

Estimation and inference for spatial models with heterogeneous coefficients: An application to US house prices

Michele Aquaro¹ | Natalia Bailey² | M. Hashem Pesaran^{3,4}

¹Joint Research Centre, European Commission, Ispra, Italy

²Department of Econometrics and Business Statistics, Monash University, Melbourne, Victoria, Australia

³Department of Economics, University of Southern California, Los Angeles, California

⁴Trinity College, University of Cambridge, Cambridge, UK

Correspondence

Natalia Bailey, Department of Econometrics and Business Statistics, Monash University, Clayton, Melbourne, Victoria 3800, Australia.

Email: natalia.bailey@monash.edu

Summary

This paper considers the estimation and inference of spatial panel data models with heterogeneous spatial lag coefficients, with and without weakly exogenous regressors, and subject to heteroskedastic errors. A quasi maximum likelihood (QML) estimation procedure is developed and the conditions for identification of the spatial coefficients are derived. The QML estimators of individual spatial coefficients, as well as their mean group estimators, are shown to be consistent and asymptotically normal. Small-sample properties of the proposed estimators are investigated by Monte Carlo simulations and results are shown to be in line with the paper's key theoretical findings, even for panels with moderate time dimensions and irrespective of the number of cross-section units. A detailed empirical application to US house price changes during the 1975–2014 period shows a significant degree of heterogeneity in spatiotemporal dynamics over the 338 Metropolitan Statistical Areas considered.

1 | INTRODUCTION

This paper considers a heterogeneous version of the standard spatial autoregressive (SAR) panel data model whereby the spatial lag coefficients are allowed to differ over the cross-section units, in addition to the fixed effects generally allowed for in the literature. We refer to this generalized specification as the heterogeneous SAR (or HSAR) model. The model also features weakly exogenous regressors, such as lagged values of the dependent variable and heteroskedastic error variances, and provides a reasonably general framework for the analysis of heterogeneous interactions, where it is important to distinguish between the average intensity of spillover effects as characterized by standard spatial models, and the heterogeneity of such effects over different geographical units such as counties, regions or countries. Importantly, the framework studied in this paper allows for spatial dependence directly through contemporaneous dependence of individual units on their connections, and indirectly through possible cross-sectional dependence in the regressors. The econometric analysis of HSAR models presents new technical difficulty for identification as well as the estimation of a large set of spatial lag coefficients that must be estimated simultaneously.

Our analysis builds on the existing literature on SAR models, pioneered by Whittle (1954) and Cliff and Ord (1973), and further advanced in a number of important directions. The maximum likelihood approach of Cliff and Ord, which was developed for a pure spatial model, has been extended to cover panel data models with fixed effects and dynamics. Other estimation and testing techniques, such as the generalized method of moments (GMM), also have been proposed. Some of the key references to this literature include Upton and Fingleton (1985), Anselin (1988), Cressie (1993),

Kelejian and Robinson (1993), Ord and Getis (1995), Anselin and Bera (1998), and, more recently, Haining (2003), Lee (2004), Kelejian and Prucha (1999, 2010), Lin and Lee (2010), Lee and Yu (2010), LeSage and Pace (2010), Arbia (2010), Cressie and Wikle (2011), and Elhorst (2014). Extensions to dynamic panels are provided by Anselin (2001), Baltagi, Song, and Kon (2003), Kapoor, Kelejian, and Prucha (2007), Baltagi, Song, Jung, and Kon (2007), and Yu, de Jong, and Lee (2008, 2012). Spatial techniques also have proved useful when analyzing network effects as can be seen in the pioneering work of Case (1991) and Manski (1993).

Almost all these contributions (whether in the context of spatial or network models) assume that, apart from possibly fixed effects, spatial spillover or network effects are homogeneous. However, even if all units in a network have the exact same number of connections, it can be the case that not all units are equally important or influential. Therefore, the assumption of a homogeneous spatial coefficient is likely to be restrictive, and should be relaxed when T is large. As shown in this paper, when T is large the heterogeneous spatial model can be estimated for any N and it is not required that $N \rightarrow \infty$, which is needed for estimation of the traditional SAR model with a single spatial coefficient when T is small.

Examples of panel data sets with T large include panels that cover counties, regions, or countries in the analysis of economic variables such as house prices, real wages, employment and income. For instance, in the empirical applications by Baltagi and Levin (1986) on demand for tobacco consumption, and by Holly, Pesaran, and Yamagata (2010) on house price diffusion across states in the USA, it is interesting to investigate whether the maintained assumption that spillover effects from neighboring states are the same across all the 48 mainland states in fact holds, particularly considering the large size of the USA and the uneven distribution of economic activity across it.

While estimation of HSAR panel data models can be carried out using MLE and GMM approaches, in this paper we focus on the former and discuss identification, estimation, and inference using the quasi maximum likelihood (QML) method. We derive conditions under which the QML estimators of the *individual* parameters are locally identified, and establish consistency and asymptotic normality of the estimators under certain regularity conditions. Asymptotic covariances of the QML estimators are derived under both Gaussian and non-Gaussian errors, and consistent estimators of these covariances are proposed. Alternative estimation methods based on our HSAR model include the Bayesian Markov chain Monte Carlo approach of LeSage and Chih (2018a) and the generalized Yule–Walker estimation method of Dou, Parrella, and Yao (2016).

Although the estimation of individual coefficients of the HSAR model can be carried out for *any* N when T is large, the estimation of the mean of the coefficients across the units requires both N and T to be large. Accordingly, we propose an estimator of the cross-section mean of the individual parameters (also known in the literature as the mean group (MG) estimator) assuming a random coefficient model, and show that the MG estimators are consistent and asymptotically normal if both N and T tend to infinity jointly, such that $\sqrt{N}/T \rightarrow 0$, and the spatial dependence is sufficiently weak. Such estimators are helpful in two respects. They provide an overall average estimator of the spatial effects that could be compared to corresponding estimates obtained using standard homogeneous SAR models. They can also be used to obtain average estimators across subspatial groupings such as states or regions, or subgroups within a production or financial network, such as industry types.

The small-sample performance of the QML estimator is investigated by Monte Carlo simulations for different values of N and T and alternative choices of the spatial weight matrices. The simulation results are in line with the paper's key theoretical findings, and show that the proposed estimators have good small-sample properties for panels with moderate time dimensions and irrespective of the number of cross-section units in the panel, although under non-Gaussian errors, tests based on QML estimators of the spatial parameters can show slight size distortions when the time dimension is relatively small. We also investigate the small-sample performance of the MG estimator and find its performance to be satisfactory with biases that are universally negligible, and root mean square errors (RMSEs) that decline with T and quite rapidly with N . Regarding size and power, tests based on the MG estimator exhibit some downward size distortions when T is small, but such distortions disappear as T rises for all values of N . The small-sample bias of the MG estimator can be reduced using the half-Jackknife procedure, as discussed in Chudik and Pesaran (2019).

We provide an empirical application by modeling the spatial and temporal dimensions of quarterly US house prices changes over the period 1975:Q1–2014:Q4 and across metropolitan statistical areas (MSAs). Not surprisingly, we find a considerable degree of heterogeneity across the MSA-specific estimates. As to be expected, the estimates of net spatial coefficients (contemporaneous and lagged) are mostly positive and statistically significant, suggesting a high degree of spillover effects of house price changes to neighboring MSAs. There were only 18 MSAs (out of 338 considered in our analysis) with statistically significant negative net spatial effects, and included Pittsfield (Massachusetts), Minneapolis (Minnesota), and Memphis (Arkansas). These MSAs tend to be relatively remote with outward migratory flows to the neighboring regions.

We also consider mean estimates for six US regions, and find the contemporaneous spatial lag coefficients to be all positive and statistically significant for all the regions. However, the net spatial effects, computed as the sum of the coefficients of contemporaneous and lagged spatial variables are positive in all regions but statistically significant only in the case of half of the regions, namely Great Lakes, South East and Far West. This result clearly shows the importance of allowing for dynamics in the analysis of spatial effects, and is to be contrasted with the large (net) spatial effect of around 0.65 found by Yang (2020), who considered a homogeneous and static SAR specification. We also find positive and statistically significant effects of population and income growth on own-region house price changes, again with a high degree of heterogeneity across the regions. Using techniques analogous to the work of LeSage and Chih (2016), we also report direct and indirect partial effects of changes in population and income growth on house prices over time. The results continue to exhibit a high degree of heterogeneity across MSAs and regions with direct effects dominating and the outcomes decaying quite rapidly.

The rest of the paper is organized as follows: Section 2 introduces the first-order spatial autoregressive model with heterogeneous coefficients and some useful generalizations, and derives its log-likelihood function. Section 3 sets out the assumptions of the model, derives the identification conditions and proves consistency and asymptotic normality of the QML estimator when the time dimension is large. Section 4 outlines the MG estimator derived from the heterogeneous spatial coefficients of the HSAR model. Section 5 presents the Monte Carlo design and reports small-sample results (bias, RMSEs, size, and power) of the QML and MG estimators for different parameter values and sample size combinations. Section 6 reports the results of the empirical application to the US house price changes across MSAs. Some concluding remarks are provided in Section 7. Mathematical proofs, data sources, and additional empirical and Monte Carlo results are provided in the online Supporting Information.

Notations: We denote the largest and the smallest eigenvalues of the $N \times N$ matrix $\mathbf{A} = (a_{ij})$ by $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$, respectively, its trace by $\text{tr}(\mathbf{A}) = \sum_{i=1}^N a_{ii}$, its spectral radius by $\rho(\mathbf{A}) = |\lambda_{\max}(\mathbf{A})|$, its spectral norm by $\|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, its maximum absolute column sum norm by $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq N} \left(\sum_{i=1}^N |a_{ij}| \right)$, and its maximum absolute row sum norm by $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq N} \left(\sum_{j=1}^N |a_{ij}| \right)$. $\text{Diag}(\mathbf{A}) = \text{Diag}(a_{11}, a_{22}, \dots, a_{NN})$ represents an $N \times N$ diagonal matrix formed by the diagonal elements of \mathbf{A} , while $\text{diag}(\mathbf{A}) = (a_{11}, a_{22}, \dots, a_{NN})'$ denotes an $N \times 1$ vector. We denote the ℓ_p -norm of the random variable x by $\|x\|_p = E(|x|^p)^{1/p}$ for $p \geq 1$, assuming that $E(|x|^p) < K$. \odot stands for Hadamard product or element-wise matrix product operator, \rightarrow_p denotes convergence in probability, $\xrightarrow{\text{a.s.}}$ almost sure convergence, \rightarrow_d convergence in distribution, and $\stackrel{a}{\sim}$ asymptotic equivalence in distribution. K and c will be used to denote finite large and nonzero small positive numbers, respectively, that do not depend on N and/or T .

2 | A HETEROGENEOUS SPATIAL AUTOREGRESSIVE MODEL (HSAR)

2.1 | Model specification

We consider the following SAR model with heterogeneous slopes:

$$y_{it} = \psi_{i0} \left(\sum_{j=1}^N w_{ij} y_{jt} \right) + \beta'_{i0} \mathbf{x}_{it} + \varepsilon_{it}, \text{ for } i = 1, 2, \dots, N; t = 1, 2, \dots, T, \quad (1)$$

where y_{it} is the dependent variable for unit i observed at time t , $\mathbf{x}_{it} = (x_{i1,t}, x_{i2,t}, \dots, x_{ik,t})'$ is a $k \times 1$ vector of (weakly) exogenous regressors, with the associated $k \times 1$ vector of slope parameters, and $\beta_{i0} = (\beta_{i1,0}, \beta_{i2,0}, \dots, \beta_{ik,0})'$. ε_{it} is the unexplained component of y_{it} , which we refer to as the error of the i th cross-section unit, or the “error” for short, with variance $\text{var}(\varepsilon_{it}) = \sigma_{i0}^2$. Finally, $y_{it}^* = \sum_{j=1}^N w_{ij} y_{jt} = \mathbf{w}_i' \mathbf{y}_t$ is the average effect of other units on unit i at time t , where $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{Nt})'$ and \mathbf{w}_i' is the i th row of the $N \times N$ spatial weight matrix, $\mathbf{W} = (w_{ij})$, with w_{ij} , for $i, j = 1, 2, \dots, N$ being the spatial weights. Without loss of generality we set $w_{ii} = 0$, for all i , assume that $w_{ij} \geq 0$, and normalize the spatial weights so that $\sum_{j=1}^N w_{ij} = 1$.¹ When the weights are not normalized, Equation 1 continues to hold with ψ_{i0} redefined as ψ_{i0}/v_i ,

¹Strictly speaking, the weights, w_{ij} , are N -dependent and should be denoted as $w_{ij,N}$. The same also applies to y_{it} , β_{i0} , and ε_{it} . But we abstract from including the subscript N when denoting w_{ij} , y_{it} , and ε_{it} , to keep the notations simple and manageable.

where $\sum_{j=1}^N w_{ij} = v_i$. Consequently, in the heterogeneous case the normalization of the weights is innocuous, and can be viewed as an identifying restriction, so that ψ_{i0} can be distinguished from v_i , which is achieved by setting $v_i = 1$. The same is not true in the homogeneous case where $\psi_{i0} = \psi_0$ for all i , and the use of non-normalized weights is equivalent to setting $\psi_{i0} = \psi_0/v_i$, which is not an innocuous restriction. The HSAR model (Equation 1) can also be viewed as a generalization of the random coefficient panel data model reviewed, for example, by Hsiao and Pesaran (2008). However, this is not a straightforward generalization due to the endogeneity of $y_{it}^* = \mathbf{w}'_i \mathbf{y}_t$ in Equation 1.

The assumption of non-negative weights ($w_{ij} \geq 0$) can be relaxed by replacing \mathbf{W} with two weight matrices: one for positive weights, $\mathbf{W}^+ = (w_{ij}^+)$, where $w_{ij}^+ = w_{ij}$ if $w_{ij} > 0$ and zero otherwise, and one for negative weights, $\mathbf{W}^- = (w_{ij}^-)$, where $w_{ij}^- = -w_{ij}$ if $w_{ij} < 0$ and zero otherwise. Further, since regressors are allowed to be weakly exogenous, our analysis covers quite general spatiotemporal models, such as the following generalization of Equation 1:

$$y_{it} = \sum_{q=1}^{h_\lambda} \lambda_{iq0} y_{i,t-q} + \sum_{q=0}^{h_\psi^+} \psi_{iq0}^+ \left(\sum_{j=1}^N w_{ij}^+ y_{j,t-q} \right) + \sum_{q=0}^{h_\psi^-} \psi_{iq0}^- \left(\sum_{j=1}^N w_{ij}^- y_{j,t-q} \right) + \tilde{\beta}'_{i0} \tilde{\mathbf{x}}_{it} + \varepsilon_{it}, \quad (2)$$

where h_λ , h_ψ^+ and h_ψ^- are fixed, and the slope coefficients, λ_{iq0} , ψ_{iq0}^+ and ψ_{iq0}^- measure the temporal effects and spatial impact effects for positively and negatively connected units. Such a model is analyzed in Bailey, Holly, and Pesaran (2016).²

The HSAR model can be generalized further in two important directions. First, the assumption of zero diagonal elements for the weight matrix can be relaxed, which could be of interest in linking spatial models more closely to the global autoregression (GVAR) approach, as discussed in Elhorst, Gross, and Tereanu (2018). Consider Equation 1, with $w_{ii} \neq 0$ and $\sum_{j=1}^N w_{ij} = v_i$, where v_i are known constants. Then Equation 1 can be reparametrized as

$$y_{it} = \dot{\psi}_{i0} \left(\sum_{j=1}^N w_{ij} y_{jt} \right) + \dot{\beta}'_{i0} \mathbf{x}_{it} + \dot{\varepsilon}_{it}, \text{ for } i = 1, 2, \dots, N; t = 1, 2, \dots, T, \quad (3)$$

where $\dot{\psi}_{i0} = (1 - \psi_{i0} w_{ii})^{-1} \psi_{i0} v_i$, $\dot{\beta}_{i0} = (1 - \psi_{i0} w_{ii})^{-1} \beta_{i0}$, and $\text{var}(\dot{\varepsilon}_{it}) = \dot{\sigma}_{i0}^2 = (1 - \psi_{i0} w_{ii})^{-2} \sigma_{i0}^2$, respectively.³ Second, the weights, w_{ij} , can be estimated so long as each unit has a finite number of known neighbors. In such a setting the HSAR model can be written as

$$y_{it} = \sum_{j=1}^N \psi_{ij0} I(w_{ij}) y_{jt} + \beta'_{i0} \mathbf{x}_{it} + \varepsilon_{it}, \quad (4)$$

where $I(w_{ij}) = 1$ if $w_{ij} \neq 0$ and 0 otherwise, and $\sup_i \sum_{j=1}^N |\psi_{ij0}| I(w_{ij}) < K$. This specification only exploits the qualitative information contained in $I(w_{ij})$ which departs from the conventional homogeneous spatial model. In what follows we focus on the basic HSAR specification given by Equation 1 and note that estimation and inference for models (2), (3) or (4) can be conducted along the lines set out in this paper.

