798	0.1707	0.1112	0.0755
σ_i^2	0.3738	0.2999	0.2654	0.2652	0.2354	0.1637	0.2457	0.1731	0.1347
$N = 100$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	-0.0006	-0.0011	-0.0012	0.0081	0.0089	0.0078	0.0015	0.0017	0.0006
ϕ_i	0.0019	0.0012	0.0004	-0.0020	-0.0020	-0.0017	0.0006	0.0001	-0.0005
λ_i	-0.0178	-0.0059	-0.0005	-0.0175	-0.0083	-0.0044	-0.0193	-0.0102	-0.0054
α_i	0.0226	0.0124	0.0083	0.0192	0.0119	0.0085	0.0197	0.0128	0.0095
β_i	-0.0016	-0.0006	0.0004	-0.0065	-0.0051	-0.0036	0.0003	0.0006	0.0011
σ_i^2	0.2093	0.1739	0.1584	0.1135	0.0767	0.0644	0.0827	0.0425	0.0239
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.3870	0.3815	0.3721	0.3139	0.3069	0.2992	0.2619	0.2225	0.1754
ϕ_i	0.3526	0.3450	0.3349	0.2861	0.2739	0.2619	0.2394	0.1981	0.1552
λ_i	0.2025	0.1695	0.1525	0.1748	0.1281	0.1016	0.1646	0.1111	0.0769
α_i	0.4378	0.4121	0.4034	0.3449	0.2969	0.2770	0.3308	0.2524	0.1951
β_i	0.1817	0.1232	0.0867	0.1712	0.1148	0.0800	0.1708	0.1116	0.0756
σ_i^2	0.3720	0.2930	0.2613	0.2631	0.1789	0.1557	0.2470	0.1568	0.1141

Notes: True parameter values are generated by sampling from $\psi_i \sim \text{iidU}[-0.4, 0.4]$, $\phi_i \sim \text{iidU}[-0.4, 0.4]$, $\lambda_i \sim \text{iidU}[|\psi_i+\phi_i|-1, 1-|\psi_i+\phi_i|]$, $\alpha_i \sim \text{iidU}[0, 1]$, $\beta_i \sim \text{iidU}[0, 1]$, $\sigma_i^2 \sim \text{iidU}(0, 1]$, $\psi_i^{(1)} \sim \text{iidU}[-0.4, 0.4]$, $\psi_i^{(2)} = \psi_i - \psi_i^{(1)}$, $\phi_i^{(1)} \sim \text{iidU}[-0.4, 0.4]$, and $\phi_i^{(2)} = \phi_i - \phi_i^{(1)}$. Average bias and RMSE are computed as $N^{-1} \sum_{i=1}^{N} R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)$ and $N^{-1} \sum_{i=1}^{N} \sqrt{R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)^2}$, respectively, where $R = 1,000$ is the number of simulations.

Table 6: Comparisons among specifications

Number of MSAs with	Model A	Model B $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_{g2}$	Model C $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_{g2}$	Model B $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_{n}$	Model C $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_{n}$
positive $\hat{\psi}_i + \hat{\phi}_i$	253	264	263	301	304
significantly positive $\hat{\psi}_i + \hat{\phi}_i$	153	151	140	186	172
negative $\hat{\psi}_i + \hat{\phi}_i$	85	74	75	37	34
significantly negative $\hat{\psi}_i + \hat{\phi}_i$	33	20	17	9	9
positive $\hat{\lambda}_i$	337	338	338	337	338
significantly positive $\hat{\lambda}_i$	335	336	336	337	337
negative $\hat{\lambda}_i$	1	0	0	1	0
significantly negative $\hat{\lambda}_i$	0	0	0	0	0
positive $\hat{\beta}_i^{pop}$	260	266	253	264	261
significantly positive $\hat{\beta}_i^{pop}$	90	102	96	107	101
negative $\hat{\beta}_i^{pop}$	78	72	85	74	77
significantly negative $\hat{\beta}_i^{pop}$	9	8	9	7	10
positive $\hat{\beta}_i^{inc}$	248	249	243	244	250
significantly positive $\hat{\beta}_i^{inc}$	79	73	68	78	78
negative $\hat{\beta}_i^{inc}$	90	89	95	94	88
significantly negative $\hat{\beta}_i^{inc}$	2	1	4	1	2
DIC	78070	74453	70076	57830	54283

Note: Statistical significance is based on the 95% highest posterior density Bayesian credible intervals.

Table 7: Comparisons among specifications: Marginal effects

Number of MSAs with	Model A	Model B $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_{g2}$	Model C $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_{g2}$	Model B $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_{n}$	Model C $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_{n}$
Population growth					
Direct impact:					
positive DI	258	265	253	265	261
significantly positive DI	86	94	93	104	91
negative DI	80	73	85	73	77
significantly negative DI	4	7	6	4	6
Indirect spillin impact:					
positive SII	236	246	241	291	296
significantly positive SII	65	77	71	123	109
negative SII	102	92	97	47	42
significantly negative SII	10	8	9	4	4
Indirect spillout impact:					
positive SOI	222	232	218	240	232
significantly positive SOI	30	27	22	45	38
negative SOI	116	106	120	98	106
significantly negative SOI	5	2	2	0	2
Income growth					
Direct impact:					
positive DI	247	250	243	245	251
significantly positive DI	68	71	60	74	70
negative DI	91	88	95	93	87
significantly negative DI	2	1	3	1	2
Indirect spillin impact:					
positive SII	247	247	238	282	287
significantly positive SII	45	58	51	102	95
negative SII	91	91	100	56	51
significantly negative SII	10	5	4	0	3
Indirect spillout impact:					
positive SOI	215	218	217	220	227
significantly positive SOI	23	17	16	32	32
negative SOI	123	120	121	118	111
significantly negative SOI	5	2	1	2	4

Note: Statistical significance is based on the 95% highest posterior density Bayesian credible intervals.

Table 8: Mean estimates of γ_i and δ_i by region under Model B with $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$

Region	Number of MSAs	Mean estimates of γ_i	Mean estimates of δ_i
Far West	43	0.5105	0.5426
		(0.0127)	(0.0135)
Great Lakes	54	0.5052	0.5072
		(0.0146)	(0.0146)
Mideast	40	0.3536	0.4586
		(0.0152)	(0.0136)
New England	13	0.2736	0.2978
		(0.0409)	(0.0381)
Plains	24	0.4701	0.6030
		(0.0213)	(0.0233)
Rocky Mountain	11	0.5397	0.5652
		(0.0287)	(0.0283)
Southeast	114	0.5120	0.5440
		(0.0034)	(0.0088)
Southwest	24	0.2989	0.4224
		(0.0314)	(0.0264)

Notes: Posterior standard deviations in parentheses. MSAs without any neighbors defined in the non-geographic weight matrix $\boldsymbol{W}_n$ have been excluded before the mean estimates of combination weights are computed.