Stacking the observations by the N individual units for each time period t , Equation 1 can be written more compactly as

$$(\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W}) \mathbf{y}_{\cdot t} = \mathbf{B}_0 \mathbf{x}_{\cdot t} + \boldsymbol{\varepsilon}_{\cdot t}, \quad t = 1, 2, \dots, T, \quad (5)$$

where $\mathbf{y}_{\cdot t} = (y_{1t}, y_{2t}, \dots, y_{Nt})'$, \mathbf{I}_N is an $N \times N$ identity matrix, $\boldsymbol{\Psi}_0 = \text{Diag}(\boldsymbol{\Psi}_0)$ with $\boldsymbol{\Psi}_0 = (\psi_{10}, \psi_{20}, \dots, \psi_{N0})'$, and \mathbf{B}_0 is the $N \times kN$ block diagonal matrix with elements β'_{i0} , $i = 1, 2, \dots, N$, on the main diagonal and zeros elsewhere, and $\mathbf{x}_{\cdot t} = (\mathbf{x}'_{1t}, \mathbf{x}'_{2t}, \dots, \mathbf{x}'_{Nt})'$ is the $kN \times 1$ vector of observations on the exogenous regressors. Finally, $\text{var}(\boldsymbol{\varepsilon}_{\cdot t}) = \boldsymbol{\Sigma}_0 = \text{Diag}(\boldsymbol{\sigma}_0^2)$, with $\boldsymbol{\sigma}_0^2 = (\sigma_{10}^2, \sigma_{20}^2, \dots, \sigma_{N0}^2)'$. We set $\mathbf{S}(\boldsymbol{\Psi}_0) = \mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W}$, and assuming that $\mathbf{S}(\boldsymbol{\Psi}_0)$ is invertible, the reduced form of Equation 5 can be expressed as

²It is also possible to allow for spatial effects in the errors and the regressors. For example, ε_{it} can be replaced by $\varepsilon_{it} = \phi_{i0} \left(\sum_{j=1}^N w_{e,ij} \varepsilon_{jt} \right) + \nu_{it}$, and the regressors augmented with spatial effects such as $\tilde{x}_{i\ell,t}^* = \sum_{j=1}^N w_{\ell,ij} \tilde{x}_{j\ell,t}$, for $\ell = 1, 2, \dots, k'$, where $w_{e,ij}$ and $w_{\ell,ij}$ are the spatial weights. To simplify the exposition in this paper we abstract from spatial error and regressor processes and focus on the contemporaneous spatial effects in the dependent variable, y_{it} .

³The derivation for this modified HSAR specification can be found in the online supplement D (Supporting Information).

$$\mathbf{y}_{\circ t} = \mathbf{S}^{-1}(\boldsymbol{\psi}_0)[\mathbf{B}_0 \mathbf{x}_{\circ t} + \boldsymbol{\varepsilon}_{\circ t}], \quad t = 1, 2, \dots, T. \quad (6)$$

2.2 | The log-likelihood function

To estimate the unit-specific coefficients we collect all the parameters of the N units in the $N(k+2) \times 1$ vector $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}', \boldsymbol{\sigma}^2)'$, where $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_N)'$, $\boldsymbol{\beta} = (\beta'_1, \beta'_2, \dots, \beta'_N)'$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)'$, and denote the associated vector of true values by $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}'_0, \boldsymbol{\beta}'_0, \boldsymbol{\sigma}'^2_0)'$. The log-likelihood function of (6) can be written as

$$\ell_T(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) = -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^N \ln \sigma_i^2 + \frac{T}{2} \ln |\mathbf{S}'(\boldsymbol{\psi}) \mathbf{S}(\boldsymbol{\psi})| - \frac{1}{2} \sum_{t=1}^T [\mathbf{S}(\boldsymbol{\psi}) \mathbf{y}_{\circ t} - \mathbf{B} \mathbf{x}_{\circ t}]' \boldsymbol{\Sigma}^{-1} [\mathbf{S}(\boldsymbol{\psi}) \mathbf{y}_{\circ t} - \mathbf{B} \mathbf{x}_{\circ t}], \quad (7)$$

where $\boldsymbol{\Sigma} = \text{Diag}(\boldsymbol{\sigma}^2)$, $\boldsymbol{\Psi} = \text{Diag}(\boldsymbol{\psi})$, and $\mathbf{S}(\boldsymbol{\psi}) = \mathbf{I}_N - \boldsymbol{\Psi} \mathbf{W}$.

The quasi maximum likelihood estimators (QMLE), $\hat{\boldsymbol{\theta}}$, are the extreme value estimators obtained by maximization of Equation 7. When the error terms, $\boldsymbol{\varepsilon}_{\circ t}(\boldsymbol{\theta}_0) = \mathbf{S}(\boldsymbol{\psi}_0) \mathbf{y}_{\circ t} - \mathbf{B} \mathbf{x}_{\circ t}$, are normally distributed, then vector $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ while, under non-Gaussian errors, $\hat{\boldsymbol{\theta}}$ is the QMLE of $\boldsymbol{\theta}$.

3 | ASYMPTOTIC PROPERTIES OF QML ESTIMATORS

3.1 | Assumptions

In order to investigate the conditions under which $\boldsymbol{\theta}_0$ is identified, and to establish consistency and the asymptotic normality of $\hat{\boldsymbol{\theta}}$, we make the following assumptions, using the filtration $\mathcal{F}_t = (\mathbf{x}_{\circ t}, \mathbf{x}_{\circ t-1}, \mathbf{x}_{\circ t-2}, \dots)$, where $\mathbf{x}_{\circ t} = (\mathbf{x}'_{1t}, \mathbf{x}'_{2t}, \dots, \mathbf{x}'_{Nt})'$, which could also contain lagged values of $\mathbf{y}_{\circ t}$:

Assumption 1. The error terms $\{\varepsilon_{it}, i = 1, 2, \dots, N; t = 1, 2, \dots, T\}$ are independently distributed over i and t ; $E(\varepsilon_{it} | \mathcal{F}_t) = 0$, $E(\varepsilon_{it}^2 | \mathcal{F}_t) = \sigma_{i0}^2$, for $i = 1, 2, \dots, N$, where $\inf_i(\sigma_{i0}^2) > c > 0$, $\sup_i(\sigma_{i0}^2) < K < \infty$, and $E(|\varepsilon_{it}|^p | \mathcal{F}_t) = E(|\varepsilon_{it}|^p) = \varpi_{ip} < K$, for all i and t , where ϖ_{ip} is a time-invariant constant, $1 \leq p \leq 4 + \epsilon$, for some $\epsilon > 0$.

Assumption 2. (a) $\mathbf{x}_{\circ t}$ are stationary processes with mean zero, that satisfy the moment condition $\sup_{i,\ell,t} E(|x_{i\ell,t}|^{2+c}) < K$, for some $c > 0$, $i = 1, 2, \dots, N$, $\ell = 1, 2, \dots, k$, and $t = 1, 2, \dots, T$. (b) $E(\mathbf{x}_{\circ t} \mathbf{x}_{\circ t}') = \boldsymbol{\Sigma}_{xx} = (\boldsymbol{\Sigma}_{ij})$, where $\boldsymbol{\Sigma}_{ij} = E(\mathbf{x}_{it} \mathbf{x}'_{jt})$ exists for all i and j , such that $\sup_{i,j} \|\boldsymbol{\Sigma}_{ij}\| < K$, and $\boldsymbol{\Sigma}_{ii}$ is a $k \times k$ nonsingular matrix with $\inf_i [\lambda_{\min}(\boldsymbol{\Sigma}_{ii})] > c > 0$, and $\sup_i [\lambda_{\max}(\boldsymbol{\Sigma}_{ii})] < K$. (c) $T^{-1} \sum_{t=1}^T \mathbf{x}_{\circ t} \mathbf{x}_{\circ t}' \xrightarrow{a.s.} \boldsymbol{\Sigma}_{xx}$, as $T \rightarrow \infty$.

Assumption 3. The $N(k+2) \times 1$ parameter vector $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}', \boldsymbol{\sigma}^2)'$ belongs to $\boldsymbol{\Theta} = \boldsymbol{\Theta}_{\boldsymbol{\psi}} \times \boldsymbol{\Theta}_{\boldsymbol{\beta}} \times \boldsymbol{\Theta}_{\boldsymbol{\sigma}} \subset \mathbb{R}^N \times \mathbb{R}^{Nk} \times \mathbb{R}^N$, a subset of the $N(k+2)$ dimensional Euclidean space, $\mathbb{R}^{N(k+2)}$. $\boldsymbol{\Theta}$ is a closed and bounded (compact) set and includes the true value of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}_0$, which is an interior point of $\boldsymbol{\Theta}$.

Assumption 4. (a) The weight matrix $\mathbf{W} = (w_{ij})$ is known and time-invariant with $w_{ii} = 0$, for $i = 1, 2, \dots, N$. (b) \mathbf{W} has bounded maximum row sum norm, $\|\mathbf{W}\|_{\infty} < K < \infty$, and

$$\sup_{\boldsymbol{\psi}_i \in \boldsymbol{\Theta}_{\boldsymbol{\psi}}} |\psi_i| < \frac{1}{\|\mathbf{W}\|_{\infty}}. \quad (8)$$

Remark 1. Assumption 1 implies that $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma_{i0}^2$, for $i = 1, 2, \dots, N$, and does not allow for conditional heteroskedasticity. But it is possible to allow for time variations in $E(|\varepsilon_{it}|^{4+\epsilon} | \mathcal{F}_t)$ by relaxing the moment conditions on ε_{it} and $x_{i\ell,t}$.

Remark 2. The HSAR model, (Equation 1), is quite general and encompasses many other models in the literature. Assumption 2 allows the regressors to be weakly exogenous and cross-sectionally correlated, namely the model can contain lagged dependent variables and observable common factors, such as time trends. It can also be

modified to include an intercept (fixed effects) by setting one of the elements of \mathbf{x}_{it} to unity, at the expense of complicating the algebra. It applies both when N is small or large, so long as T is sufficiently large. Small-sample evidence on such settings is presented in Section 5.

Remark 3. Assumption 4 is sufficiently general and allows the spatial weights to take negative values. But, as noted above, in empirical applications one might wish to distinguish between positive and negative connections as they might have differential effects on the outcomes. This assumption does not require the weights to be normalized either, so long as condition (8) is met. In the case when a dense inverse distance matrix is adopted, for example when w_{ij} are set in terms of geodesic distance, d_{ij} , between the i and j units such that $\mathbf{W} = (1/d_{ij}^\delta)$, then for Assumption 4 to hold it is necessary that $\delta > 1$ and is sufficiently large. See Elhorst (2014).

Remark 4. As shown in Lemma 1 in the online supplement A (Supporting Information), under Assumption 4 we have $\lambda_{\min}[\mathbf{S}(\boldsymbol{\psi})] > 0$ and $|\lambda_{\max}[\mathbf{S}(\boldsymbol{\psi})]| < 2$, and boundedness of $\|\mathbf{W}\|_1$ is not needed. Condition (8) is required simply to ensure invertibility of matrix $\mathbf{S}(\boldsymbol{\psi})$.

3.2 | Identification

Here we focus on the problem of identification of the individual parameters in $N(k + 2) \times 1$ vector $\boldsymbol{\theta}_0$ for a given N , and as $T \rightarrow \infty$. To highlight the main issues involved in the identification of spatial parameters under the heterogeneous setting, first we consider the HSAR model (Equation 5) without the exogenous regressors. Under Assumption 4(b), with $\mathbf{B}_0 = \mathbf{0}$, we have (see Equation 6)

$$\mathbf{y}_{\circ t} = \mathbf{S}^{-1}(\boldsymbol{\psi})\boldsymbol{\epsilon}_{\circ t}, \quad t = 1, 2, \dots, T.$$

With a slight abuse of notation let $\boldsymbol{\theta} = (\boldsymbol{\psi}', \sigma^2)'$, and note that in this case the log-likelihood function is given by

$$\ell_T(\boldsymbol{\theta}) = -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^N \ln \sigma_i^2 + \frac{T}{2} \ln |\mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})| - \frac{1}{2} \sum_{t=1}^T \mathbf{y}'_{\circ t} \mathbf{S}'(\boldsymbol{\psi}) \boldsymbol{\Sigma}^{-1} \mathbf{S}(\boldsymbol{\psi}) \mathbf{y}_{\circ t}.$$

It is also helpful to write the associated average log-likelihood function as

$$\bar{\ell}_T(\boldsymbol{\theta}) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln \sigma_i^2 + \frac{1}{2} \ln |\mathbf{V}(\boldsymbol{\psi})| - \frac{1}{2} \left(\frac{1}{T} \sum_{t=1}^T \mathbf{y}'_{\circ t} \mathbf{P}(\boldsymbol{\theta}) \mathbf{y}_{\circ t} \right),$$

where

$$\mathbf{V}(\boldsymbol{\psi}) = \mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi}), \quad \mathbf{P}(\boldsymbol{\theta}) = \mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi}), \text{ and } \mathbf{S}(\boldsymbol{\psi}) = \mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W}.$$

Let $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})$, in which $\bar{\ell}_T(\boldsymbol{\theta}_0)$ is $\bar{\ell}_T(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Then, for a given N , and as $T \rightarrow \infty$, we have (see Lemma 3 of online supplement A (Supporting Information) when setting $\mathbf{B} = \mathbf{0}$ in A.6) $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) - E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] \xrightarrow{\text{a.s.}} 0$, where

$$E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] = E_0[\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})] = -\frac{1}{2} \sum_{i=1}^N \ln(\sigma_{i0}^2 / \sigma_i^2) - \frac{N}{2} + \frac{1}{2} \left[\ln \left(\frac{|\mathbf{V}(\boldsymbol{\psi}_0)|}{|\mathbf{V}(\boldsymbol{\psi})|} \right) \right] + \frac{1}{2} \text{tr}[\mathbf{P}(\boldsymbol{\theta}) \mathbf{P}^{-1}(\boldsymbol{\theta}_0)]. \quad (9)$$

Consider now the problem of identification of $\boldsymbol{\psi}_0$, which is the parameter vector of interest. Note that

$$\begin{aligned} |\mathbf{V}(\boldsymbol{\psi}_0)| / |\mathbf{V}(\boldsymbol{\psi})| &= |\mathbf{S}(\boldsymbol{\psi}_0)|^2 / |\mathbf{S}(\boldsymbol{\psi})|^2 = |\mathbf{S}(\boldsymbol{\psi}_0)\mathbf{S}^{-1}(\boldsymbol{\psi})|^2 = |\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)|^{-2}, \\ \text{tr}[\mathbf{P}(\boldsymbol{\theta}) \mathbf{P}^{-1}(\boldsymbol{\theta}_0)] &= \text{tr}[\mathbf{S}'(\boldsymbol{\psi}) \boldsymbol{\Sigma}^{-1} \mathbf{S}(\boldsymbol{\psi}) \mathbf{S}^{-1}(\boldsymbol{\psi}_0) \mathbf{S}_0 \mathbf{S}'^{-1}(\boldsymbol{\psi}_0)] = \text{tr}[\boldsymbol{\Sigma}^{-1/2} \mathbf{S}(\boldsymbol{\psi}) \mathbf{S}^{-1}(\boldsymbol{\psi}_0) \mathbf{S}_0 \mathbf{S}'^{-1}(\boldsymbol{\psi}_0) \mathbf{S}'(\boldsymbol{\psi}) \boldsymbol{\Sigma}^{-1/2}], \end{aligned}$$

and rewrite Equation 9 as

$$E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] = -\frac{N}{2} - \frac{1}{2} \sum_{i=1}^N \ln(\sigma_{i0}^2/\sigma_i^2) - [\ln(|\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)|)] + \frac{1}{2} \text{tr}[\boldsymbol{\Sigma}^{-1/2} \mathbf{S}(\boldsymbol{\psi}) \mathbf{S}^{-1}(\boldsymbol{\psi}_0) \mathbf{S}_0 \mathbf{S}'^{-1}(\boldsymbol{\psi}_0) \mathbf{S}'(\boldsymbol{\psi}) \boldsymbol{\Sigma}^{-1/2}].$$

Further, we note that $\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0) = \mathbf{I}_N - \mathbf{D}\mathbf{G}_0$, where $\mathbf{G}_0 = \mathbf{W}(\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1}$, and $\mathbf{D} = \boldsymbol{\Psi} - \boldsymbol{\Psi}_0$, is a diagonal matrix with elements $d_i = \psi_i - \psi_{i0}$. Using these results, the above expression for $E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})]$ can be written equivalently as

$$\begin{aligned} E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] &= -\frac{N}{2} - \frac{1}{2} \sum_{i=1}^N \ln(\sigma_{i0}^2/\sigma_i^2) - \ln|\mathbf{I}_N - \mathbf{D}\mathbf{G}_0| + \frac{1}{2} \text{tr}[\boldsymbol{\Sigma}^{-1/2} (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)' \mathbf{S}_0 (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0) \boldsymbol{\Sigma}^{-1/2}] \\ &= A_N + B_N, \end{aligned}$$

where

$$A_N = \frac{1}{2} \sum_{i=1}^N [(\sigma_{i0}^2/\sigma_i^2) - \ln(\sigma_{i0}^2/\sigma_i^2) - 1] - \ln|\mathbf{I}_N - \mathbf{D}\mathbf{G}_0| - \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_0 \mathbf{D}\mathbf{G}_0), \quad (10)$$

$$B_N = \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1/2} \mathbf{G}_0' \mathbf{D} \boldsymbol{\Sigma}_0 \mathbf{D} \mathbf{G}_0 \boldsymbol{\Sigma}^{-1/2}). \quad (11)$$