Figure 1: Estimates of $\psi_i + \phi_i$ by MSA under Model A

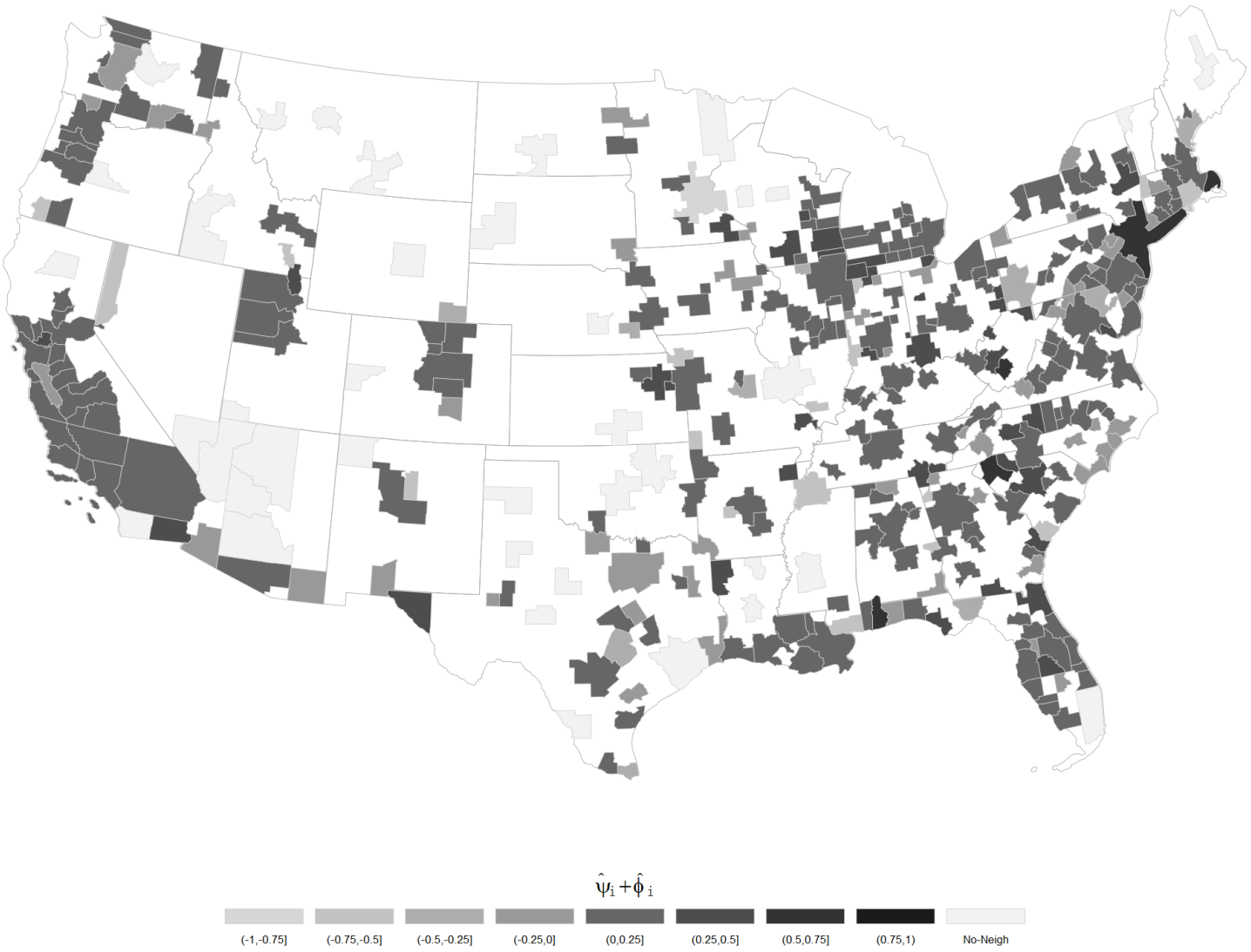


Figure 2: Estimates of $\psi_i + \phi_i$ by MSA under Model B with $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$

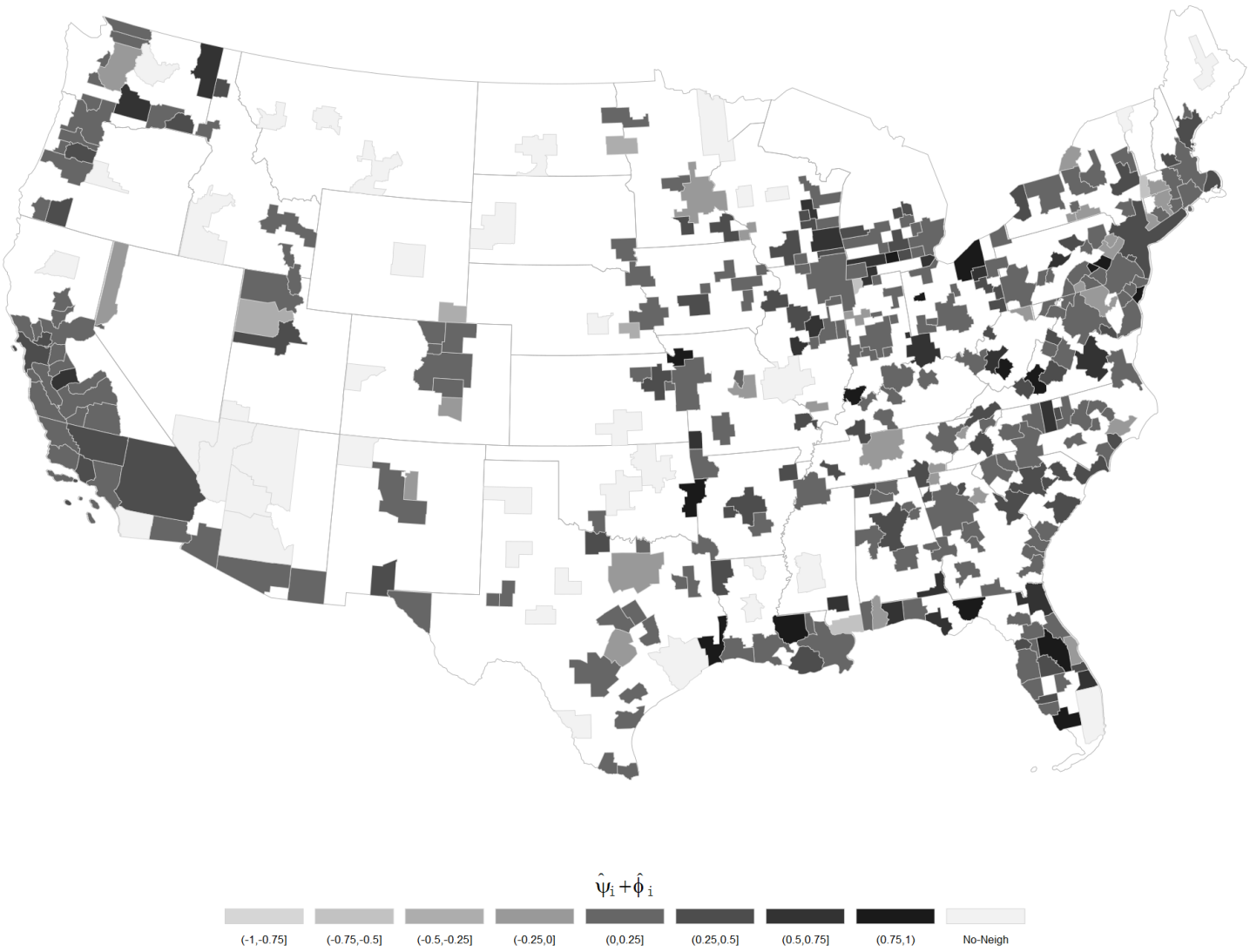


Figure 3: Estimates of $\psi_i + \phi_i$ by MSA under Model C with $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$

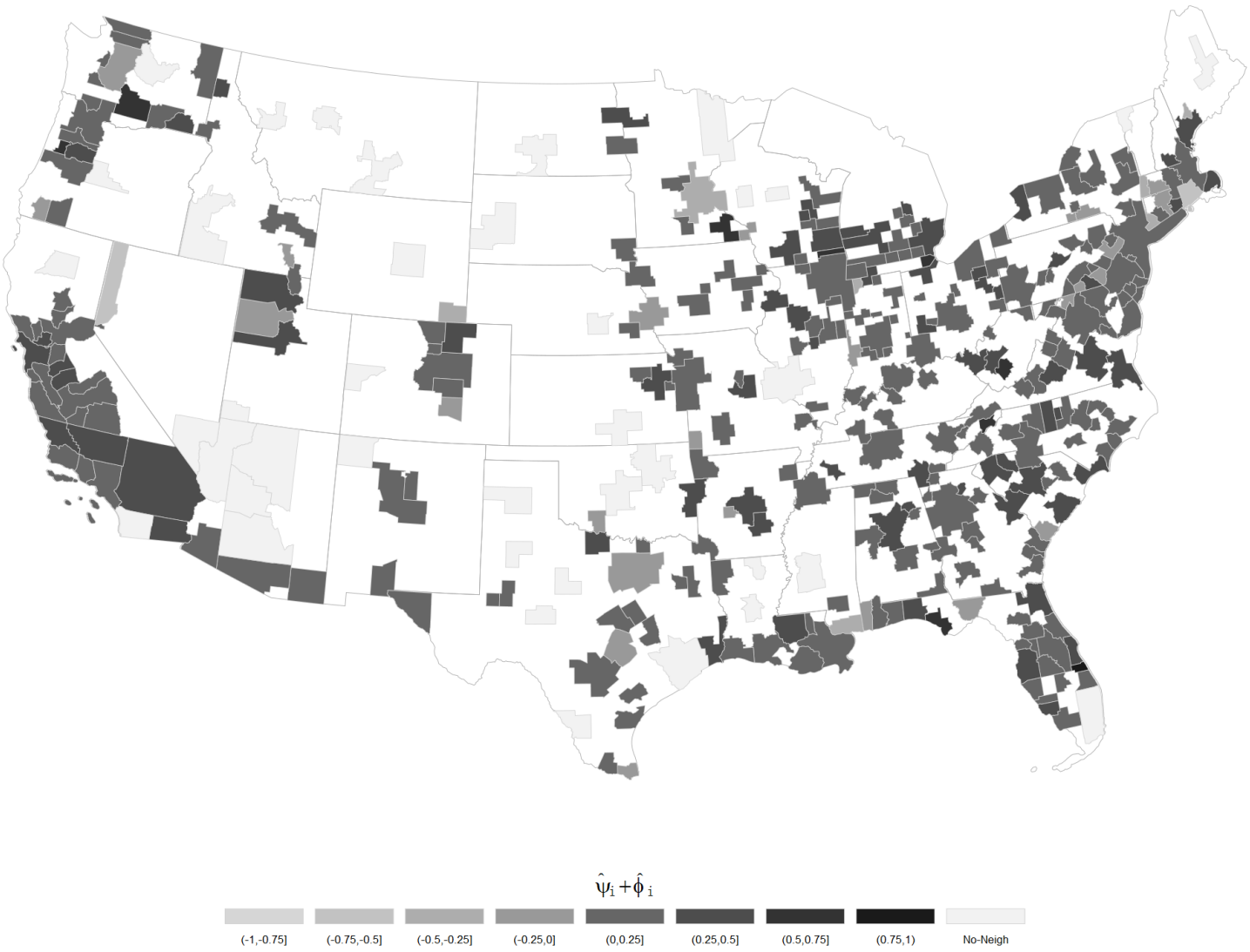


Figure 4: Estimates of γ_i by MSA under Model B with $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$