We first note that $B_N \geq 0$, since we can write $B_N = (1/2)\text{tr}(\mathbf{A}_0' \mathbf{A}_0)$, with $\mathbf{A}_0 = \boldsymbol{\Sigma}_0^{1/2} \mathbf{D}\mathbf{G}_0 \boldsymbol{\Sigma}^{-1/2}$. Consider now A_N , denote the i th eigenvalue of $\mathbf{D}\mathbf{G}_0$ by μ_i , and note that since $\mathbf{I}_N - \mathbf{D}\mathbf{G}_0 = \mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)$, then the eigenvalues of $\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)$ are also given by $1 - \mu_i$, for $i = 1, 2, \dots, N$. Further, by Lemma 1 in the online supplement A (Supporting Information), $\lambda_{\min}[\mathbf{S}(\boldsymbol{\psi})] > 0$ for all ψ_i that satisfy condition (8). Hence we must also have $1 - \mu_i > 0$, for all i . Using these results, A_N can now be written as

$$A_N = \frac{1}{2} \sum_{i=1}^N \left[\frac{\sigma_{i0}^2}{\sigma_i^2} - \ln(\sigma_{i0}^2/\sigma_i^2) - 1 \right] - \sum_{i=1}^N \ln(1 - \mu_i) - \sum_{i=1}^N \left(\frac{\sigma_{i0}^2}{\sigma_i^2} \right) \mu_i.$$

Let $\delta_{\sigma i} = \sigma_{i0}^2/\sigma_i^2 > 0$ and $\delta_{\psi i} = (1 - \mu_i) > 0$, for all i . Then, write $E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})]$ as

$$\begin{aligned} E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] &= A_N + B_N = \frac{1}{2} \sum_{i=1}^N [\delta_{\sigma i} - \ln(\delta_{\sigma i}) - 1] - \sum_{i=1}^N \ln \delta_{\psi i} - \sum_{i=1}^N \delta_{\sigma i} (1 - \delta_{\psi i}) + B_N \\ &= \frac{1}{2} \sum_{i=1}^N [\delta_{\sigma i} - \ln(\delta_{\sigma i}) - 1] + \left[\sum_{i=1}^N \delta_{\sigma i} (\delta_{\psi i} - \ln \delta_{\psi i} - 1) \right] \\ &\quad + \left[\sum_{i=1}^N (\delta_{\sigma i} - 1) \ln \delta_{\psi i} \right] + B_N = A_{1,N} + A_{2,N} + A_{3,N} + B_N. \end{aligned}$$

Since $\delta_{\sigma i} > 0$, and $\delta_{\psi i} > 0$ for all i , then $\delta_{\sigma i} - \ln(\delta_{\sigma i}) - 1 \geq 0$, and $\delta_{\psi i} - \ln \delta_{\psi i} - 1 \geq 0$ for all i , with equalities holding if and only if $\delta_{\sigma i} = 1$ and $\delta_{\psi i} = 1$ for all i . Hence, $A_{1,N} \geq 0$, and $A_{2,N} \geq 0$ for all values of N , and global identification of σ_{i0}^2 will be possible only if we are able to show that $A_{3,N} + B_N$ is nonnegative. But it is easily seen that the non-negativity of $A_{3,N} + B_N$ can not be guaranteed without further restrictions. This follows since

$$A_{3,N} = \sum_{i=1}^N (\delta_{\sigma i} - 1) \ln \delta_{\psi i},$$

and there are values of $\delta_{\sigma i}$ and $\delta_{\psi i}$ in $\boldsymbol{\Theta} = \boldsymbol{\Theta}_{\psi} \times \boldsymbol{\Theta}_{\sigma}$ for which $A_{3,N} < 0$. Considering $A_{3,N} + B_N$ somewhat weakens the requirement since $B_N \geq 0$, but still does not guarantee that $(A_{3,N} + B_N) \geq 0$, for all values of $\delta_{\sigma i} > 0$ and $\delta_{\psi i} > 0$. Therefore,

global identification of ψ_0 can not be guaranteed. To investigate the possibility of local identification we introduce the following definition:

Definition 1. Consider the set $\mathcal{N}_c(\sigma_0^2)$ in the closed neighborhood of σ_0^2 defined by

$$\mathcal{N}_c(\sigma_0^2) = \{\sigma_0^2 \in \Theta_\sigma, |\sigma_{i0}^2/\sigma_i^2 - 1| < c_i, \text{for } i = 1, 2, \dots, N\},$$

for some $c_i > 0$, $i = 1, 2, \dots, N$, where Θ_σ is a compact subset of \mathbb{R}^N .

We now show that $\theta_0 = (\psi'_0, \sigma_0^2)'$ is identified on $\Theta_c = \Theta_\psi \times \mathcal{N}_c(\sigma_0^2)$. Consider values of $\delta_{\sigma i}$ within the local neighborhood of $\delta_{\sigma i} = 1$ for all i . Recall that $A_{1,N} + A_{2,N} \geq 0$, and the boundary values $A_{1,N} = 0$ or $A_{2,N} = 0$ can occur if and only if $\delta_{\sigma i} = 1$ and $\delta_{\psi i} = 1$ for all i , respectively. Therefore, $A_N \geq 0$ if $\delta_{\sigma i} = 1$, otherwise $A_{1,N} > 0$. Similarly, $A_N \geq 0$ if $\delta_{\psi i} = 1$, otherwise $A_{2,N} > 0$. Therefore, there must exist $\mathbf{c} = (c_1, c_2, \dots, c_N)' > \mathbf{0}$, such that $A_N = 0$ on Θ_c if and only if $\theta = \theta_0$, which in turn establishes that $\bar{\ell}_T(\theta_0) - \bar{\ell}_T(\theta) \xrightarrow{a.s.} 0$, as $T \rightarrow \infty$, on the set Θ_c if and only if $\theta = \theta_0$.

Next, consider the HSAR model (Equation 5) with exogenous regressors. The average log-likelihood in this case is given by (see Equation 7)

$$\bar{\ell}_T(\theta) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^N \ln \sigma_i^2 + \frac{1}{2} \ln |\mathbf{V}(\psi)| - \frac{1}{2} \left(\frac{1}{T} \sum_{t=1}^T [\mathbf{S}(\psi) \mathbf{y}_{\cdot t} - \mathbf{B} \mathbf{x}_{\cdot t}]' \Sigma^{-1} [\mathbf{S}(\psi) \mathbf{y}_{\cdot t} - \mathbf{B} \mathbf{x}_{\cdot t}] \right), \quad (12)$$

where θ is now defined by $\theta = (\psi', \beta', \sigma^2)'$ and \mathbf{B} has the same form as that used in Equation 5. Following a similar line of reasoning as in the case without exogenous regressors (see Lemma 3 in the online supplement A (Supporting Information)), we have that $Q_T(\theta_0, \theta) = \bar{\ell}_T(\theta_0) - \bar{\ell}_T(\theta)$, where $\bar{\ell}_T(\theta)$ is now given by Equation 12, and $Q_T(\theta_0, \theta) - E_0[Q_T(\theta_0, \theta)] \xrightarrow{a.s.} 0$, (as $T \rightarrow \infty$) where

$$E_0[Q_T(\theta_0, \theta)] = A_N + B_N + C_N. \quad (13)$$

A_N and B_N are defined as before by Equations 10 and 11, and C_N is given by

$$C_N = \frac{1}{2} \sum_{i=1}^N [(\beta_i - \beta_{i0})' \Sigma_{ii} (\beta_i - \beta_{i0})] / \sigma_i^2 + \text{tr}[\Sigma^{-1/2} (\mathbf{B} - \mathbf{B}_0) \Sigma_{xx} \Xi_0'] + \frac{1}{2} \text{tr}(\Sigma_{xx} \Xi_0' \Xi_0) = C_{1,N} + C_{2,N} + C_{3,N}, \quad (14)$$

where $\Xi_0 = \Sigma^{-1/2} \mathbf{D} \mathbf{G}_0 \mathbf{B}_0$, and as before $\mathbf{D} = \text{Diag}(\psi - \psi_0)$. Consider now $C_{3,N}$ and note that since $\Sigma_{xx} = E(\mathbf{x}_{\cdot t} \mathbf{x}_{\cdot t}')$ and $\Xi_0' \Xi_0$ are both $kN \times kN$ positive semi-definite matrices, then by result (9) on p. 44 of Lütkepohl (1996),

$$\frac{1}{kN} \text{tr}(\Sigma_{xx} \Xi_0' \Xi_0) \geq [\det(\Sigma_{xx})]^{1/kN} [\det(\Xi_0' \Xi_0)]^{1/kN} \geq 0,$$

and hence $C_{3,N} \geq 0$. Also, as shown above, on the subset $\Theta_c = \Theta_\psi \times \Theta_\beta \times \mathcal{N}_c(\sigma_0^2)$, $A_N + B_N = 0$ if and only if $\mathbf{D} = \text{Diag}(\psi - \psi_0) = 0$, and hence it must also follow that $C_{2,N} = 0$ on Θ_c . Thus, overall $\bar{\ell}_T(\theta_0) - \bar{\ell}_T(\theta) \xrightarrow{a.s.} 0$ on Θ_c if and only if

$$\sum_{i=1}^N (\beta_i - \beta_{i0})' \Sigma_{ii} (\beta_i - \beta_{i0}) / \sigma_i^2 = 0. \quad (15)$$

This equality holds for all N if and only if $(\beta_i - \beta_{i0})' \Sigma_{ii} (\beta_i - \beta_{i0}) = 0$, for all i , and since under Assumption 2(b) Σ_{ii} is a positive definite matrix this can occur if and only if $\beta_i = \beta_{i0}$ for all i .

Before we state the identification result for the general model (Equation 5), we require the following modification of Assumption 3:

Assumption 5. The $N(k+2) \times 1$ parameter vector $\theta = (\psi', \beta', \sigma^2)'$ belongs to $\Theta_c = \Theta_\psi \times \Theta_\beta \times \mathcal{N}_c(\sigma_0^2)$, where Θ_ψ and Θ_β are compact subsets of \mathbb{R}^N and \mathbb{R}^{Nk} , respectively, $\mathcal{N}_c(\sigma_0^2)$ is given in Definition 1, and Θ_c is a subset of the $N(k+2)$ dimensional Euclidean space, $\mathbb{R}^{N(k+2)}$.

The main identification result of the paper is summarized in the following proposition:

Proposition 1. Consider the heterogeneous spatial autoregressive (HSAR) model given by Equation 5 with the associated log-likelihood function given by Equation 7. Suppose that Assumptions 1-5 hold. Then for a fixed N and k , the $N(k+2)$ dimensional true parameter vector $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}'_0, \boldsymbol{\beta}'_0, \sigma^2_0)'$ is almost surely locally identified on Θ_c .

3.3 | Consistency and asymptotic normality

We are now in a position to consider consistency and asymptotic normality of the QML estimator of $\boldsymbol{\theta}$, given by $\hat{\boldsymbol{\theta}} = \arg_{\boldsymbol{\theta}} \bar{\ell}_T(\boldsymbol{\theta})$, where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\psi}}', \hat{\boldsymbol{\beta}}', \hat{\sigma}^2)'$, which is estimated simultaneously. We establish the results for a given N , and as $T \rightarrow \infty$. First, we focus on the proof of consistency. Under Assumptions 1, 2, 4, and 5, we have: (i) Θ_c , being a subset of Θ , is compact; (ii) $\boldsymbol{\theta}_0$ is an interior point of Θ_c ; (iii) $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \xrightarrow{a.s.} E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})]$, with $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})$ and $E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] = A_N + B_N + C_N$, where A_N , B_N , and C_N are given by Equation 10, 11, and 14, respectively; and (iv) $\boldsymbol{\theta}_0$ is a unique maximum of $E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})]$ on Θ_c . The last result follows from the identification analysis of Section 3.2. It is clear that all conditions of theorem 9.3.1 of Davidson (2000) are satisfied; therefore almost sure local consistency of $\hat{\boldsymbol{\theta}}$ is ensured, with $\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$ on Θ_c , as $T \rightarrow \infty$. To establish asymptotic normality of $\hat{\boldsymbol{\theta}}$, we apply the mean value theorem to $\bar{\ell}_T(\boldsymbol{\theta})$ such that

$$\bar{\ell}_T(\boldsymbol{\theta}) - \bar{\ell}_T(\boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \bar{\mathbf{s}}_T(\boldsymbol{\theta}_0) - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \bar{\mathbf{H}}_T(\bar{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \quad (16)$$

where $\bar{\mathbf{s}}_T(\boldsymbol{\theta}) = \partial \bar{\ell}_T(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$, $\bar{\mathbf{H}}_T(\boldsymbol{\theta}) = -\partial^2 \bar{\ell}_T(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$, and $\bar{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. By Lemma 5 of the online supplement A (Supporting Information) we have $\bar{\mathbf{s}}_T(\boldsymbol{\theta}_0) \xrightarrow{a.s.} \mathbf{0}$, and by the results of Section 3.2 we also have $\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta}) \xrightarrow{a.s.} E_0[\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})] \geq 0$. Hence, in view of Equation 16, it must also hold that (as $T \rightarrow \infty$)

$$(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \bar{\mathbf{H}}_T(\bar{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \xrightarrow{a.s.} E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})],$$

where $E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})]$ is given by Equation 13. But we have already established that on Θ_c , the right-hand side of the above expression can be equal to zero if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and hence it must be that $\bar{\mathbf{H}}_T(\bar{\boldsymbol{\theta}}) \xrightarrow{a.s.} \bar{\mathbf{H}}(\boldsymbol{\theta}_0)$, where $\bar{\mathbf{H}}(\boldsymbol{\theta}_0)$ must be a positive definite matrix given by $\bar{\mathbf{H}}(\boldsymbol{\theta}_0) = \lim_{T \rightarrow \infty} E_0[-\partial^2 \bar{\ell}_T(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}']$. Next, for a given N we apply the mean value theorem to $\bar{\mathbf{s}}_T(\boldsymbol{\theta}) = 1/\sqrt{T} \mathbf{s}_T(\boldsymbol{\theta})$ so that

$$0 = \sqrt{T} \bar{\mathbf{s}}_T(\hat{\boldsymbol{\theta}}) = \sqrt{T} \bar{\mathbf{s}}_T(\boldsymbol{\theta}_0) - \bar{\mathbf{H}}_T(\check{\boldsymbol{\theta}}) \sqrt{T} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\mathbf{s}_T(\boldsymbol{\theta}) = \partial \ell_T(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$, $\mathbf{H}_T(\boldsymbol{\theta}) = -\frac{1}{T} \partial^2 \ell_T(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$, and $\check{\boldsymbol{\theta}}$ lies between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Therefore, $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathbf{H}_T^{-1}(\check{\boldsymbol{\theta}})[\sqrt{T} \mathbf{s}_T(\boldsymbol{\theta}_0)]$, and since $\hat{\boldsymbol{\theta}}$ is consistent on Θ_c , then

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{a} \mathbf{H}^{-1}(\boldsymbol{\theta}_0) [\sqrt{T} \mathbf{s}_T(\boldsymbol{\theta}_0)],$$

where $\mathbf{H}(\boldsymbol{\theta}_0) = \lim_{T \rightarrow \infty} E_0[-\frac{1}{T} \partial^2 \ell_T(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}']$, with

$$E_0 \left[-\frac{1}{T} \partial^2 \ell_T(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' \right] = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \mathbf{H}_{13} \\ \mathbf{H}'_{12} & \mathbf{H}_{22} & \mathbf{H}_{23} \\ \mathbf{H}'_{13} & \mathbf{H}'_{23} & \mathbf{H}_{33} \end{pmatrix}_{N(k+2) \times N(k+2)}.$$

The expressions for \mathbf{H}_{ij} can be obtained using the partial derivatives $\partial^2 \ell_T(\boldsymbol{\theta}_0) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ given in the online supplement B (Supporting Information). Specifically we have

$$\mathbf{H}(\boldsymbol{\theta}_0) = \begin{pmatrix} (\mathbf{G}_0 \odot \mathbf{G}'_0) + \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0 \Sigma_0 \mathbf{G}'_0) + \Delta_{\beta_0} & \mathbb{E}_{\beta_0} & \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0) \\ \mathbb{E}_{\beta_0} & \mathbf{Z}_0 & \mathbf{0} \\ \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0) & \mathbf{0}' & \frac{1}{2} \Sigma_0^{-2} \end{pmatrix}, \quad (17)$$

where Δ_{β_0} , \mathbb{E}_{β_0} , and \mathbf{Z}_0 are diagonal matrices given by

$$\begin{aligned}\Delta_{\beta_0} &= \text{Diag}\left[\sigma_{i0}^{-2} \sum_{r=1}^N \sum_{s=1}^N g_{0,rs} g_{0,ir} \boldsymbol{\beta}'_{r0} \Sigma_{rs} \boldsymbol{\beta}_{s0}, i = 1, 2, \dots, N\right], \\ \mathbb{E}_{\beta_0} &= \text{Diag}\left[\sigma_{i0}^{-2} \sum_{s=1}^N g_{0,is} \boldsymbol{\beta}'_{s0} \Sigma_{is}, i = 1, 2, \dots, N\right], \\ \mathbf{Z}_0 &= \text{Diag}\left[\sigma_{i0}^{-2} \Sigma_{ii}, i = 1, 2, \dots, N\right].\end{aligned}\quad (18)$$

Again by Lemma 5 of the online supplement A (Supporting Information), we have that

$$\left[\frac{1}{\sqrt{T}} \mathbf{s}_T(\boldsymbol{\theta}_0) \right] \xrightarrow{d} N[\mathbf{0}, \mathbf{J}(\boldsymbol{\theta}_0, \gamma)]$$

where

$$\mathbf{J}(\boldsymbol{\theta}_0, \gamma) = \lim_{T \rightarrow \infty} \begin{pmatrix} (\mathbf{G}_0 \odot \mathbf{G}'_0) + \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0 \Sigma_0 \mathbf{G}'_0) + \Delta_{\beta_0} & \mathbb{E}_{\beta_0} \frac{\gamma}{2} \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0) \\ +(\gamma-2) \text{Diag}(\mathbf{G}_0 \odot \mathbf{G}'_0) & \mathbf{Z}_0 \quad \mathbf{0} \\ \mathbb{E}_{\beta_0} & \mathbf{0}' \quad \frac{\gamma}{4} \Sigma_0^{-2} \\ \frac{\gamma}{2} \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0) & \end{pmatrix} \quad (19)$$

and

$$\gamma = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \text{var}(\zeta_{it}^2) = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T [E(\zeta_{it}^4) - 1], \quad (20)$$

with $\zeta_{it} = \varepsilon_{it}/\sigma_{i0}$, for $i = 1, 2, \dots, N$. Hence $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{\boldsymbol{\theta}})$, where $\mathbf{V}_{\boldsymbol{\theta}}$ has the usual sandwich formula

$$\mathbf{V}_{\boldsymbol{\theta}} = \mathbf{H}^{-1}(\boldsymbol{\theta}_0) \mathbf{J}(\boldsymbol{\theta}_0, \gamma) \mathbf{H}^{-1}(\boldsymbol{\theta}_0). \quad (21)$$

In the case where the errors, ε_{it} , are Gaussian, $\gamma = 2$ and, as to be expected, $\mathbf{H}(\boldsymbol{\theta}_0) = \mathbf{J}(\boldsymbol{\theta}_0, 2)$. This is easily verified by referring back to Equation 17, which is equal to $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ defined by Equation 19 for $\gamma = 2$, as required.