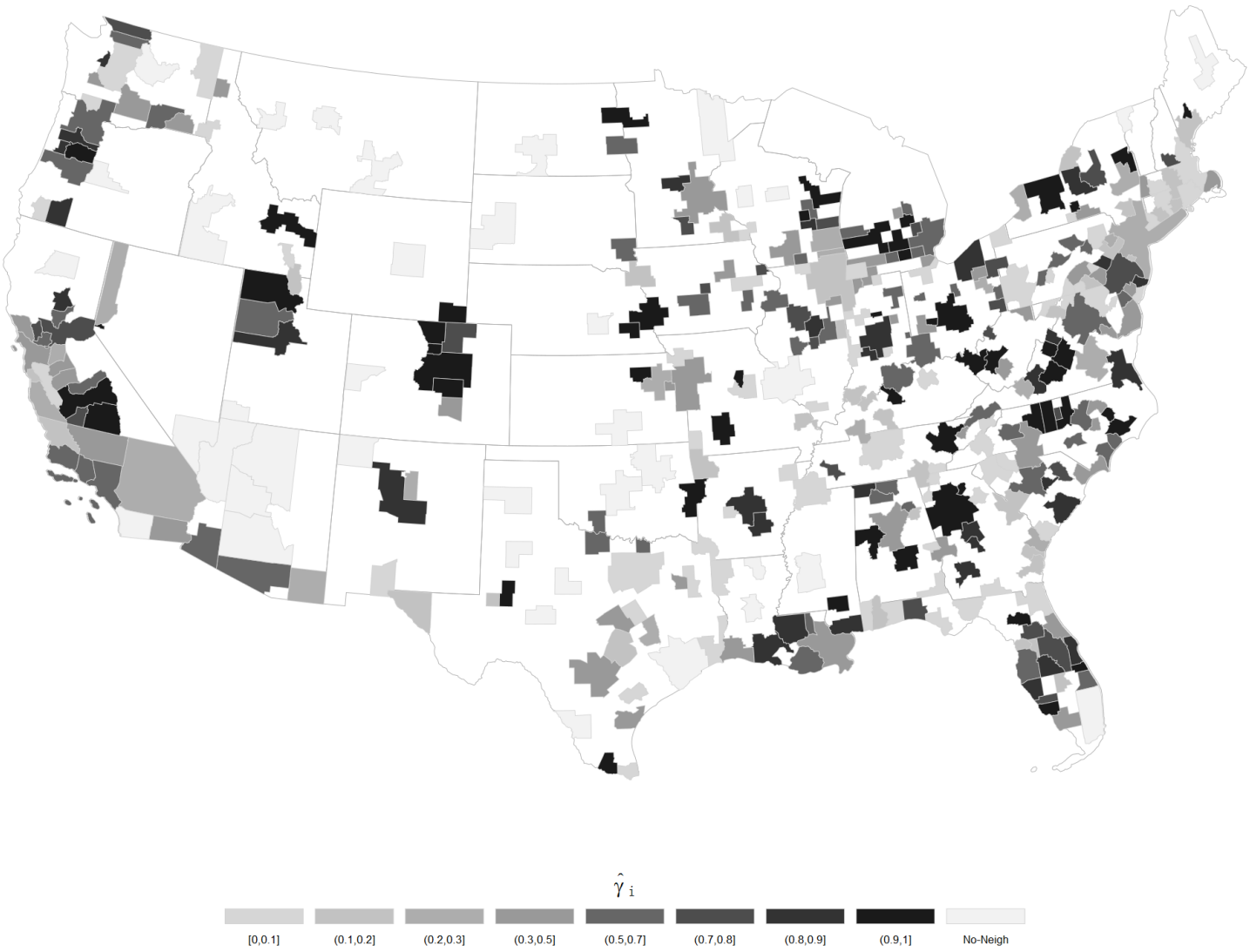
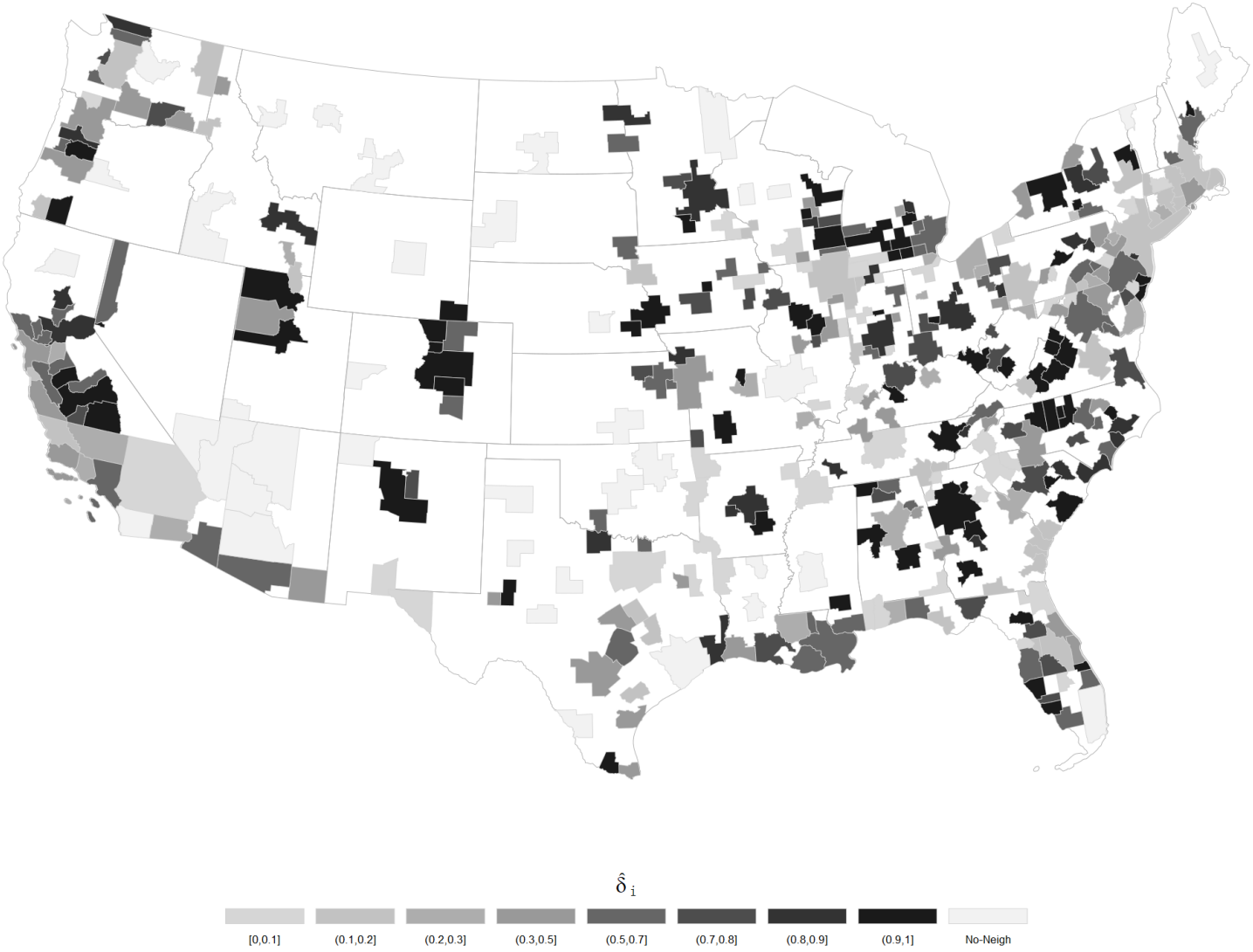


Figure 5: Estimates of δ_i by MSA under Model B with $\boldsymbol{W}_{g1}$ and $\boldsymbol{W}_n$



Heterogeneous Spatial Dynamic Panels with an Application to US Housing Data

A: Additional Simulation Results under Different Weight Matrices

In this section, we further evaluate the finite-sample performance of the proposed procedure when the DGP is the combined model (2), namely, Model B. We consider situations where the candidate weight matrices are relatively highly correlated or when both sparse and dense weight matrices are present. For this purpose, we specify a binary contiguity matrix $\boldsymbol{W}_3$ which characterizes both first-order and second-order neighbors, as defined respectively by $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ in the main text. We also specify distance-based weight matrices. Let the distance between each pair of units (i, j), denoted by d_{ij}, be randomly sampled from independent uniform distributions on the interval $(0, 100]$. One can think of d_{ij} as the distance between two cities in miles. Then we construct $\boldsymbol{W}_4$ and $\boldsymbol{W}_5$ as the 4-nearest-neighbor and 5-nearest-neighbor weight matrices, respectively. We further construct an exponential distance decay matrix $\boldsymbol{W}_e$ such that its (i, j)-th element is given by $\exp(-d_{ij})$. We row-normalize all the weight matrices. As in the main text, we consider $T = 25, 50, 100$ and $N = 25, 50, 100$. The experimental design is the same as discussed in the main text except that we are no longer using $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ as the candidate weight matrices.

We first consider $\{\boldsymbol{W}_1, \boldsymbol{W}_3\}$ as the set of candidate matrices, which we label as Case 1. The highest correlation measure between $\boldsymbol{W}_1$ and $\boldsymbol{W}_3$ (again as defined in LeSage and Pace (2014)) across all the N-T combinations is 0.7158. In Case 2, we consider $\boldsymbol{W}_4$ and $\boldsymbol{W}_5$ as the candidate weight matrices. The highest correlation measure between these two nearest-neighbor matrices is 0.8944, much higher than that in Case 1. We adopt the sparse binary contiguity matrix $\boldsymbol{W}_1$ and the dense exponential distance decay matrix $\boldsymbol{W}_e$ in Case 3. The highest correlation measure under this design is 0.0741, indicating that the degree of similarity between $\boldsymbol{W}_1$ and $\boldsymbol{W}_e$ is fairly low.

Table A1 presents the model selection results under the three cases. We can see that the performance of DIC is negatively impacted in Case 2 when two highly similar weight

matrices $\boldsymbol{W}_4$ and $\boldsymbol{W}_5$ are adopted, compared to the results presented in Table 2 in the main text. This suggests that Model B may be subject to identification issue when we have two weight matrices that are very highly correlated.

Tables A2 to A4 present the average bias and RMSE results, corresponding to Cases 1 to 3, respectively. We can see from Table A2 that when the DGP is Model B but the candidate weight matrices have a relatively high similarity, the RMSEs from Model B are only slightly lower than those when Model A and Model C are fitted to the data. (Recall from Table 4 in the main text the outstanding performance of the estimation procedure under Model B when $\{\boldsymbol{W}_1, \boldsymbol{W}_2\}$ is adopted.) However, according to Table A3, when the two weight matrices are very highly correlated, the RMSEs from the true model are basically as high as (if not higher than) those from estimates under Models A and C. These results raise concerns regarding the identification of Model B when two candidate weight matrices have a very high degree of similarity. Finally, We can see from Table A4 that the average bias and RMSE results are comparable to the previous results when $\{\boldsymbol{W}_1, \boldsymbol{W}_2\}$ is used as the set of weight matrices (see Table 4 in the main text). It seems that including a dense weight matrix wouldn't affect the identification of Model B as long as there is no issue of high degree of similarity.

Table A1: DIC results under all N-T combinations when DGP is Model B

	$N = 25$			$N = 50$			$N = 100$		
Case 1: $\boldsymbol{W}_1$, $\boldsymbol{W}_3$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
% B $\succ$ A	97.1	97.6	97.1	99.7	99.6	98.1	100	100	99.2
% B $\succ$ C	96.3	98.9	96.8	99.4	99.7	97.5	100	99.9	99.0
Case 2: $\boldsymbol{W}_4$, $\boldsymbol{W}_5$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
% B $\succ$ A	87.6	89.4	88.1	95.6	95.0	88.6	98.6	96.9	92.2
% B $\succ$ C	72.2	90.6	92.8	78.5	96.9	93.4	89.0	98.3	96.9
Case 3: $\boldsymbol{W}_1$, $\boldsymbol{W}_e$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
% B $\succ$ A	100	100	99.9	100	100	100	100	100	100
% B $\succ$ C	99.9	99.4	95.7	100	99.8	95.6	100	99.9	98.6

Note: % B $\succ$ A, for instance, denotes the percentage (out of 1,000 simulations) that DIC favors Model B over Model A.

Table A2: Average bias and RMSE for region-specific parameters when DGP is Model B, Case 1: $\boldsymbol{W}_1$, $\boldsymbol{W}_3$

	Model A			Model B			Model C		
$N = 25$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0029	0.0004	0.0004	0.0075	0.0048	0.0012	0.0067	0.0055	0.0020
ϕ_i	-0.0007	-0.0005	-0.0006	-0.0013	-0.0013	-0.0003	-0.0019	-0.0008	0.0004
λ_i	-0.0236	-0.0108	-0.0061	-0.0225	-0.0108	-0.0064	-0.0230	-0.0111	-0.0063
α_i	0.0262	0.0149	0.0100	0.0220	0.0124	0.0099	0.0227	0.0116	0.0094
β_i	-0.0014	0.0000	0.0001	-0.0032	-0.0012	-0.0005	-0.0013	0.0008	0.0010
σ_i^2	0.0716	0.0367	0.0230	0.0518	0.0241	0.0143	0.0535	0.0299	0.0201
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2488	0.2002	0.1546	0.2429	0.1974	0.1540	0.2438	0.2118	0.1765
ϕ_i	0.2188	0.1739	0.1349	0.2273	0.1818	0.1388	0.2261	0.1916	0.1535
λ_i	0.1643	0.1099	0.0766	0.1620	0.1078	0.0750	0.1658	0.1106	0.0762
α_i	0.3060	0.2330	0.1798	0.2942	0.2266	0.1729	0.3129	0.2426	0.1907
β_i	0.1634	0.1081	0.0743	0.1603	0.1072	0.0738	0.1646	0.1091	0.0748
σ_i^2	0.2244	0.1386	0.0935	0.2098	0.1313	0.0929	0.2142	0.1368	0.0982
$N = 50$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0007	0.0005	0.0005	0.0054	0.0036	0.0024	0.0050	0.0040	0.0034
ϕ_i	0.0017	0.0002	0.0002	-0.0002	-0.0007	0.0002	0.0001	-0.0018	-0.0002
λ_i	-0.0219	-0.0112	-0.0056	-0.0206	-0.0111	-0.0060	-0.0216	-0.0115	-0.0060
α_i	0.0240	0.0132	0.0086	0.0211	0.0118	0.0085	0.0212	0.0113	0.0080
β_i	-0.0020	0.0002	-0.0003	-0.0037	-0.0010	-0.0009	-0.0022	0.0008	0.0005
σ_i^2	0.0700	0.0382	0.0209	0.0512	0.0257	0.0129	0.0523	0.0312	0.0178
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2482	0.2004	0.1543	0.2436	0.1994	0.1537	0.2439	0.2124	0.1754
ϕ_i	0.2191	0.1749	0.1373	0.2277	0.1833	0.1412	0.2270	0.1940	0.1552
λ_i	0.1623	0.1103	0.0761	0.1601	0.1086	0.0747	0.1643	0.1110	0.0756
α_i	0.2995	0.2323	0.1797	0.2924	0.2271	0.1772	0.3095	0.2437	0.1885
β_i	0.1622	0.1082	0.0742	0.1595	0.1070	0.0737	0.1635	0.1093	0.0746
σ_i^2	0.2212	0.1407	0.0926	0.2081	0.1351	0.0977	0.2124	0.1391	0.0946
$N = 100$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0007	0.0004	0.0003	0.0042	0.0036	0.0024	0.0041	0.0042	0.0031
ϕ_i	-0.0003	0.0000	-0.0002	-0.0006	-0.0002	-0.0001	-0.0007	-0.0008	-0.0007
λ_i	-0.0222	-0.0112	-0.0054	-0.0209	-0.0110	-0.0058	-0.0213	-0.0112	-0.0058
α_i	0.0235	0.0151	0.0093	0.0205	0.0132	0.0094	0.0205	0.0125	0.0089
β_i	-0.0024	-0.0010	0.0000	-0.0042	-0.0020	-0.0005	-0.0026	-0.0002	0.0007
σ_i^2	0.0694	0.0372	0.0210	0.0508	0.0249	0.0132	0.0522	0.0305	0.0176
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2498	0.1991	0.1543	0.2450	0.1981	0.1531	0.2450	0.2132	0.1733
ϕ_i	0.2200	0.1746	0.1367	0.2285	0.1830	0.1407	0.2274	0.1928	0.1546
λ_i	0.1625	0.1098	0.0764	0.1600	0.1081	0.0748	0.1643	0.1106	0.0757
α_i	0.3010	0.2315	0.1801	0.2921	0.2250	0.1774	0.3087	0.2423	0.1893
β_i	0.1624	0.1081	0.0740	0.1597	0.1070	0.0736	0.1637	0.1092	0.0746
σ_i^2	0.2200	0.1396	0.0948	0.2072	0.1334	0.1016	0.2114	0.1406	0.0959