Remark 5. When no exogenous regressors are included in the HSAR specification (Equation 1), then the asymptotic variance, $\mathbf{V}_{\boldsymbol{\theta}} = \mathbf{H}^{-1}(\boldsymbol{\theta}_0) \mathbf{J}(\boldsymbol{\theta}_0, \gamma) \mathbf{H}^{-1}(\boldsymbol{\theta}_0)$, simplifies so that

$$\mathbf{H}(\boldsymbol{\theta}_0) = \begin{pmatrix} (\mathbf{G}_0 \odot \mathbf{G}'_0) + \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0 \Sigma_0 \mathbf{G}'_0) & \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0) \\ \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0) & \frac{1}{2} \Sigma_0^{-2} \end{pmatrix}_{2N \times 2N},$$

and

$$\mathbf{J}(\boldsymbol{\theta}_0, \gamma) = \begin{pmatrix} (\mathbf{G}_0 \odot \mathbf{G}'_0) + \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0 \Sigma_0 \mathbf{G}'_0) & \frac{\gamma}{2} \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0) \\ +(\gamma-2) \text{Diag}(\mathbf{G}_0 \odot \mathbf{G}'_0) & \frac{\gamma}{4} \Sigma_0^{-2} \\ \frac{\gamma}{2} \Sigma_0^{-1} \text{Diag}(\mathbf{G}_0) & \end{pmatrix}.$$

Again, under Gaussian errors we have $\mathbf{J}(\boldsymbol{\theta}_0, 2) = \mathbf{H}(\boldsymbol{\theta}_0)$.

The main result of this section is summarized in the following proposition:

Proposition 2. Consider the heterogeneous spatial autoregressive (HSAR) model given by Equation 1. Suppose that Assumptions 1-5 hold. Let N and k be fixed constants, and denote the $N(k+2)$ dimensional (quasi-) maximum likelihood estimator of $\boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \bar{\ell}_T(\boldsymbol{\theta})$, where $\bar{\ell}_T(\boldsymbol{\theta})$ is given by Equation 12. Then, $\hat{\boldsymbol{\theta}}$ is almost surely locally consistent for $\boldsymbol{\theta}_0$ on Θ_c , and has the following asymptotic distribution:

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rightarrow_d N(\mathbf{0}, \mathbf{V}_{\boldsymbol{\theta}}),$$

where $\mathbf{V}_{\boldsymbol{\theta}} = \mathbf{H}^{-1}(\boldsymbol{\theta}_0) \mathbf{J}(\boldsymbol{\theta}_0, \gamma) \mathbf{H}^{-1}(\boldsymbol{\theta}_0)$, and $\mathbf{H}(\boldsymbol{\theta}_0)$ and $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ are defined by Equations 17 and 19, respectively.

Proof. See the online supplement B (Supporting Information).

Focusing on the QML estimators of the spatial lag coefficients, $\hat{\boldsymbol{\psi}}$, we introduce the following partitioning of $\mathbf{H}(\boldsymbol{\theta}_0)$:

$$\mathbf{H}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{H}_{11} & \mathcal{H}_{12} \\ \mathcal{H}'_{12} & \mathcal{H}_{22} \end{pmatrix},$$

where $\mathcal{H}_{12} = (\mathbf{H}_{12}, \mathbf{H}_{13})$ is an $N \times (Nk+N)$ matrix, and since $\mathbf{H}_{23} = \mathbf{H}_{32} = \mathbf{0}$, then $\mathcal{H}_{22} = \text{Diag}(\mathbf{H}_{22}, \mathbf{H}_{33})$, which is an $(Nk+N) \times (Nk+N)$ matrix. Then, the inverse of $\mathbf{H}(\boldsymbol{\theta}_0)$ is given by

$$\mathbf{H}^{-1}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathcal{H}_{11.2}^{-1} & -\mathcal{H}_{11.2}^{-1} \mathcal{H}_{12} \mathcal{H}_{22}^{-1} \\ -\mathcal{H}_{22}^{-1} \mathcal{H}_{21} \mathcal{H}_{11.2}^{-1} & \mathcal{H}_{22}^{-1} + \mathcal{H}_{22}^{-1} \mathcal{H}_{21} \mathcal{H}_{11.2}^{-1} \mathcal{H}_{12} \mathcal{H}_{22}^{-1} \end{pmatrix},$$

and the inverse of the $N \times N$ information matrix $\mathcal{H}_{11.2}$ corresponds to the asymptotic covariance matrix of $\hat{\boldsymbol{\psi}}$. This result is summarized in the following corollary:

Corollary 1. Consider the heterogeneous spatial autoregressive (HSAR) model given by Equation 1. Suppose that Assumptions 1-5 hold. Then for any fixed N and k , the $N \times N$ information matrix

$$\begin{aligned} \mathcal{H}_{11.2} &= (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{Diag} \left[-g_{0,ii}^2 + \sum_{s=1, s \neq i}^N (\sigma_{s0}^2 / \sigma_{i0}^2) g_{0,is}^2, i = 1, 2, \dots, N \right] \\ &\quad + \text{Diag} \left[\sigma_{i0}^{-2} \sum_{r=1}^N \sum_{s=1}^N g_{0,is} g_{0,ir} \boldsymbol{\beta}'_{r0} (\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is}) \boldsymbol{\beta}_{s0}, i = 1, 2, \dots, N \right], \end{aligned} \tag{22}$$

is full rank, where $\mathbf{G}_0 = \mathbf{W}(\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} = (g_{0,ij})$, $\boldsymbol{\Psi}_0 = \text{Diag}(\boldsymbol{\psi}_0)$, $\boldsymbol{\psi}_0 = (\psi_{10}, \psi_{20}, \dots, \psi_{N0})'$, and \mathbf{W} is the spatial weight matrix, and ε_{it} IIDN($0, \sigma_{i0}^2$). Then the maximum likelihood estimator of $\boldsymbol{\psi}_0$, denoted by $\hat{\boldsymbol{\psi}}$ and computed by maximizing (A.21) in online supplement A (Supporting Information), has the following asymptotic distribution:

$$\sqrt{T}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) \rightarrow_d N(\mathbf{0}, \mathbf{V}_{\boldsymbol{\psi}}), \tag{23}$$

where $\mathbf{V}_{\boldsymbol{\psi}} = [\mathcal{H}_{11.2}]^{-1}$.

Proof. See online supplement B (Supporting Information).

Remark 6. When $T \rightarrow \infty$, estimation of the HSAR model (Equation 1) can be conducted for any N . Proposition 2 describes the asymptotic distribution of each individual parameter in vector $\hat{\boldsymbol{\theta}}$. Yu et al. (2008), who studied a similar panel data model to ours, with fixed effects but with homogeneous spatial and slope parameters, considered three cases for N : fixed, asymptotically proportional to T , and asymptotically large relative to T , as $T \rightarrow \infty$. The interest in distinguishing between these cases in their paper arises from the fact that different biases arise in the computation of their proposed QML estimators depending on the relative size of N and T , and due to the homogeneity assumption imposed on their spatial and slope coefficients. On the other hand, the estimated fixed effects under their model specification converge to their respective true values at rate \sqrt{T} irrespective of N —see theorem 4 in Yu, de Jong and Lee (2008).

Remark 7. When the time dimension T is short and fixed, as N rises we are likely to encounter the well-known “incidental parameter” problem, since the number of parameters rise with N and the standard asymptotic results do not hold—an issue originally highlighted by Neyman and Scott (1948). But the incidental parameter problem will not be present if T is large relative to N , irrespective of whether N is fixed or rises with T . This is because the unit-specific parameters are estimated individually consistently, and the impact of initial conditions on the parameter estimates becomes negligible as $T \rightarrow \infty$. But, as discussed below in Section 4, if the object of interest is the mean of the individual coefficients, then we require both N and T to rise together such that $N/T \rightarrow \kappa$, for some strictly positive κ .

Remark 8. In the case where ε_{it} are non-Gaussian but $E(|\varepsilon_{it}|^{4+\epsilon}) < K$ holds for some $\epsilon > 0$, the quasi maximum likelihood estimator, $\hat{\boldsymbol{\psi}}$, continues to be normally distributed but its asymptotic covariance matrix is given by the upper $N \times N$ partition of $\mathbf{H}^{-1}(\boldsymbol{\theta}_0)\mathbf{J}(\boldsymbol{\theta}_0, \gamma)\mathbf{H}^{-1}(\boldsymbol{\theta}_0)$, where $\mathbf{H}(\boldsymbol{\theta}_0)$ and $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ are defined by Equations 17 and 19, respectively. Recall that γ is defined by Equation 20, and under Gaussian errors it takes the value of $\gamma = 2$, so that we have $\mathbf{J}(\boldsymbol{\theta}_0, 2) = \mathbf{H}(\boldsymbol{\theta}_0)$.

Remark 9. The conditions that we have derived for local/global identification and consistency of the QML estimators have parallels in the GMM estimation of spatial models and correspond to the high-level assumptions made under GMM requiring the moment conditions to have a unique solution (which might not be met and is often difficult to check). In practice, when computing QML and GMM estimators it is advisable that a number of different initial parameter vectors, $\boldsymbol{\theta}_{in}$, are considered in the optimization procedure to make sure that the resultant estimates correspond to the global optimum.

3.3.1 | Consistent estimation of $V_{\boldsymbol{\theta}}$

The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ can be consistently estimated using the expressions given by Equations 17 and 19, yielding the following standard and sandwich formulae:

$$\mathbf{V}_{\boldsymbol{\theta}} = \mathbf{H}^{-1}(\boldsymbol{\theta}_0) \text{ and } \mathbf{V}_{\boldsymbol{\theta}} = \mathbf{H}^{-1}(\boldsymbol{\theta}_0)\mathbf{J}(\boldsymbol{\theta}_0, \gamma)\mathbf{H}^{-1}(\boldsymbol{\theta}_0),$$

with the information matrix equality holding in the case of ε_{it} IIDN(0, σ_{i0}^2) and $\gamma = 2$. Consistent estimators of $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ and $\mathbf{H}(\boldsymbol{\theta}_0)$ can be obtained by replacing $\boldsymbol{\theta}_0$ with its QML estimator, $\hat{\boldsymbol{\theta}}$, and estimating γ by $\hat{\gamma} = (NT)^{-1}\sum_{t=1}^T\sum_{i=1}^N(\hat{\varepsilon}_{it}/\hat{\sigma}_i)^4 - 1$, where $\hat{\varepsilon}_{it} = y_{it} - \hat{\psi}_i\sum_{j=1}^N w_{ij}y_{jt} - \hat{\beta}'_i x_{it}$, with $\hat{\sigma}_i$, $\hat{\beta}_i$ and $\hat{\psi}_i$ being the QML estimators of σ_{i0} , β_{i0} and ψ_{i0} , respectively.

Alternatively, one can use the sample counterparts of $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ and $\mathbf{H}(\boldsymbol{\theta}_0)$ and estimate the covariance matrix of the QML estimator consistently by

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}} = \hat{\mathbf{H}}_T^{-1}(\hat{\boldsymbol{\theta}}) \text{ and } \hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}} = \hat{\mathbf{H}}_T^{-1}(\hat{\boldsymbol{\theta}})\hat{\mathbf{J}}_T(\hat{\boldsymbol{\theta}}, \hat{\gamma})\hat{\mathbf{H}}_T^{-1}(\hat{\boldsymbol{\theta}}), \quad (24)$$

where $\hat{\mathbf{J}}_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^T(\partial\ell_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta})(\partial\ell_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta})'$, $\ell_t(\boldsymbol{\theta})$ is defined by (A.20) in the online supplement A (Supporting Information) and $\hat{\mathbf{H}}_T(\boldsymbol{\theta}) = -T^{-1}[\partial^2\ell_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}]$. Consistency of $\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}, \hat{\gamma})$ for $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ follows from consistency of $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$, of $\hat{\gamma}$ for γ and the independence of $\partial\ell_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ over t , as shown in Lemma 5 of the online supplement A (Supporting Information). The first and second derivatives are provided in the online supplement C (Supporting Information).

4 | MEAN GROUP ESTIMATORS

So far we have focused on estimation of the unit-specific parameters and have derived the asymptotic results for a given N and as $T \rightarrow \infty$. But in practice it is often of interest to obtain average estimates across all the units or a subgroup of the units in the panel, assuming that the individual coefficients follow a random coefficient model. In the context of the HSAR model (Equation 1), suppose that $\{\psi_{i0}, \beta_{i0}, i = 1, 2, \dots, N\}$ are randomly distributed around the common means, ψ_0 and β_0 , such that

$$\psi_{i0} = \psi_0 + \eta_{i\psi}, \text{ and } \beta_{i0} = \beta_0 + \eta_{i\beta}, \text{ for } i = 1, 2, \dots, N, \quad (25)$$

where $\boldsymbol{\eta}_i = (\eta_{i\psi}, \eta'_{i\beta})'$ ~ IID $(\mathbf{0}, \Omega_\eta)$, $\Omega_\eta > 0$ is a positive definite matrix, and it is assumed that $E\|\boldsymbol{\eta}_i\|^{2+c} < K$, for some $c > 0$. The parameters of interest are ψ_0 and β_0 which are the population means of spatial lag coefficients and slope parameters of the underlying HSAR model. For consistent estimation of ψ_0 and β_0 we now need both N and T sufficiently large. Large T is required to consistently estimate the unit-specific coefficients, and large N is required for estimation of the common means, ψ_0 and β_0 . It is also possible to apply this procedure to subsets of the units, so long as the number of units in each set is reasonably large.

Consistent estimators of ψ_0 and β_0 are given by the MG estimators:

$$\hat{\psi}_{MG} = N^{-1} \sum_{i=1}^N \hat{\psi}_i, \text{ and } \hat{\beta}_{MG} = N^{-1} \sum_{i=1}^N \hat{\beta}_i, \quad (26)$$

where $\hat{\psi}_i$ and $\hat{\beta}_i$ are the underlying unit-specific estimators. The MG estimator was originally developed by Pesaran and Smith (1995), who showed that in the standard case where $\hat{\psi}_i$ and $\hat{\beta}_i$ are independently distributed, then $\hat{\psi}_{MG}$ and $\hat{\beta}_{MG}$ will be consistent and asymptotically normal. Recently, Chudik and Pesaran (2019) extended this analysis and considered MG estimators based on possibly cross-correlated estimators, and showed that the standard MG estimation continued to apply so long as the underlying unit-specific estimators were weakly cross-correlated.

The main result of this section is summarized in the following proposition:

Proposition 3. Consider the heterogeneous spatial autoregressive (HSAR) model given by Equation 1, where the coefficients $\{\psi_{i0}, \beta_{i0}, i = 1, 2, \dots, N\}$ are distributed randomly around the common means ψ_0 and β_0 following Equation 25. Suppose that Assumptions 1-5 hold. Then for a fixed k , and as $N, T \rightarrow \infty$ jointly such that $\sqrt{N}/T \rightarrow 0$, the mean group estimators, $\hat{\psi}_{MG}$ and $\hat{\beta}_{MG}$, defined by (26) have the following asymptotic distributions

$$\sqrt{N}(\hat{\psi}_{MG} - \psi_0) \xrightarrow{a} N(0, \omega_\psi^2), \text{ and } \sqrt{N}(\hat{\beta}_{MG} - \beta_0) \xrightarrow{a} N(\mathbf{0}, \Omega_\beta),$$

with consistent estimators of ω_ψ^2 and Ω_β given by

$$\hat{\omega}_\psi^2 = \frac{1}{N-1} \sum_{i=1}^N (\hat{\psi}_i - \hat{\psi}_{MG})^2, \text{ and } \hat{\Omega}_\beta = \frac{1}{N-1} \sum_{i=1}^N (\hat{\beta}_i - \hat{\beta}_{MG})(\hat{\beta}_i - \hat{\beta}_{MG})', \quad (27)$$

respectively, where $\hat{\psi}_i$ and $\hat{\beta}_i$ are the underlying unit-specific estimators.

Proof. See the online supplement B (Supporting Information).

Remark 10. The asymptotic distributions of $\hat{\psi}_{MG}$ and $\hat{\beta}_{MG}$ shown in Proposition 3 are carried out assuming β_i and ψ_i are heterogeneous, such that $\text{var}(\beta_i)$ and $\text{var}(\psi_i)$ are strictly nonzero, and do not apply to the case where the slopes are assumed to be homogeneous. For a formal exposition of the properties of the MG estimators under heterogeneous (\sqrt{N} -consistent) and homogeneous slope coefficients (\sqrt{NT} -consistent); see Pesaran and Tosetti (2011).

Remark 11. In principle, it is possible to test the hypothesis of slope homogeneity using Hausman type tests. However, as shown in Pesaran and Yamagata (2008), such tests are likely to lack power. The development of more powerful tests of slope homogeneity in the context of HSAR models is beyond the scope of this paper and would be an interesting topic of further research.

5 | SMALL-SAMPLE PROPERTIES OF THE QMLE

We investigate the small-sample properties of the proposed QML estimator and the associated MG estimator using Monte Carlo simulations. We consider the following data-generating process (DGP):

$$y_{it} = a_i + \psi_i \sum_{j=1}^N w_{ij} y_{jt} + \beta_i x_{it} + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \quad (28)$$

We include one exogenous regressor, x_{it} , with coefficient β_i as well as fixed effects, a_i , in unit-specific regressions. Stacking these regressions we have

$$\mathbf{y}_{\cdot t} = \mathbf{a} + \Psi \mathbf{W} \mathbf{y}_{\cdot t} + \mathbf{B} \mathbf{x}_{\cdot t} + \boldsymbol{\varepsilon}_{\cdot t}, \quad t = 1, 2, \dots, T,$$

where $\mathbf{a} = (a_1, a_2, \dots, a_N)'$, $\Psi = \text{Diag}(\psi)$ and $\psi = (\psi_1, \psi_2, \dots, \psi_N)'$, $\mathbf{W} = (w_{ij})$, $i, j = 1, 2, \dots, N$, $\mathbf{B} = \text{Diag}(\beta)$, with $\beta = (\beta_1, \beta_2, \dots, \beta_N)'$, $\mathbf{x}_{\cdot t} = (x_{1t}, x_{2t}, \dots, x_{Nt})'$, and $\boldsymbol{\varepsilon}_{\cdot t} = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$. Note that since we explicitly account for fixed effects which we separate out from the remaining regressors included in $\mathbf{x}_{\cdot t}$, the unknown parameters are summarized in vector $\boldsymbol{\theta}$, as follows: $\boldsymbol{\theta} = (\mathbf{a}', \psi', \beta', \sigma^2)'$, $\sigma^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)'$. In total there are $4N$ unknown parameters.

We allow for spatial dependence in the regressors, x_{it} , and generate them as

$$\mathbf{x}_{\cdot t} = (\mathbf{I}_N - \Phi \mathbf{W}_x)^{-1} \mathbf{v}_{\cdot t}, \quad (29)$$

where $\Phi = \text{Diag}(\phi_1, \phi_2, \dots, \phi_N)$, and $\mathbf{v}_{\cdot t} = (v_{1t}, v_{2t}, \dots, v_{Nt})'$, with v_{it} IIDN($0, \sigma_v^2$). We set $\phi_i = 0.5$ (representing a moderate degree of spatial dependence), and set $\sigma_v^2 = N/\text{tr}[(\mathbf{I}_N - \Phi \mathbf{W}_x)^{-1} (\mathbf{I}_N - \Phi \mathbf{W}_x)^{-1}]$, which ensures that $N^{-1} \sum_{i=1}^N \text{var}(x_{it}) = 1$. We set $\mathbf{W}_x = \mathbf{W} = (w_{ij})$, $i, j = 1, 2, \dots, N$, and use the 4-connection spatial matrix described below.

We consider Gaussian errors such that $\varepsilon_{it}/\sigma_{i0}$ IIDN(0, 1), and non-Gaussian errors such that $\varepsilon_{it}/\sigma_{i0}$ IID[$\chi^2(2) - 2$]/2, for $i = 1, 2, \dots, N$, and $t = 1, 2, \dots, T$, where $\chi^2(2)$ is a chi-squared variate with 2 degrees of freedom. σ_{i0}^2 are generated as independent draws from $\chi^2(2)/4 + 0.50$, for $i = 1, 2, \dots, N$, and kept fixed across the replications.

For the weight matrix, $\mathbf{W} = (w_{ij})$, first we use contiguity criteria to generate the non-normalized weights matrices, $\mathbf{W}^0 = (w_{ij}^0)$, and then row normalize these to obtain w_{ij} . More specifically, we consider \mathbf{W} matrices with 2, 4, and 10 connections.⁴ Since by construction $\|\mathbf{W}\|_\infty = 1$, then condition 8 is satisfied if $\sup_i |\psi_i| < 1$, and ensures that $\mathbf{I}_N - \Psi \mathbf{W}$ is invertible. We generate the unit-specific coefficients of the HSAR model as a_{i0} IIDN(1, 1), β_{i0} IIDU(0, 1), and ψ_{i0} IIDU(0, 0.8), for $i = 1, 2, \dots, N$.⁵ Given the DGP in Equation 28, values of y_{it} are now generated as $\mathbf{y}_{\cdot t} = (\mathbf{I}_N - \Psi \mathbf{W})^{-1} (\mathbf{a} + \mathbf{B} \mathbf{x}_{\cdot t} + \boldsymbol{\varepsilon}_{\cdot t})$.

Initially, to illustrate that our proposed estimator applies to both cases where N is small and large, we considered the two polar cases of $N = 5$ and $N = 100$, and set $T = 25, 50, 100, 200$, thus including (N, T) combinations such that $N < T$, $N = T$ and $N > T$, respectively. We then considered a more comprehensive set of N values, namely $N = 25, 50, 75, 100$. For each experiment we used $R = 2,000$ replications. Across the replications, $\boldsymbol{\theta}_0$, and the weight matrix, \mathbf{W} , are kept fixed, while the errors and the regressors, $\boldsymbol{\varepsilon}_{\cdot t}$ and $\mathbf{x}_{\cdot t}$ (and hence y_{it}), are regenerated randomly in each replication. Note that, as N increases, supplementary units are added to the original vector $\boldsymbol{\theta}_0$ generated initially for $N = 5$. Due to the problem of simultaneity, the degree of time variations in $y_{it}^* = \sum_{j=1}^N w_{ij} y_{jt}$ for each unit i depends on the choice of \mathbf{W} and the number of cross-section units, N . Naturally, this is reflected in the performance of the estimators and the power properties of the tests based on them.

We report bias and RMSE of the QML estimators for individual cross-section units, as well as their corresponding empirical sizes. In addition, we report power functions for three units with true spatial autoregressive parameters, ψ_{i0} , selected to be low, medium, and large in magnitude. The experiments are carried out for spatial weight matrices, \mathbf{W} , with two, four, and ten connections. The mean of simulated parameter estimates are computed as $\hat{\psi}_{i(R)} = R^{-1} \sum_{r=1}^R \hat{\psi}_{i,r}$, and $\hat{\beta}_{i(R)} = R^{-1} \sum_{r=1}^R \hat{\beta}_{i,r}$, where $\hat{\psi}_{i,r}$ and $\hat{\beta}_{i,r}$ refer to the QML estimates of ψ_i and β_i in the r^{th} replication. The QML estimators are computed using the log-likelihood function (7). We also report small sample results for the MG estimators of ψ_0 and β_0 , defined by Equation 26, using the experiment described in Section 5.2 below.

⁴We generate $\mathbf{W}^0 = (w_{ij}^0)$ such that: (i) $w_{ij}^0 = 1$ if $j = i - 1, i + 1$, and zero otherwise (2 connections), (ii) $w_{ij}^0 = 1$ if $j = i - 2, i - 1, i + 1, i + 2$, and zero otherwise (4 connections), and (iii) $w_{ij}^0 = 1$ if $j = i - 5, i - 4, \dots, i - 1, i + 1, i + 2, \dots, i + 5$, and zero otherwise (10 connections). By construction, the first and last units have fewer neighbors as compared to the other units.

⁵We also carried out experiments without exogenous regressors with $\beta_{i0} = 0$, for all i , corresponding to the simplified version of model (5) discussed in Section 3.2. The results of these experiments are available upon request.

5.1 | Results for individual estimates

Since the results based on the Gaussian and non-Gaussian errors are very close, in what follows we only report the results for the non-Gaussian case where the errors are generated as IID $\chi^2(2)$ random variables, and use the sandwich formula (Equation 24) to compute standard errors. Also, to save space, we focus on results based on the spatial weight matrix, \mathbf{W} , with four connections.⁶ Initially, to highlight the applicability of the proposed estimators to small- as well as large-dimensional HSAR panels, we provide detailed results for the experiments with $N = 5$ and $N = 100$.

5.1.1 | Two polar cases: $N = 5$ and $N = 100$

Table 1 reports the bias, RMSE, empirical size, and power of the individual parameters, ψ_{i0} and β_{i0} , $i = 1, 2, \dots, N$, for the experiments with $N = 5$. The bias of estimating ψ_{i0} tends to be small but negative when $T = 25$, while estimates of β_{i0} show an upward bias when $T = 25$. But the biases of both estimators fall quite rapidly with T , for all i . A similar pattern can be seen in the RMSEs, again declining with T reasonably quickly. Turning to size and power of the tests based on the QML estimates, there is some evidence of over-rejection when T is small ($T = 25$). But the size distortion gets eliminated as T is increased, with the tests having the correct size for values of $T \geq 50$. This pattern is shared by both ψ_{i0} and β_{i0} . Similarly, power is low when $T = 25$ but improves markedly for all 5 units as T is increased.⁷ Overall the small sample results are in line with our theoretical findings, and give satisfactory results for values of $T \geq 50$ —a property which is repeated for other experiments considered in this paper.

For $N = 100$ we report the results only for a selected number of units, namely units with the three smallest and largest population values for ψ_{i0} and a few in between, and the associated β_{i0} values. The small-sample results for these experiments are summarized in Table 2, and are qualitatively similar to those reported in Table 1 for $N = 5$, indicating that the theoretical framework of Section 3 can be applied equally to data sets with small and large numbers of cross-section units.

5.1.2 | RMSE, size, and power for all N and T combinations

We now turn to the rest of the results and consider all combinations of $N \in \{25, 50, 75, 100\}$ and $T \in \{25, 50, 100, 200\}$. To save space we use box plots to summarize the results for RMSE and size, and use empirical rejection frequency plots for power.⁸ All results are shown in the online supplement G (Supporting Information). The RMSE box plots for ψ_{i0} and β_{i0} are given in Figures G1 and G2, respectively. Overall, the RMSE values are small for both parameters and fall with T but are not affected by changes in the cross-section dimension, N , which is in line with the theory developed in Section 3.

The box plots for the size of the tests based on the QML estimates of ψ_{i0} and β_{i0} are given in Figures G3 and G7, respectively. These results are based on the sandwich covariance matrix formula given by Equation 24. As can be seen, in general the tests are correctly sized at 5% for T relatively large, although for small values of T there are some size distortions. Once again, the size estimates are not affected by N , and tend to 5% as T increases, irrespective of the value of N .

To save space we only report the empirical power functions of the tests for three cross-section units with low, medium, and high parameter values. The power plots are computed for different values of ψ_i and β_i defined by $\psi_i = \psi_{i0} + \delta$, and $\beta_i = \beta_{i0} + \delta$, for $i = 1, 2, \dots, N$, where $\delta = -0.800, -0.791, \dots, 0.791, 0.800$. We only consider values of ψ_i that satisfy the condition $|\psi_i| < 1$.⁹

The power results for the spatial parameters, ψ_{i0} , are displayed in Figures G4–G6 in the online supplement G (Supporting Information), which correspond to the low value ($\psi_{i0} = 0.3374$), the medium value ($\psi_{i0} = 0.5059$), and the high value ($\psi_{i0} = 0.7676$), respectively. As to be expected, the power depends on the choice of ψ_{i0} and rises with T , but

⁶Results for Gaussian errors and other choices of spatial weight matrices are available upon request.

⁷Clearly, improvements in power can be achieved by reducing the error variances, σ_{i0}^2 . Some supporting evidence is provided in Tables G1 and G2 in the online supplement G (Supporting Information).

⁸The box plots for bias of the estimators are similar to those of RMSE and are available upon request. The corresponding tables that show bias and RMSE results for the individuals estimates ($\hat{\psi}_{i(R)}$, and $\hat{\beta}_{i(R)}$, $i = 1, 2, \dots, N$) are also available upon request.

⁹The empirical power functions are computed using the sandwich formula for the covariance matrix of the underlying estimators.

TABLE 1 Bias, RMSE, size, and power for parameters of individual units in the HSAR(1) model with one exogenous regressor and non-Gaussian errors for $N = 5$ and $T \in \{25, 50, 100, 200\}$

T Parameter	25		50		100		200	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ψ_{i0}								
$\psi_{1,0} = 0.1261$	-0.0056	0.1891	0.0005	0.1230	-0.0023	0.0851	0.0010	0.0592
$\psi_{2,0} = 0.3883$	-0.0051	0.2495	-0.0058	0.1687	-0.0006	0.1148	-0.0003	0.0803
$\psi_{3,0} = 0.4375$	-0.0115	0.2436	-0.0022	0.1499	0.0034	0.1041	-0.0003	0.0743
$\psi_{4,0} = 0.5059$	0.0050	0.1769	-0.0040	0.1221	-0.0028	0.0820	-0.0010	0.0571
$\psi_{5,0} = 0.7246$	-0.0109	0.2089	-0.0031	0.1502	-0.0009	0.1071	0.0006	0.0721
β_{i0}								
$\beta_{1,0} = 0.9649$	0.0125	0.2236	0.0069	0.1472	0.0024	0.1008	-0.0020	0.0717
$\beta_{2,0} = 0.9572$	0.0100	0.2674	0.0068	0.1833	-0.0022	0.1272	-0.0025	0.0892
$\beta_{3,0} = 0.2785$	0.0078	0.2908	-0.0012	0.1806	-0.0026	0.1252	0.0022	0.0907
$\beta_{4,0} = 0.9134$	-0.0020	0.2195	0.0072	0.1461	0.0012	0.1000	0.0000	0.0684
$\beta_{5,0} = 0.8147$	0.0104	0.2842	0.0108	0.1950	0.0081	0.1341	0.0003	0.0911
T Parameter	25	50	100	200	25	50	100	200
	Size				Power			
ψ_{i0}								
$\psi_{1,0} = 0.1261$	0.1040	0.0675	0.0535	0.0515	0.3410	0.4665	0.7060	0.9065
$\psi_{2,0} = 0.3883$	0.0950	0.0690	0.0560	0.0580	0.2515	0.3525	0.4900	0.7315
$\psi_{3,0} = 0.4375$	0.0935	0.0620	0.0560	0.0510	0.2245	0.3355	0.5115	0.7975
$\psi_{4,0} = 0.5059$	0.0835	0.0740	0.0660	0.0485	0.3430	0.5025	0.7355	0.9345
$\psi_{5,0} = 0.7246$	0.0660	0.0670	0.0645	0.0530	0.2450	0.3610	0.5410	0.7975
β_{i0}								
$\beta_{1,0} = 0.9649$	0.0900	0.0645	0.0525	0.0530	0.2845	0.3825	0.5360	0.8075
$\beta_{2,0} = 0.9572$	0.0930	0.0725	0.0625	0.0570	0.2165	0.2885	0.4380	0.6535
$\beta_{3,0} = 0.2785$	0.0960	0.0710	0.0515	0.0585	0.2585	0.3000	0.4565	0.6375
$\beta_{4,0} = 0.9134$	0.0865	0.0630	0.0565	0.0485	0.3055	0.3845	0.5715	0.8245
$\beta_{5,0} = 0.8147$	0.0890	0.0705	0.0555	0.0510	0.2005	0.2570	0.3700	0.6080

Note. True parameter values are generated as $\psi_{i0} \sim \text{IIDU}(0,0.8)$, $\alpha_{i0} \sim \text{IIDN}(1,1)$, and $\beta_{i0} \sim \text{IIDU}(0,1)$ for $i = 1, 2, \dots, N$. Non-Gaussian errors are generated as $\varepsilon_{i0}/\sigma_{i0} \sim \text{IID}[\chi^2(2) - 2]/2$, with $\sigma_{i0}^2 \sim \text{IID}[\chi^2(2)/4 + 0.5]$ for $i = 1, 2, \dots, N$. The spatial weight matrix $\mathbf{W} = (w_{ij})$ has four connections so that $w_{ij} = 1$ if j is equal to: $i - 2$, $i - 1$, $i + 1$, $i + 2$, and zero otherwise, for $i = 1, 2, \dots, N$. Biases and RMSEs are computed as $R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})$ and $\sqrt{R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})^2}$ for $i = 1, 2, \dots, N$. Empirical size and empirical power are based on the sandwich formula given by Equation 24. The nominal size is set to 5%. Size is computed under $H_0: \psi_i = \psi_{i0}$, using a two-sided alternative, for $i = 1, 2, \dots, N$. Power is computed under $\psi_i = \psi_{i0} + 0.2$, for $i = 1, 2, \dots, N$. The number of replications is set to $R = 2,000$. Estimates are sorted in ascending order according to the true values of the spatial autoregressive parameters. Biases, RMSEs, sizes, and powers for β_i , $i = 1, 2, \dots, N$, are computed similarly, with power computed under $\beta_i = \beta_{i0} + 0.2$.

does not seem to be affected by N . Furthermore, perhaps not surprisingly, empirical power functions for ψ_{i0} become more and more asymmetrical as ψ_{i0} 's move closer and closer to the boundary value of 1. The power functions for the three associated values of β_{i0} are shown in Figures G8–G10 of the online supplement G (Supporting Information) for the low value of β_{i0} ($\beta_{i0} = 0.0344$), the medium value ($\beta_{i0} = 0.4898$), and the high value ($\beta_{i0} = 0.9649$), respectively. Again the empirical power functions are similar across N and improve with T .