Notes: True parameter values are generated by sampling from $\psi_i \sim \text{iidU}[-0.4, 0.4]$, $\phi_i \sim \text{iidU}[-0.4, 0.4]$, $\lambda_i \sim \text{iidU}[|\psi_i + \phi_i| - 1, 1 - |\psi_i + \phi_i|]$, $\alpha_i \sim \text{iidU}[0, 1]$, $\beta_i \sim \text{iidU}[0, 1]$, $\sigma_i^2 \sim \text{iidU}(0, 1]$, $\gamma_i \sim \text{iidU}[0.1, 0.8]$, $\delta_i \sim \text{iidU}[0.1, 0.8]$. Average bias and RMSE are computed as $N^{-1} \sum_{i=1}^{N} R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)$ and $N^{-1} \sum_{i=1}^{N} \sqrt{R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)^2}$, respectively, where $R = 1,000$ is the number of simulations.

Table A3: Average bias and RMSE for region-specific parameters when DGP is Model B, Case 2: $\boldsymbol{W}_4$, $\boldsymbol{W}_5$

	Model A			Model B			Model C		
$N = 25$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0044	0.0019	0.0001	0.0045	0.0026	-0.0001	0.0041	0.0024	-0.0003
ϕ_i	0.0007	-0.0001	0.0007	0.0003	0.0002	0.0004	0.0005	0.0001	0.0009
λ_i	-0.0204	-0.0100	-0.0059	-0.0199	-0.0101	-0.0061	-0.0208	-0.0101	-0.0060
α_i	0.0212	0.0123	0.0079	0.0209	0.0125	0.0098	0.0215	0.0124	0.0085
β_i	-0.0042	-0.0021	-0.0011	-0.0042	-0.0021	-0.0012	-0.0042	-0.0020	-0.0009
σ_i^2	0.0499	0.0236	0.0141	0.0463	0.0210	0.0126	0.0390	0.0172	0.0113
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2499	0.2029	0.1535	0.2534	0.2082	0.1618	0.2380	0.2000	0.1596
ϕ_i	0.2423	0.1946	0.1483	0.2459	0.2009	0.1564	0.2322	0.1939	0.1542
λ_i	0.1610	0.1074	0.0744	0.1608	0.1074	0.0747	0.1616	0.1081	0.0750
α_i	0.2998	0.2338	0.1815	0.2972	0.2331	0.1886	0.3006	0.2351	0.1874
β_i	0.1589	0.1066	0.0735	0.1583	0.1064	0.0734	0.1586	0.1069	0.0738
σ_i^2	0.2050	0.1288	0.0874	0.2032	0.1280	0.0891	0.2010	0.1271	0.0872
$N = 50$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0001	0.0012	-0.0003	0.0004	0.0014	0.0004	0.0003	0.0013	0.0003
ϕ_i	0.0010	-0.0013	0.0000	0.0013	-0.0010	-0.0003	0.0013	-0.0010	-0.0005
λ_i	-0.0201	-0.0106	-0.0055	-0.0195	-0.0106	-0.0058	-0.0206	-0.0108	-0.0056
α_i	0.0232	0.0134	0.0076	0.0225	0.0138	0.0095	0.0227	0.0135	0.0083
β_i	-0.0049	-0.0020	-0.0014	-0.0050	-0.0021	-0.0014	-0.0051	-0.0019	-0.0012
σ_i^2	0.0498	0.0258	0.0129	0.0465	0.0233	0.0119	0.0390	0.0197	0.0102
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2498	0.2017	0.1530	0.2532	0.2087	0.1614	0.2375	0.2007	0.1595
ϕ_i	0.2422	0.1945	0.1477	0.2463	0.2014	0.1560	0.2319	0.1936	0.1536
λ_i	0.1593	0.1084	0.0741	0.1591	0.1082	0.0744	0.1602	0.1090	0.0744
α_i	0.2978	0.2349	0.1806	0.2967	0.2361	0.1896	0.2977	0.2378	0.1882
β_i	0.1580	0.1066	0.0733	0.1576	0.1064	0.0733	0.1580	0.1070	0.0738
σ_i^2	0.2041	0.1315	0.0875	0.2023	0.1311	0.1025	0.2002	0.1316	0.0887
$N = 100$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0001	0.0007	0.0003	0.0001	0.0009	0.0003	-0.0002	0.0008	0.0004
ϕ_i	0.0025	0.0002	0.0002	0.0028	0.0004	0.0008	0.0025	0.0007	0.0003
λ_i	-0.0200	-0.0106	-0.0054	-0.0195	-0.0105	-0.0056	-0.0203	-0.0106	-0.0055
α_i	0.0214	0.0133	0.0081	0.0208	0.0133	0.0095	0.0215	0.0129	0.0087
β_i	-0.0051	-0.0027	-0.0012	-0.0051	-0.0028	-0.0011	-0.0053	-0.0027	-0.0009
σ_i^2	0.0501	0.0247	0.0130	0.0466	0.0222	0.0117	0.0391	0.0186	0.0102
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2499	0.2026	0.1533	0.2536	0.2090	0.1612	0.2381	0.2009	0.1596
ϕ_i	0.2425	0.1950	0.1474	0.2467	0.2015	0.1556	0.2326	0.1943	0.1534
λ_i	0.1587	0.1079	0.0743	0.1586	0.1077	0.0743	0.1597	0.1085	0.0745
α_i	0.2966	0.2344	0.1797	0.2956	0.2355	0.1865	0.2977	0.2365	0.1861
β_i	0.1588	0.1067	0.0734	0.1583	0.1065	0.0733	0.1588	0.1069	0.0736
σ_i^2	0.2036	0.1302	0.0874	0.2020	0.1313	0.0977	0.1997	0.1290	0.0889

Notes: True parameter values are generated by sampling from $\psi_i \sim \text{iidU}[-0.4, 0.4]$, $\phi_i \sim \text{iidU}[-0.4, 0.4]$, $\lambda_i \sim \text{iidU}[|\psi_i + \phi_i| - 1, 1 - |\psi_i + \phi_i|]$, $\alpha_i \sim \text{iidU}[0, 1]$, $\beta_i \sim \text{iidU}[0, 1]$, $\sigma_i^2 \sim \text{iidU}(0, 1]$, $\gamma_i \sim \text{iidU}[0.1, 0.8]$, $\delta_i \sim \text{iidU}[0.1, 0.8]$. Average bias and RMSE are computed as $N^{-1} \sum_{i=1}^{N} R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)$ and $N^{-1} \sum_{i=1}^{N} \sqrt{R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)^2}$, respectively, where $R = 1,000$ is the number of simulations.

Table A4: Average bias and RMSE for region-specific parameters when DGP is Model B, Case 3: $\boldsymbol{W}_1$, $\boldsymbol{W}_e$

	Model A			Model B			Model C		
$N = 25$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0034	-0.0001	0.0000	0.0040	0.0015	-0.0009	0.0036	0.0012	-0.0009
ϕ_i	-0.0021	-0.0008	-0.0008	0.0006	0.0006	0.0010	0.0002	0.0019	0.0008
λ_i	-0.0229	-0.0096	-0.0041	-0.0223	-0.0109	-0.0063	-0.0237	-0.0113	-0.0066
α_i	0.0257	0.0133	0.0103	0.0227	0.0122	0.0104	0.0245	0.0121	0.0102
β_i	-0.0009	0.0002	0.0000	-0.0034	-0.0010	-0.0003	0.0020	0.0035	0.0015
σ_i^2	0.0975	0.0660	0.0486	0.0468	0.0233	0.0140	0.0861	0.0502	0.0246
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2768	0.2347	0.1920	0.2383	0.1926	0.1498	0.2692	0.2349	0.1821
ϕ_i	0.2485	0.2110	0.1765	0.2230	0.1765	0.1363	0.2416	0.1989	0.1536
λ_i	0.1735	0.1243	0.0923	0.1626	0.1083	0.0750	0.1724	0.1134	0.0769
α_i	0.3404	0.2807	0.2376	0.2994	0.2297	0.1755	0.3390	0.2586	0.1956
β_i	0.1673	0.1106	0.0762	0.1608	0.1075	0.0741	0.1720	0.1121	0.0756
σ_i^2	0.2441	0.1624	0.1161	0.2072	0.1330	0.0906	0.2461	0.1595	0.1018
$N = 50$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0006	0.0002	0.0004	0.0017	0.0003	0.0007	0.0009	0.0009	0.0011
ϕ_i	0.0018	0.0000	0.0009	0.0015	-0.0001	0.0000	0.0016	-0.0008	0.0006
λ_i	-0.0215	-0.0098	-0.0042	-0.0206	-0.0111	-0.0058	-0.0218	-0.0114	-0.0059
α_i	0.0240	0.0123	0.0072	0.0221	0.0135	0.0095	0.0234	0.0128	0.0088
β_i	-0.0015	0.0003	-0.0004	-0.0042	-0.0012	-0.0009	0.0010	0.0016	0.0007
σ_i^2	0.0913	0.0602	0.0407	0.0456	0.0234	0.0125	0.0815	0.0398	0.0210
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2769	0.2325	0.1912	0.2441	0.1963	0.1523	0.2726	0.2294	0.1832
ϕ_i	0.2489	0.2084	0.1756	0.2301	0.1839	0.1412	0.2462	0.2032	0.1575
λ_i	0.1694	0.1203	0.0900	0.1609	0.1084	0.0751	0.1701	0.1119	0.0762
α_i	0.3334	0.2771	0.2353	0.2960	0.2321	0.1842	0.3333	0.2555	0.1959
β_i	0.1649	0.1103	0.0755	0.1599	0.1073	0.0738	0.1697	0.1110	0.0751
σ_i^2	0.2367	0.1577	0.1107	0.2047	0.1342	0.0973	0.2400	0.1485	0.0986
$N = 100$									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0004	0.0001	-0.0004	0.0012	0.0002	-0.0001	0.0012	0.0003	-0.0001
ϕ_i	-0.0005	-0.0004	-0.0004	0.0015	0.0006	0.0006	0.0015	0.0011	0.0005
λ_i	-0.0217	-0.0105	-0.0043	-0.0207	-0.0108	-0.0058	-0.0214	-0.0110	-0.0058
α_i	0.0237	0.0144	0.0091	0.0205	0.0144	0.0097	0.0199	0.0137	0.0099
β_i	-0.0017	-0.0007	0.0000	-0.0045	-0.0022	-0.0007	-0.0009	0.0002	0.0006
σ_i^2	0.0823	0.0515	0.0344	0.0432	0.0214	0.0115	0.0686	0.0363	0.0185
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2773	0.2320	0.1922	0.2510	0.2062	0.1613	0.2714	0.2363	0.1881
ϕ_i	0.2489	0.2107	0.1783	0.2401	0.1953	0.1517	0.2525	0.2137	0.1675
λ_i	0.1668	0.1161	0.0854	0.1597	0.1083	0.0749	0.1675	0.1116	0.0757
α_i	0.3293	0.2701	0.2273	0.3000	0.2363	0.1850	0.3310	0.2584	0.1999
β_i	0.1645	0.1096	0.0749	0.1596	0.1071	0.0736	0.1681	0.1103	0.0747
σ_i^2	0.2288	0.1494	0.1027	0.2025	0.1324	0.0905	0.2263	0.1459	0.0971

Notes: True parameter values are generated by sampling from $\psi_i \sim \text{iidU}[-0.4, 0.4]$, $\phi_i \sim \text{iidU}[-0.4, 0.4]$, $\lambda_i \sim \text{iidU}[|\psi_i + \phi_i| - 1, 1 - |\psi_i + \phi_i|]$, $\alpha_i \sim \text{iidU}[0, 1]$, $\beta_i \sim \text{iidU}[0, 1]$, $\sigma_i^2 \sim \text{iidU}(0, 1]$, $\gamma_i \sim \text{iidU}[0.1, 0.8]$, $\delta_i \sim \text{iidU}[0.1, 0.8]$. Average bias and RMSE are computed as $N^{-1} \sum_{i=1}^{N} R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)$ and $N^{-1} \sum_{i=1}^{N} \sqrt{R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)^2}$, respectively, where $R = 1,000$ is the number of simulations.

B: Additional Simulation Results When $N = 200$

The experimental design and parameter configurations are the same as discussed in the main text. Table A5 below presents the model selection results for each DGP-T combination under $N = 200$ and Table A6 reports the average bias and RMSE results.