TABLE 2 Bias, RMSE, size, and power for parameters of individual units in the HSAR(1) model with one exogenous regressor and non-Gaussian errors for $N = 100$ and $T \in \{25, 50, 100, 200\}$

T Parameter	25		50		100		200	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ψ_{i0}								
$\psi_{1,0} = 0.0244$	-0.0005	0.3152	-0.0049	0.2138	0.0021	0.1415	-0.0001	0.1010
$\psi_{2,0} = 0.0255$	-0.0330	0.5189	0.0034	0.3674	-0.0137	0.2641	-0.0033	0.1794
$\psi_{3,0} = 0.0397$	0.0129	0.3509	-0.0017	0.2448	-0.0014	0.1681	0.0013	0.1183
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\psi_{51,0} = 0.3927$	-0.0027	0.2912	0.0038	0.2056	0.0009	0.1395	0.0005	0.0960
$\psi_{52,0} = 0.3987$	0.0001	0.1994	-0.0031	0.1381	0.0029	0.0921	0.0005	0.0638
$\psi_{53,0} = 0.4004$	-0.0112	0.3063	0.0075	0.2049	0.0033	0.1392	-0.0015	0.0991
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\psi_{98,0} = 0.7695$	-0.0031	0.1621	0.0018	0.1149	0.0055	0.0824	-0.0003	0.0586
$\psi_{99,0} = 0.7705$	-0.0530	0.2903	-0.0126	0.1895	0.0002	0.1401	0.0003	0.1041
$\psi_{100,0} = 0.7904$	-0.0125	0.1716	-0.0094	0.1275	0.0011	0.0897	0.0008	0.0613
β_{i0}								
$\beta_{1,0} = 0.1978$	0.0089	0.2782	0.0017	0.1771	0.0007	0.1192	-0.0073	0.0824
$\beta_{2,0} = 0.7060$	0.0252	0.3699	0.0016	0.2359	-0.0005	0.1608	0.0049	0.1144
$\beta_{3,0} = 0.4173$	0.0107	0.2541	0.0034	0.1733	0.0000	0.1157	0.0028	0.0821
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\beta_{51,0} = 0.9448$	0.0060	0.1924	-0.0024	0.1294	0.0018	0.0896	0.0009	0.0634
$\beta_{52,0} = 0.1190$	0.0046	0.1824	0.0026	0.1259	-0.0005	0.0853	0.0021	0.0578
$\beta_{53,0} = 0.7127$	0.0026	0.2630	-0.0050	0.1654	0.0038	0.1201	0.0012	0.0831
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\beta_{98,0} = 0.1067$	0.0041	0.1688	-0.0031	0.1115	0.0010	0.0762	-0.0002	0.0550
$\beta_{99,0} = 0.4588$	0.0207	0.2643	0.0039	0.1788	0.0033	0.1232	0.0027	0.0888
$\beta_{100,0} = 0.3674$	0.0056	0.1691	0.0032	0.1179	0.0009	0.0830	0.0004	0.0560
T Parameter	25	50	100	200	25	50	100	200
	Size				Power			
ψ_{i0}								
$\psi_{1,0} = 0.0244$	0.0890	0.0810	0.0520	0.0590	0.1820	0.2200	0.3290	0.5480
$\psi_{2,0} = 0.0255$	0.0705	0.0595	0.0555	0.0490	0.0945	0.0895	0.1495	0.2140
$\psi_{3,0} = 0.0397$	0.0905	0.0745	0.0585	0.0575	0.1555	0.1895	0.2805	0.4450
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\psi_{51,0} = 0.3927$	0.0950	0.0645	0.0590	0.0535	0.1785	0.2625	0.3590	0.5810
$\psi_{52,0} = 0.3987$	0.0850	0.0620	0.0625	0.0505	0.3050	0.4390	0.6285	0.8660
$\psi_{53,0} = 0.4004$	0.0885	0.0785	0.0570	0.0585	0.1995	0.2490	0.3745	0.5800
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\psi_{98,0} = 0.7695$	0.0635	0.0630	0.0660	0.0610	0.3340	0.4755	0.6935	0.9145
$\psi_{99,0} = 0.7705$	0.0300	0.0285	0.0370	0.0495	0.1455	0.2045	0.3095	0.5205
$\psi_{100,0} = 0.7904$	0.0545	0.0570	0.0625	0.0505	0.3120	0.4665	0.6455	0.8845
β_{i0}								
$\beta_{1,0} = 0.1978$	0.1160	0.0700	0.0610	0.0505	0.2405	0.3040	0.4380	0.7115

TABLE 2 (Continued)

T Parameter	25	50	100	200	25	50	100	200
	Size				Power			
$\beta_{2,0} = 0.7060$	0.1025	0.0580	0.0505	0.0545	0.1725	0.2095	0.2710	0.4510
$\beta_{3,0} = 0.4173$	0.0950	0.0780	0.0550	0.0595	0.2450	0.3160	0.4655	0.7080
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\beta_{51,0} = 0.9448$	0.0910	0.0685	0.0590	0.0570	0.3260	0.4665	0.6500	0.8880
$\beta_{52,0} = 0.1190$	0.0970	0.0800	0.0505	0.0440	0.3500	0.4840	0.7030	0.9150
$\beta_{53,0} = 0.7127$	0.1075	0.0660	0.0665	0.0515	0.2420	0.3150	0.4410	0.6810
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\beta_{98,0} = 0.1067$	0.0960	0.0660	0.0530	0.0545	0.3950	0.5500	0.7605	0.9475
$\beta_{99,0} = 0.4588$	0.0725	0.0615	0.0545	0.0595	0.2015	0.2775	0.4225	0.6415
$\beta_{100,0} = 0.3674$	0.0935	0.0660	0.0695	0.0540	0.3605	0.5025	0.7255	0.9370

Note. See the notes to Table 1

5.2 | Small-sample properties of the MG estimators

We employ the same DGP, defined by Equation 28, and set $a_{i0} = a_0 + \epsilon_{1i}$, with $a_0 = 1$ and ϵ_{1i} IIDN(0, 1), $\psi_{i0} = \psi_0 + \epsilon_{2i}$, with $\psi_0 = 0.4$ and ϵ_{2i} IIDU(-0.4, 0.4) and $\beta_{i0} = \beta_0 + \epsilon_{3i}$, with $\beta_0 = 0.5$ and ϵ_{3i} IIDU(-0.5, 0.5). Parameters a_0 , ψ_0 and β_0 are fixed, while parameters a_{i0} , ψ_{i0} , and β_{i0} vary across replications, for $i = 1, 2, \dots, N$, in accordance with the random coefficients model. The MG estimators and their standard errors are computed using Equations 26 and 27, and the number of replications is set to $R = 2,000$. The small-sample properties of the mean group estimators of ψ_0 and β_0 are summarized in Table G3 of the online supplement G (Supporting Information). The top panel gives the results for Gaussian errors, and the bottom panel for non-Gaussian errors. As to be expected, the bias and RMSE of the MG estimators decline steadily with both N and T , and it does not matter whether the errors are Gaussian or not. There are some small size distortions when $N = T = 25$, but the size rapidly converges to the nominal value of 5% as N and T are increased. For example, for $T = 25$ the size is always within the simulation standard errors when $N \geq 50$.

6 | HETEROGENEOUS SPATIAL SPILLOVER EFFECTS IN THE US HOUSING MARKET

As an empirical application we estimate HSAR models for quarterly real house price changes in the USA at Metropolitan Statistical Areas (MSAs) over the period 1975:Q1–2014:Q4. Modeling and forecasting of cycles in housing markets are of paramount importance for prospective owners, investors, and real estate market participants such as insurers and mortgage lenders (Agnello, Castro, & Sousa, 2015). Some areas of interest include: (i) land use regulations which affect the elasticity of housing supply and thus the extent to which population growth translates into greater housing price growth [Saiz (2010)]; (ii) Real Estate Investment Trust (REIT) investment, by determining the optimal portfolios of real estate structures across space that have the highest returns while holding risk constant; (iii) equilibrium effects of public policies, such as the knock-on effects on real estate price growth due to local labor market distortions associated with increased import competition [Autor, Dorn, and Hanson (2013)].

Determinants of US house price changes are numerous and well documented in the literature; two prominent fundamentals are growth in real per capita disposable income and population—see, for example, Malpezzi (1999) and Gallin (2006), among others. Factors such as differences in land use regulations, construction costs, and real wages that vary rather slowly over time can be captured by fixed effects and slope heterogeneity. An important aspect of the modeling strategy is to account for the existence of co-movements in house prices within and across MSAs as well. Recently, Bailey et al. (2016) (hereafter BHP) highlighted the importance of distinguishing between types of cross-sectional dependence in the analysis of US house price changes, which if ignored can lead to biased parameter estimates. See, for example, the studies by Swoboda, Nega, and Timm (2015) and Munro (2018). BHP distinguish between spatial dependence that originates from economy-wide common shocks such as changes in interest rates, oil prices, and technology,

and the dependence across MSAs due to local spillover effects arising from differences in house prices, incomes, and demographics across MSAs.¹⁰ Here, we use an extended version of the panel data set employed by BHP and further augmented with population and per capita real income data by Yang (2020) to estimate HSAR models, after filtering out the effects of common factors on house price changes.¹¹ We provide MSA-specific estimates of spillover effects, temporal dynamics, as well as population and income elasticities of house prices and corresponding partial effects over time. Further, we report MG estimates of our individual parameter values both at the national and regional levels. As we shall see, we find considerable heterogeneity across MSAs and regions.

6.1 | Data description and transformations

There exist 381 metropolitan statistical areas (MSAs) according to the February 2013 definition provided by the US Office of Management and Budget (OMB). We consider $N = 377$ of these from the contiguous USA. Accordingly, we compile the following variables that are included in our model over the period 1975:Q2–2014:Q4 ($T = 160$ quarters): Π_{it} is the percent quarterly rate of change of real house prices of MSA i in quarter t , $GPOP_{it}$ is the percent quarterly rate of change of population, and $GINC_{it}$ is the percent quarterly rate of change in real per capita income. Details on data sources and transformations can be found in the online supplement F (Supporting Information).

Our estimation strategy requires real house price changes, Π_{it} , to be cross-sectionally weakly dependent by Assumption 4. We first apply the CD test developed in Pesaran (2004, 2015) to Π_{it} in order to assess the strength of cross-sectional dependence (CSD) in Π_{it} . The CD statistic turns out to be 1621.22, which is substantially higher than the 1.96 critical value at the 5% level. With the null hypothesis of weak CSD soundly rejected, we then estimated the exponent of cross-sectional dependence, α , due to Bailey, Holly and Pesaran (2016), which measures the degree of cross-sectional dependence of house price changes. Values of α close to unity are indicative of strong cross-sectional dependence. We obtained $\hat{\alpha} = 1.00$ (0.03), where the standard error of the estimate is given in brackets. It is clear that real house price changes, Π_{it} , are strongly correlated across MSAs, and before estimating local spillover effects using the HSAR model, we must first purge Π_{it} of the common sources of their dependence, as suggested in BHP.

Accordingly, we deseasonalize and defactor the three variables that we use to estimate the HSAR specifications, which we denote by π_{it} , $gpop_{it}$ and $ginc_{it}$, respectively.¹² The CD statistic for the filtered series, π_{it} , is -3.367 , and is substantially lower than the value of 1621.22 obtained for the unfiltered series, but is nevertheless too large, and could suggest that the filtering has not been effective in removing the strong sources of cross-sectional dependence. There are two reasons that this might not be the case. First, as Juodis and Reese (2019) pointed out, the CD test most likely over-rejects when it is applied to residuals from a regression on *unobserved* common factors. Second, as shown in Pesaran (2015), the implicit null of the CD test depends on the relative expansion rates of N and T , and for $T = O(N^\epsilon)$, $0 \leq \epsilon \leq 1$, such that the exponent of cross-sectional dependence, α , which characterizes the degree of cross-sectional dependence of π_{it} , falls in the range $0 \leq \alpha \leq (2 - \epsilon)/4$. Hence a mere rejection of the CD test does not necessarily mean that the defactored series are not weakly correlated. This viewpoint is supported by the estimates of spatial effects reported in the next subsection, which satisfy the stability conditions associated with weakly cross-correlated processes.

6.2 | Estimation of HSAR model for defactored house price changes

We now consider the following HSAR specification for π_{it} which allows for spatiotemporal effects in house price inflation:

$$\pi_{it} = a_i + \psi_{0i} \sum_{j=1}^N w_{ij} \pi_{jt} + \psi_{1i} \sum_{j=1}^N w_{ij} \pi_{j,t-1} + \lambda_i \pi_{i,t-1} + \beta_i^{pop} gpop_{it} + \beta_i^{inc} ginc_{it} + \varepsilon_{it}, \quad (30)$$

for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$. The model incorporates fixed effects and full heterogeneity in the spatial and temporal autoregressive coefficients of real house price changes ($\psi_{0i}, \psi_{1i}, \lambda_i$), as well as the slope coefficients for the two

¹⁰For a theoretical analysis of the interactions between regional house prices, migration flows and income shocks, see Cun and Pesaran (2018).

¹¹In order to decipher the relative importance of common factors and spillovers in explaining the variation in house prices by MSA, one can extend Pesaran and Chudik (2014) to the case of heterogeneous spillovers. This analysis is beyond the scope of this paper.

¹²Details of the deseasonalizing and defactoring of Π_{it} , $GPOP_{it}$ and $GINC_{it}$ can be found in the online supplement F (Supporting Information).

regressors ($\beta_i^{pop}, \beta_i^{inc}$). Innovations are assumed to be distributed as ε_{it} IID $(0, \sigma_{ie}^2)$.¹³ Equation 30 is in accordance with the theoretical model (Equation 1) analyzed in Sections 2 and 3.¹⁴

6.2.1 | Choice of weight matrix \mathbf{W}

For the weight matrix $\mathbf{W} = (w_{ij})$, we consider the distance based weighting scheme implemented in Yang (2020), which is common in the spatial econometrics literature. More precisely, we compute the geodesic distance between each pair of latitude/longitude coordinates for the MSAs included in our sample using the Haversine formula. These coordinates correspond to the center of the polygon implied by each MSA. Then, we determine a specific radius threshold, d (miles), within which MSAs are considered to be neighbors. In this case, the relevant entries in the un-normalized weight matrix \mathbf{W}^0 are set to unity. The MSAs that fall outside this radius are labeled non-neighbors and their corresponding entries in \mathbf{W}^0 are set to zero. Finally, we row-normalize \mathbf{W}^0 and obtain \mathbf{W} , which is used in Equation 30.

We consider three versions of \mathbf{W} constructed with the radius threshold values of $d = 75, 100$ and 125 miles. We name the adjacency matrices $\mathbf{W}_{75}, \mathbf{W}_{100}$, and \mathbf{W}_{125} , respectively. For brevity of exposition, in what follows we focus on the version of Equation 30 that uses \mathbf{W}_{75} , which gives a reasonably sparse weight matrix with 0.88% nonzero elements. Other types of weighting schemes can also be entertained. For example, one can make use of the inverse of bilateral geodesic distances, d_{ij} , between MSAs to construct \mathbf{W} . Another alternative was considered by BHP, who used two separate adjacency matrices determined by the statistically significant positive and negative pairwise correlations of defactored real house price changes. A further scheme was proposed by Zhou, Tu, Chen, and Wang (2017), who used a sample-based adjacency matrix to approximate the true network structure by focusing on an estimation framework that incorporates just the degree (number of connections) of each unit in the network.

6.2.2 | MSA specific estimates

First we present the estimates of individual contemporaneous and net spatial and temporal effects by MSAs. Note that when using adjacency matrix \mathbf{W}_{75} in Equation 30 there are 39 out of 377 MSAs that are completely isolated (have no neighbors) and are thus excluded from the analysis. This leaves us with a reduced sample of $N = 338$ MSAs. Out of these, 260 estimates (or 77%) of the contemporaneous spatial coefficients ($\hat{\psi}_{0i}$) were positive and statistically significant, with only 19 estimates being significantly negative. However, these positive ripple effects are negated somewhat after one quarter, with the coefficients of the lagged spatial effects ($\hat{\psi}_{1i}$) being largely negative, with 259 being negative and statistically significant, 14 significant and positive.¹⁵ As a result, the net spatial effects, computed as $\hat{\psi}_{0i} + \hat{\psi}_{1i}$, are smaller, with fewer being statistically significant. Figure 1a displays the estimates of net effects. Each net spatial estimate, $\hat{\psi}_{0i} + \hat{\psi}_{1i}$, is matched to its corresponding MSA on the map of the USA. MSAs colored in blue depict positive net spatial lag coefficients, with different shades of blue corresponding to differing ranges within which each $(\hat{\psi}_{0i} + \hat{\psi}_{1i})$ falls: Lighter shades refer to ranges closer to zero, whereas darker shades relate to net spatial lag coefficient estimates closer to the boundary value of unity. Similarly, red areas are associated with negative net spatial lag coefficient estimates, with the lighter shade of red indicating $(\hat{\psi}_{0i} + \hat{\psi}_{1i})$ falling in ranges closer to zero, whereas darker red areas refer to more sizable net spatial coefficient estimates. Similarly, Figure 1b displays the estimates of the temporal coefficients, $\hat{\lambda}_i$, which are generally positive and highly significant.¹⁶

Although, on average, spatial (contemporaneous and lagged) effects across the USA net out at a value of 0.08 (0.013), it is evident from Figure 1a that there are significant differences in these effects across individual MSAs. Indeed,

¹³In performing the data transformations of Section 6.1, we abstract from the sampling uncertainty related to using defactored series when estimating HSAR models. In principle, one could estimate the common and local effects simultaneously, instead of the two-stage procedure being followed. However, such an endeavor is beyond the scope of the present paper.

¹⁴We have considered alternative models to Equation 30—one that allows time lags in the exogenous variables as well, and another assuming no spatially and temporally lagged dependent variable. Overall, the results convey a similar message to that from running regression 30. For brevity of exposition, these results are not included in the paper, but are available upon request.

¹⁵A visual representation of the individual estimates, $\hat{\psi}_{0i}$ and $\hat{\psi}_{1i}$, is given in Figures F1(a) and F1(b) of the online supplement F (Supporting Information).

¹⁶Forty-two MSAs have coefficient estimates that hit the upper or lower bound of 0.994/−0.994 set in our optimization procedure. This occurs when $|\hat{\psi}_{0i}| > 0.994$ or $|\hat{\psi}_{1i}| > 0.994$ or $|\hat{\lambda}_i| > 0.994$. These MSAs are shown as a separate category in Figures 1a and 1b.

263 MSAs have positive net spatial lag coefficients, of which 147 are statistically significant, while 75 MSAs have negative net spatial lag coefficients, only 18 of which are significant. Overall, these net spatial lag coefficients point to the existence of important spillover effects in the US housing market even when the influence of national (common) factors is filtered out. It is easy to show spillover effects in house price changes across MSAs without defactoring, but such evidence suffers from the conjunctions of national and local influences, and can be misleading. The spatial display of the estimates in Figure 1a shows how the strength and sign of local spillover effects changes as we move from less economically developed MSAs towards more vibrant neighboring hubs. A distinction in relative spatial effects is evident between the sparsely populated areas in the middle of the USA (Plains, Rocky Mountains and South West), and the two coastal areas (South East, Mid East and Far West) which have a much higher population density—see Table F1 of the online supplement F (Supporting Information). Next, Figure 1b shows that the temporal dynamics in house price changes are universally positive, with 338 MSAs having positive temporal lag coefficients, of which 328 are statistically significant. In general, these estimates are also reasonably large, considering that defactoring is likely to have removed most of the common dynamics in house price fluctuations. Still, heterogeneity across MSAs is evident, with parts of Mid East, South West and Far West showing stronger temporal feedback effects when compared to other US regions.

Similar differences can also be seen in the estimates of the elasticities of house prices to population and real per capita income, as shown in Figures 2a and 2b. We observe that in 302 MSAs the contemporaneous population or income variables have a positive impact on house price changes, although the population effects tend to be more significant and sizable. Of these, more than 70% tend to coincide with areas also reporting positive estimates for the net spatial lag coefficients. Important examples of such MSAs include Los Angeles and San Francisco, Kansas City, and New York–Newark–Jersey.¹⁷ In contrast, the MSAs with negative estimates spread evenly across the USA and correspond to economically less active areas, such as Cheyenne (Wyoming), Pocatello (Idaho), Pittsfield (Massachusetts), Minneapolis (Minnesota) and Memphis (Arkansas). Interestingly, out of these 18 MSAs around half have in fact experienced relatively muted rise, stagnant or outright decline in population over our sample period, potentially contributing to their negative ($\psi_{0i} + \psi_{1i}$) estimates.

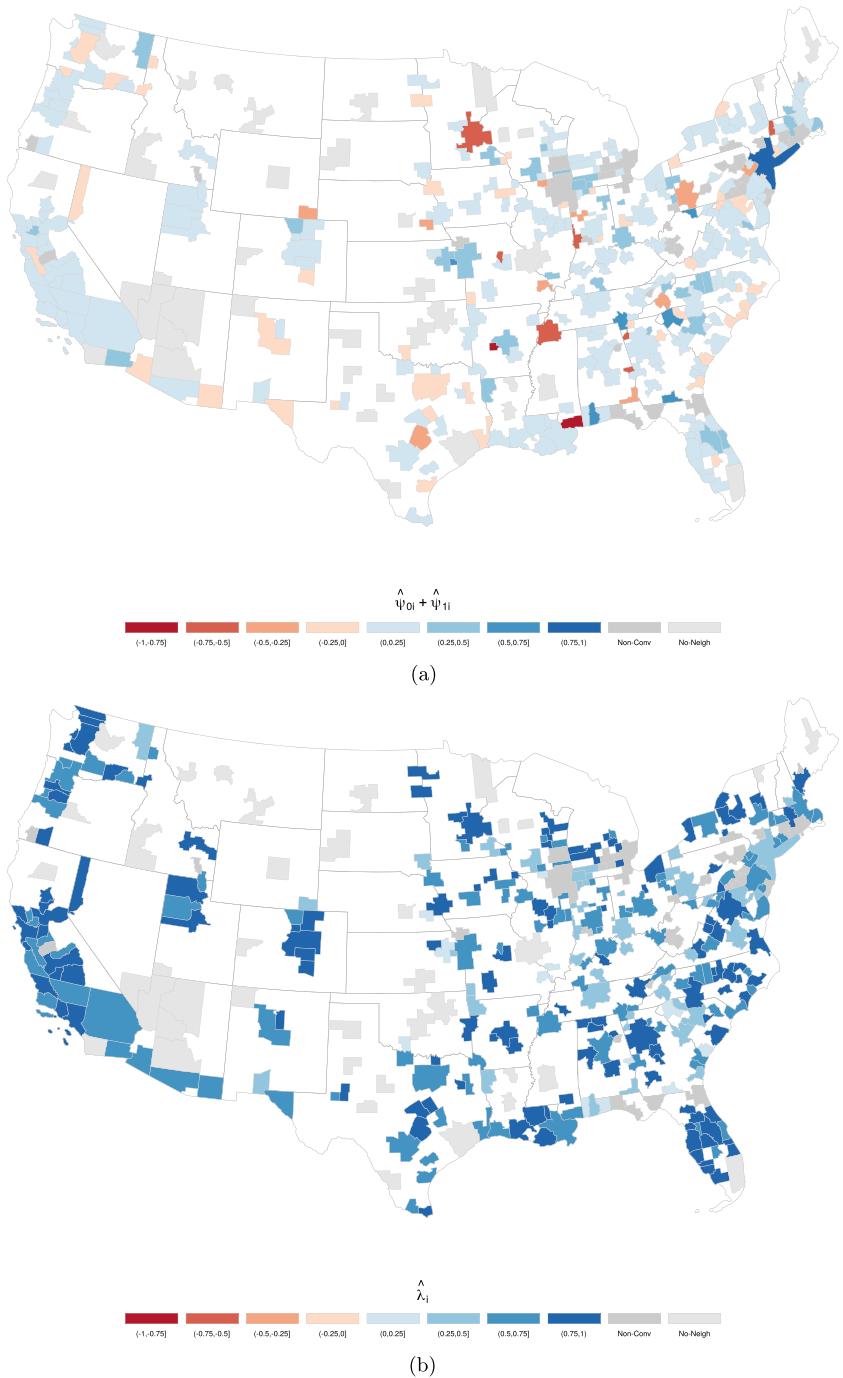
6.2.3 | Regional estimates

The heterogeneity in the estimates we observe across the MSAs continues to be present at the regional level. Table 3 reports the mean group estimates of the parameters by six regions. We started with the standard eight regional classification, but combined New England and the Mid East, and South West and Rocky Mountains to ensure a reasonable number of MSAs (N_r in Table 3) per each region. As can be seen, the MG estimates of the contemporaneous and lagged spatial coefficients are quite close for Great Lakes, South East and Far West, but differ markedly for the other three regions, namely New England & Mid East, Plains, and South West & Rocky Mountains. These differences largely reflect the different degrees of population density across the USA.

Temporal effects are positive and significant across all regions, and cluster within two broad groupings, namely (i) New England & Mid East, South West & Rocky Mountains and Far West, and (ii) Great Lakes, Plains and South East. We notice even larger differences in the MG estimates of the coefficients of population and real income variables across the regions, with much larger estimates for the effects of the population variable on house price changes as compared to the effects of the income variable. For the USA as a whole, the MG estimate of the net spatial and temporal effects amount to 0.088 (0.013) and 0.667 (0.010) respectively, where the net spatial effects are decomposed into the contemporaneous MG estimate 0.603 (0.027) and lagged spatial MG estimate -0.515 (0.020). These estimates point to the existence of non-negligible spatiotemporal effects in the USA even after conditioning on factors that generate strong correlation between MSA-specific house price fluctuations over time and space. These results clearly show the importance of including dynamics in the analysis of spatial effects, which if omitted can lead to exaggeration of these effects. For example, Yang (2020) found (net) spatial effects of around 0.65 when considering a homogeneous and static SAR specification. Finally, the MG estimates of the contemporaneous effects of population and income variables for the USA as a whole are 0.250 (0.029) and 0.050 (0.006), respectively, both being statistically significant, with sizable long-run effects.¹⁸

¹⁷Adding lagged population and real per capita income variables in Equation 30 produces estimates that are generally small and less statistically significant as compared to their contemporaneous effects. Hence these estimates are omitted.

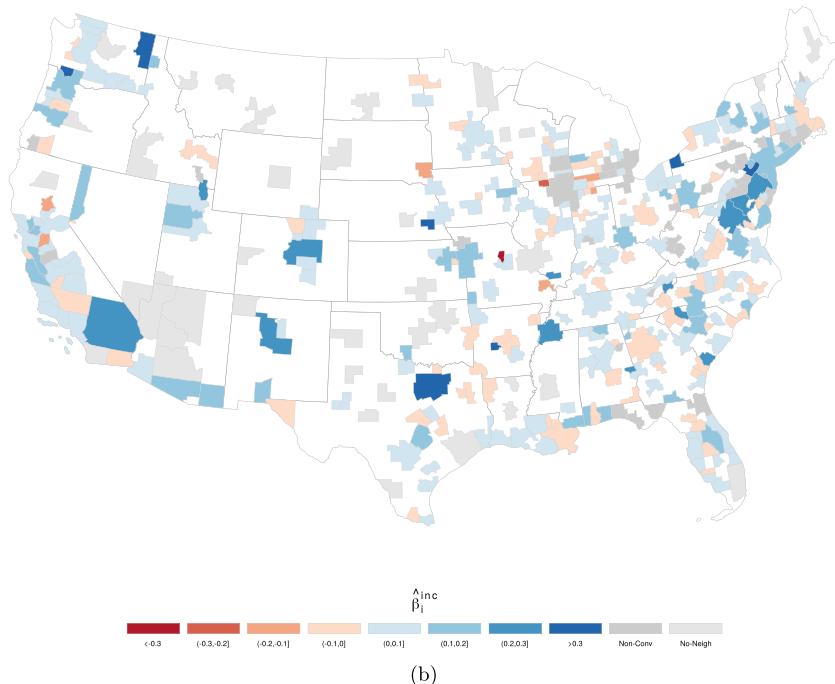
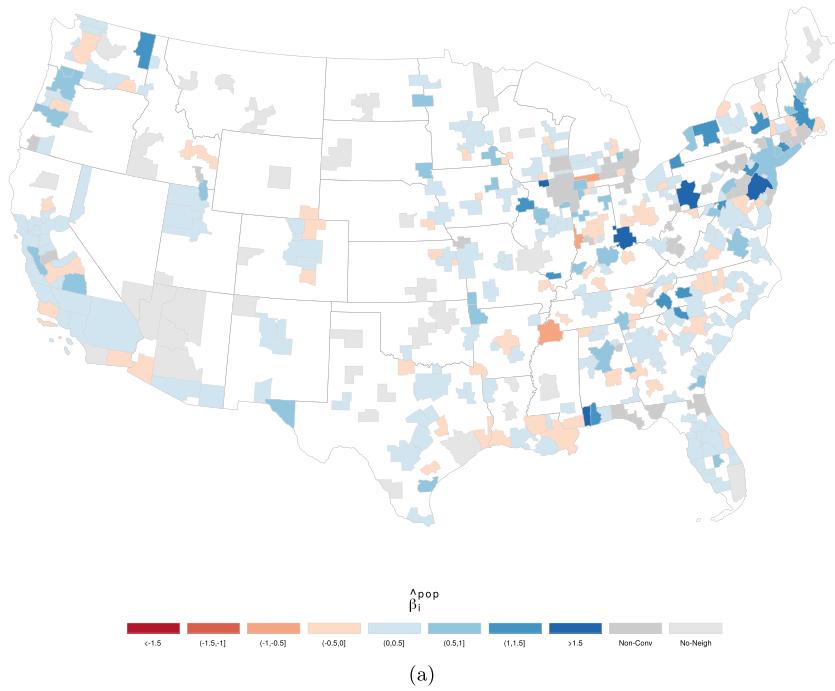
FIGURE 1 (a) net spatial parameter estimates ($\hat{\psi}_{0i} + \hat{\psi}_{1i}$), and (b) autoregressive parameter estimates ($\hat{\lambda}_i$) for Metropolitan Statistical Areas in the USA. Each $\hat{\psi}_{0i} + \hat{\psi}_{1i}$ and $\hat{\lambda}_i$ is mapped to a Metropolitan Statistical Area (MSA) in the USA. A total of 338 MSAs are included. MSAs colored in blue correspond to positive net spatial and temporal parameter estimates, while MSAs colored in red match to negative net spatial and temporal parameter estimates. Darker shades of blue or red indicate more sizable $\hat{\psi}_{0i} + \hat{\psi}_{1i}$ and $\hat{\lambda}_i$, while lighter shades related to $\hat{\psi}_{0i} + \hat{\psi}_{1i}$ and $\hat{\lambda}_i$ closer to zero in absolute terms. Category “Non-conv” includes MSAs whose $\hat{\psi}_{0i}$, $\hat{\psi}_{1i}$, or $\hat{\lambda}_i$ estimates hit the upper/lower bound in the optimization procedure, while category “No-Neigh” includes MSAs that have no neighbors when using \mathbf{W}_{75} [Colour figure can be viewed at wileyonlinelibrary.com]



6.2.4 | Direct and indirect spatial effects at different horizons

The estimated spatiotemporal model (Equation 30) can also be used for impulse response analysis that focuses on the effects of MSA-specific shocks (e_{it}), and/or can be used to investigate the effects of changes in the exogenous variables, namely income and population. To save space we focus on the latter exercise. We closely follow Debarsy, Ertur, and LeSage (2012) and extend their analysis to our heterogeneous parameter specification. Writing Equation 30 in matrix notation and solving for $\boldsymbol{\pi}_{t+h} = (\pi_{1,t+h}, \pi_{2,t+h}, \dots, \pi_{N,t+h})'$, we have

¹⁸We repeated the above empirical analysis using as nonzero elements of the weight matrix \mathbf{W}_0 the inverse of the geodesic pairwise distances between MSAs instead. For brevity of exposition we report the regional MG parameter estimates of model 30 in Table F3 of the online supplement F (Supporting Information).