Table A5: DIC results when $N = 200$

	$T = 25$	$T = 50$	$T = 100$
DGP: Model A			
% A $\succ$ B	13.6	91.5	99.8
% A $\succ$ C	100	100	100
DGP: Model B			
% B $\succ$ A	100	100	100
% B $\succ$ C	100	100	100
DGP: Model C			
% C $\succ$ A	100	100	100
% C $\succ$ B	99.9	100	100

Note: % A $\succ$ B, for instance, denotes the percentage (out of $1,000$ simulations) when DIC favors Model A over Model B.

Table A6: Average bias and RMSE for region-specific parameters when $N = 200$

	Model A			Model B			Model C		
DGP: Model A									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	-0.0002	-0.0003	-0.0004	0.0045	0.0034	0.0021	0.0042	0.0040	0.0027
ϕ_i	0.0009	0.0006	0.0001	-0.0001	-0.0003	-0.0001	-0.0015	-0.0013	-0.0005
λ_i	-0.0215	-0.0114	-0.0064	-0.0192	-0.0104	-0.0059	-0.0213	-0.0113	-0.0064
α_i	0.0248	0.0161	0.0101	0.0204	0.0136	0.0096	0.0220	0.0131	0.0092
β_i	-0.0030	-0.0005	-0.0007	-0.0054	-0.0019	-0.0015	-0.0002	0.0016	0.0005
σ_i^2	0.0641	0.0310	0.0156	0.0536	0.0282	0.0169	0.0772	0.0402	0.0212
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2290	0.1782	0.1308	0.2559	0.2144	0.1725	0.2713	0.2325	0.1834
ϕ_i	0.2011	0.1553	0.1136	0.2414	0.1993	0.1562	0.2466	0.2064	0.1598
λ_i	0.1607	0.1087	0.0746	0.1613	0.1104	0.0769	0.1683	0.1124	0.0764
α_i	0.2859	0.2117	0.1558	0.3034	0.2394	0.1878	0.3328	0.2545	0.1951
β_i	0.1621	0.1074	0.0737	0.1620	0.1077	0.0742	0.1701	0.1107	0.0750
σ_i^2	0.2195	0.1373	0.0926	0.2107	0.1360	0.0972	0.2374	0.1476	0.1036
DGP: Model B									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0003	0.0003	0.0001	0.0039	0.0026	0.0015	0.0043	0.0046	0.0029
ϕ_i	0.0017	0.0008	0.0003	0.0013	0.0005	0.0003	0.0003	-0.0006	-0.0002
λ_i	-0.0219	-0.0105	-0.0047	-0.0208	-0.0113	-0.0060	-0.0218	-0.0116	-0.0062
α_i	0.0240	0.0148	0.0082	0.0210	0.0137	0.0089	0.0214	0.0132	0.0083
β_i	-0.0032	-0.0006	-0.0006	-0.0056	-0.0020	-0.0012	-0.0008	0.0011	0.0004
σ_i^2	0.0851	0.0540	0.0382	0.0437	0.0217	0.0122	0.0755	0.0392	0.0205
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.2795	0.2394	0.1999	0.2455	0.2006	0.1553	0.2727	0.2349	0.1849
ϕ_i	0.2519	0.2139	0.1834	0.2345	0.1892	0.1453	0.2479	0.2073	0.1609
λ_i	0.1675	0.1168	0.0858	0.1605	0.1083	0.0749	0.1693	0.1118	0.0762
α_i	0.3308	0.2723	0.2303	0.2984	0.2328	0.1794	0.3336	0.2548	0.1949
β_i	0.1648	0.1096	0.0749	0.1600	0.1068	0.0733	0.1697	0.1103	0.0744
σ_i^2	0.2302	0.1504	0.1039	0.2033	0.1335	0.0933	0.2336	0.1456	0.1011
DGP: Model C									
Average Bias	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.0010	0.0009	0.0013	0.0087	0.0089	0.0078	0.0024	0.0017	0.0009
ϕ_i	0.0009	0.0001	-0.0001	-0.0025	-0.0021	-0.0008	-0.0004	-0.0002	0.0005
λ_i	-0.0180	-0.0069	-0.0013	-0.0170	-0.0085	-0.0044	-0.0194	-0.0104	-0.0056
α_i	0.0224	0.0125	0.0063	0.0193	0.0118	0.0076	0.0211	0.0132	0.0086
β_i	-0.0017	0.0010	0.0004	-0.0069	-0.0037	-0.0040	-0.0003	0.0018	0.0007
σ_i^2	0.2051	0.1729	0.1567	0.1109	0.0769	0.0637	0.0801	0.0427	0.0232
Average RMSE	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
ψ_i	0.3844	0.3804	0.3708	0.3115	0.3056	0.2986	0.2609	0.2207	0.1749
ϕ_i	0.3535	0.3450	0.3361	0.2859	0.2724	0.2617	0.2395	0.1977	0.1550
λ_i	0.2018	0.1678	0.1506	0.1744	0.1271	0.1007	0.1647	0.1108	0.0769
α_i	0.4369	0.4123	0.4003	0.3447	0.2970	0.2758	0.3308	0.2513	0.1930
β_i	0.1820	0.1228	0.0861	0.1717	0.1145	0.0794	0.1709	0.1114	0.0754
σ_i^2	0.3670	0.2927	0.2615	0.2608	0.1775	0.1567	0.2449	0.1542	0.1111

Notes: True parameter values are generated by sampling from $\psi_i \sim$ iidU$[-0.4, 0.4]$, $\phi_i \sim$ iidU$[-0.4, 0.4]$, $\lambda_i \sim$ iidU$[|\psi_i + \phi_i| - 1, 1 - |\psi_i + \phi_i|]$, $\alpha_i \sim$ iidU$[0, 1]$, $\beta_i \sim$ iidU$[0, 1]$, $\sigma_i^2 \sim$ iidU$(0, 1]$. When DGP is Model B, $\gamma_i \sim$ iidU$[0.1, 0.8]$, $\delta_i \sim$ iidU$[0.1, 0.8]$. When DGP is Model C, $\psi_i^{(1)} \sim$ iidU$[-0.4, 0.4]$, $\psi_i^{(2)} = \psi_i - \psi_i^{(1)}$, $\phi_i^{(1)} \sim$ iidU$[-0.4, 0.4]$, and $\phi_i^{(2)} = \phi_i - \phi_i^{(1)}$. Average bias and RMSE are computed as $N^{-1} \sum_{i=1}^{N} R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)$ and $N^{-1} \sum_{i=1}^{N} \sqrt{R^{-1} \sum_{r=1}^{R} (\hat{\psi}_i - \psi_i)^2}$, respectively, where $R = 1,000$ is the number of simulations.

REFERENCES

LeSage, J. P., & Pace, R. K. (2014). The biggest myth in spatial econometrics. *Econometrics*, *2(4)*, 217–249. https://doi.org/10.3390/econometrics2040217