$$\boldsymbol{\pi}_{t+h} = \boldsymbol{\Phi}^{h+1} \boldsymbol{\pi}_{t-1} + \left(\sum_{s=0}^{h-1} \boldsymbol{\Phi}^s \right) \mathbf{c} + \sum_{s=0}^{h-1} \boldsymbol{\Phi}^s \mathbf{A} \mathbf{x}_{t+h-s} + \sum_{s=0}^{h-1} \boldsymbol{\Phi}^s \mathbf{u}_{t+h-s},$$

where $\mathbf{x}_t = (\mathbf{x}'_{1t}, \mathbf{x}'_{2t}, \dots, \mathbf{x}'_{Nt})'$, $\mathbf{x}_{it} = (x_{i1,t}, x_{i2,t})' = (gpop_{it}, ginc_{it})'$,

$$\begin{aligned} \mathbf{c} &= (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} \mathbf{a}, \quad \boldsymbol{\Phi} = (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} (\boldsymbol{\Psi}_1 \mathbf{W} + \boldsymbol{\Lambda}), \\ \mathbf{A} &= (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} \mathbf{B}, \text{ and } \mathbf{u}_t = (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} \boldsymbol{\varepsilon}_t. \end{aligned}$$

The marginal effect of a unit change in $x_{j\ell,t}$ on $\pi_{i,t+h}$ is given by

FIGURE 2 Contemporaneous elasticities of house price changes to (a) population growth estimates ($\hat{\beta}_i^{pop}$), and (b) real income growth estimates ($\hat{\beta}_i^{inc}$) for Metropolitan Statistical Areas in the USA. Each $\hat{\beta}_i^{pop}$ and $\hat{\beta}_i^{inc}$ is mapped to a Metropolitan Statistical Area (MSA) in the USA. A total of 338 MSAs are included. MSAs colored in blue correspond to positive slope parameter estimates, while MSAs colored in red match to negative slope parameter estimates. Darker shades of blue or red indicate more sizable $\hat{\beta}_i^{pop}$ and $\hat{\beta}_i^{inc}$, while lighter shades related to $\hat{\beta}_i^{pop}$ and $\hat{\beta}_i^{inc}$ closer to zero in absolute terms. Category “Non-conv” includes MSAs whose $\hat{\psi}_{0i}$, $\hat{\psi}_{1i}$ or $\hat{\lambda}_i$ estimates hit the upper/lower bound in the optimization procedure, while category “No-Neigh” includes MSAs that have no neighbors when using \mathbf{W}_{75} [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Mean group estimates (MGE) of spatial and temporal coefficients as well as elasticities of house price changes to population and real income growth by six major US regions, and the USA as a whole

r	Name	N_r	$\hat{\psi}_{MG,r}$	$\hat{\psi}_{MG0,r}$	$\hat{\psi}_{MG1,r}$	$\hat{\lambda}_{MG,r}$	$\hat{\beta}_{MG,r}^{pop}$	$\hat{\beta}_{MG,r}^{inc}$
1 & 2	New England & Mideast	35	0.067	0.499 [*] (0.044)	-0.432 [*] (0.087)	0.645 [‡] (0.069)	0.629 [‡] (0.026)	0.085 [‡] (0.172)
3	Great Lakes	48	0.115 [‡] (0.030)	0.714 [‡] (0.064)	-0.599 [‡] (0.047)	0.629 [‡] (0.026)	0.224 [‡] (0.068)	0.025 [‡] (0.012)
4	Plains	26	0.040	0.525 [‡] (0.058)	-0.484 [‡] (0.083)	0.599 [‡] (0.060)	0.252 [‡] (0.046)	0.039 (0.073)
5	Southeast	106	0.105 [‡] (0.024)	0.669 [‡] (0.044)	-0.564 [‡] (0.032)	0.655 [‡] (0.019)	0.194 [‡] (0.034)	0.038 [‡] (0.008)
6 & 7	Southwest & Rocky Mountain	40	0.017	0.325 [*] (0.021)	-0.308 [*] (0.078)	0.713 [‡] (0.064)	0.162 [‡] (0.020)	0.072 [‡] (0.038)
8	Far West	41	0.126 [‡] (0.017)	0.711 [‡] (0.040)	-0.585 [‡] (0.035)	0.759 [‡] (0.012)	0.183 [‡] (0.047)	0.067 [‡] (0.018)
	USA	296	0.088 [‡] (0.013)	0.603 [‡] (0.027)	-0.515 [‡] (0.020)	0.667 [‡] (0.010)	0.250 [‡] (0.029)	0.050 [‡] (0.006)

Note. * $p < 0.1$; $^{\dagger}p < 0.05$; $^{\ddagger}p < 0.01$. Nonparametric robust standard errors in parentheses (see below). For $r = 1, \dots, 6$, $N_r^{-1} \sum_{i \in I_r} (\hat{\psi}_{0i} + \hat{\psi}_{1i})$ s.e.($\hat{\psi}_{MG,r}$) = $\sqrt{[N_r(N_r-1)]^{-1} \sum_{i \in I_r} [(\hat{\psi}_{0i} + \hat{\psi}_{1i}) - \hat{\psi}_{MG,r}]^2}$, $\hat{\psi}_{MGj,r} = N_r^{-1} \sum_{i \in I_r} \hat{\psi}_{ji}$, and s.e.($\hat{\psi}_{MGj,r}$) = $\sqrt{[N_r(N_r-1)]^{-1} \sum_{i \in I_r} (\hat{\psi}_{ji} - \hat{\psi}_{MGj,r})^2}$, for $j = 0, 1$. I_r is the set of units belonging to region r , $I_r = \{i : i \text{ is in region } r\}$, and N_r is the number of units per region, $N_r = \#(I_r)$. New England (9 MSAs) and Mid East (26 MSAs) as well as South West (26 MSAs) and Rocky Mountains (14 MSAs) have been merged in order to obtain a sufficiently large number of MSAs in the two broader regions. For the USA as a whole: $\hat{\psi}_{MG,US} = N^{-1} \sum_{i=1}^N (\hat{\psi}_{0i} + \hat{\psi}_{1i})$, and s.e.($\hat{\psi}_{MG,US}$) = $\sqrt{[N(N-1)]^{-1} \sum_{i=1}^N [(\hat{\psi}_{0i} + \hat{\psi}_{1i}) - \hat{\psi}_{MG,US}]^2}$. The MGE of coefficient estimates of lagged house price changes ($\hat{\lambda}_{MG,r}$), as well as house price changes to population and real income changes ($\hat{\beta}_{MG,r}^{pop}$ and $\hat{\beta}_{MG,r}^{inc}$) are computed similarly to $\hat{\psi}_{MGj,r}, j = 0, 1$. The computations of all MG estimates exclude the MSAs whose spatial lag coefficients hit the upper/lower bound in the optimization procedure.

$$\frac{\partial \pi_{t+h}}{\partial x_{j\ell,t}} = [\Phi^h(I_N - \Psi_0 W)^{-1} e_j] \beta_{j\ell}, \text{ for } i, j = 1, 2, \dots, N; \ell = 1, 2; h = 0, 1, \dots$$

The average marginal direct and indirect effects of a unit change in $x_{j\ell,t}$ are now given by

$$D_N(h, \ell) = \frac{1}{N} \sum_{i=1}^N \frac{\partial \pi_{i,t+h}}{\partial x_{i\ell,t}}, \text{ and } ID_N(h, \ell) = \frac{1}{N(N-1)} \sum_{i \neq j}^N \frac{\partial \pi_{i,t+h}}{\partial x_{j\ell,t}},$$

for $h = 0, 1, \dots$, and $\ell = 1, 2$. Further, the average indirect effects can be decomposed into average spill-in and spill-out effects. For further details on the derivation and interpretation of these effects see the online supplement E (Supporting Information), and related contributions by LeSage and Chih (2016, 2018b) and LeSage, Chih, and Vance (2019).

The direct and indirect effects (decomposed into spill-in and spill-out effects) at the MSA levels are displayed in Figures F2, F3 and F4 of the online supplement F (Supporting Information). These figures give the estimates on impact, and at horizons 3 and 6 quarters following a 1% increase in population and real income growth. The relative importance of these effects across time and space is evident, also when we compute the equivalent average regional metrics, namely the within-region direct and indirect effects, and the between-region spill-in and spill-out effects as characterized by Equations (E.63), (E.64), (E.65), and (E.66) of the online supplement E (Supporting Information), respectively. Any significant effects on house prices from changes in either the population and income variables of own or neighboring MSAs across regions and horizons are concentrated within-region while the between-region effects are by comparison negligible (universally less than 0.3% of direct effects). These are shown in Table F2 of the online supplement F (Supporting Information). Focusing on the within-region effects, these are more sizable across the board following a

change in population growth as compared to the real income growth, and decay quite rapidly over time. For both population and real income variables, direct effects dominate over indirect effects across the six regions, but the ratio of indirect to direct effects remains remarkably similar for the population and income variables in each region. Nevertheless, there continues to be considerable heterogeneity across the regions. For example, indirect effects tend to be larger in more densely populated regions of New England & Mid East, Great Lakes and Far West, especially on impact (see Table F1 of the online supplement).

7 | CONCLUSION

Standard spatial econometric models assume a single parameter to characterize the intensity or strength of spatial dependence across all units. In the case of pure cross-sectional models or panel data models with a short time dimension, this assumption is inevitable. However, in a data-rich environment where both the time (T) and cross-section (N) dimensions are large, this can be relaxed. This paper investigates a spatial autoregressive panel data model with fully heterogeneous spatial parameters (HSAR) where the spatial dependence can arise directly through contemporaneous dependence of individual units on their neighbors, and indirectly through possible cross-sectional dependence in the regressors.

The asymptotic properties of the quasi maximum likelihood estimator are analyzed assuming a sparse spatial structure, with each individual unit having at least one connection. Conditions under which the QML estimator of spatial parameters are consistent and asymptotically normal are derived. It is also shown that under certain conditions on spatial coefficients and the spatial weights, the asymptotic properties of the individual estimates are not affected by the size of cross-section dimension N . An estimator of the cross-section mean of the individual parameters (MG estimators) is also analyzed, which can be used for comparisons with outcomes from standard homogeneous SAR models. It is shown that MG estimators are consistent and asymptotically normal as N and $T \rightarrow \infty$, jointly, so long as $\sqrt{N}/T \rightarrow 0$, and the spatial dependence is sufficiently weak. Monte Carlo simulation results provided are supportive of the theoretical findings. As an application of the HSAR model we investigate the potential heterogeneity in ripple effects, and the spatio-temporal direct and indirect effects of changes in population and income in the US housing market at the MSA level.

ACKNOWLEDGMENTS

The authors would like to thank Alex Chudik, Bernard Fingleton, Harry Kelejian, Matt Kahn, Ron Smith, Michael Weber, Cynthia Fan Yang, Herman van Dijk (Co-editor), four anonymous referees, and participants at the following conferences: 2015 (EC²) conference: On Theory and Practice of Spatial Econometrics, Edinburgh, UK; 2015 conference on Endogenous Financial Networks and Equilibrium Dynamics: Addressing Challenges of Financial Stability and Monetary Policy, Bank of France, Paris, France; 2016 SEA conference, Rome, Italy; 2018 conference on Housing, Urban Development, and the Macroeconomy, USC, California, USA; 2018 Billfest conference, Melbourne, Australia; 2018 ESAM conference, Auckland, New Zealand; 2019 IAAE conference, Nicosia, Cyprus; and 2019 XIII World SEA-NARSC conference, Pittsburgh, USA for valuable comments and suggestions. Financial support under ESRC Grant ES/I031626/1 is also gratefully acknowledged. All errors and omissions are our own. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this paper.

OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [<http://qed.econ.queensu.ca/jae/datasets/aquaro001/>].

REFERENCES

- Agnello, L., Castro, V., & Sousa, R. M. (2015). Booms, busts, and normal times in the housing market. *Journal of Business and Economic Statistics*, 33(1), 25–45.
- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht, Netherlands: Kluwer Academic.
- Anselin, L. (2001). Spatial econometrics. In Baltagi, B. H. (Ed.), *A companion to theoretical econometrics* (pp. 310–330). Hoboken, NJ: Wiley.
- Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah, A., & Giles, D. E. A. (Eds.), *Handbook of applied economic statistics* (pp. 100–118). New York: Marcel Dekker.

- Arbia, G. (2010). *Spatial econometrics: Statistical foundations and applications to regional convergence*. Berlin, Germany: Springer.
- Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, 103(6), 2121–2168.
- Bailey, N., Holly, S., & Pesaran, M. H. (2016). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*, 31(1), 249–280.
- Baltagi, B., & Levin, D. (1986). Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation and advertising reconsidered. *Review of Economics and Statistics*, 68, 148–155.
- Baltagi, B., Song, S. H., Jung, B. C., & Kon, W. (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics*, 140, 5–51.
- Baltagi, B., Song, S. H., & Kon, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, 117, 123–150.
- Case, A. C. (1991). Spatial patterns in household demand. *Econometrica*, 59, 953–965.
- Chudik, A., & Pesaran, M. H. (2019). Mean group estimator in presence of weakly cross-correlated estimators. *Economics Letters*, 175, 101–105.
- Cliff, A. D., & Ord, J. K. (1973). *Spatial autocorrelation*. London, UK: Pion.
- Cressie, N. (1993). *Statistics for spatial data*. New York, NY: Wiley.
- Cressie, N., & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. New York, NY: Wiley-Blackwell.
- Cun, W., & Pesaran, M. H. (2018). Land use regulations, migration and rising house price dispersion in the U.S. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3162399
- Davidson, J. (2000). *Econometric theory*. Oxford, UK: Blackwell.
- Debarsy, N., Ertur, C., & LeSage, J. P. (2012). Interpreting dynamic space-time panel data models. *Statistical Methodology*, 9, 158–171.
- Dou, B., Parrella, M. L., & Yao, Q. (2016). Generalized Yule–Walker estimation for spatio-temporal models with unknown diagonal coefficients. *Journal of Econometrics*, 194(2), 369–382.
- Elhorst, J. P. (2014). *Spatial econometrics: From cross-sectional data to spatial panels*. Berlin, Germany: Springer.
- Elhorst, J. P., Gross, M., & Tereanu, E. (2018). Spillovers in space and time: Where spatial econometrics and global VAR models meet (ECB Working Paper No. 2134). Frankfurt, Germany: European Central Bank.
- Gallin, J. (2006). The long run relationship between housing prices and income: Evidence from local housing markets. *Real Estate Economics*, 34(3), 417–438.
- Haining, R. (2003). *Spatial data analysis: Theory and practice*. Cambridge, UK: Cambridge University Press.
- Holly, S., Pesaran, M. H., & Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, 158, 160–173.
- Hsiao, C., & Pesaran, M. H. (2008). Random coefficient models(3rd ed.). In Matyas, L., & Sevestre, P. (Eds.), *The econometrics of panel data* (pp. 185–213). Berlin, Germany: Springer.
- Juodis, A., & Reese, S. (2019). The incidental parameters problem in testing for remaining cross-section correlation. Available at: arXiv: 1810.03715[econ.EM].
- Kapoor, M., Kelejian, H. H., & Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140, 97–130.
- Kelejian, H. H., & Prucha, I. R. (1999). A generalised moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40, 509–533.
- Kelejian, H. H., & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157, 53–67.
- Kelejian, H. H., & Robinson, D. (1993). A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. *Papers in Regional Science*, 72, 297–312.
- Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72, 1899–1925.
- Lee, L.-F., & Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154, 165–185.
- LeSage, J. P., & Chih, Y.-Y. (2016). Interpreting heterogeneous coefficient spatial autoregressive panel models. *Economic Letters*, 142, 1–5.
- LeSage, J. P., & Chih, Y.-Y. (2018a). A Bayesian spatial panel model with heterogeneous coefficients. *Regional Science and Urban Economics*, 72, 58–73.
- LeSage, J. P., & Chih, Y.-Y. (2018b). A matrix exponential spatial panel data model with heterogeneous coefficients. *Geographical Analysis*, 50, 422–453.
- LeSage, J. P., Chih, Y.-Y., & Vance, C. (2019). Markov chain Monte Carlo estimation of spatial dynamic panel models for large samples. *Computational Statistics and Data Analysis*, 138, 107–125.
- LeSage, J. P., & Pace, K. (2010). Spatial econometrics. *Web book of regional science* (Vol. 50, pp. 1014–1015). Morgantown, WV: Regional Research Institute.
- Lin, X., & Lee, L.-F. (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*, 157, 34–52.
- Lütkepohl, H. (1996). *Handbook of matrices*. New York, NY: Wiley.
- Malpezzi, S. (1999). A simple error correction model of house prices. *Journal of Housing Economics*, 8(1), 27–62.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3), 531–542.

- Munro, A. (2018). Economic valuation of a rise in ambient radiation risks: Hedonic pricing evidence from the accident in Fukushima. *Journal of Regional Science*, 58, 635–658.
- Neyman, J., & Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16, 1417–1426.
- Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, 27, 286–306.
- Pesaran, M. H. (2004). General diagnostic tests for cross-section dependence in panels (CESifo working paper series No. 1229). Available at <http://ssrn.com/abstract=572504>
- Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, 34(6–10), 1089–1117.
- Pesaran, M. H., & Chudik, A. (2014). Aggregation in large dynamic panels. *Journal of Econometrics*, 178, 273–285.
- Pesaran, M. H., & Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68, 79–113.
- Pesaran, H. P., & Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161, 182–202.
- Pesaran, M. H., & Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, 142(1), 50–93.
- Saiz, A. (2010). The geographic determinants of housing supply. *Quarterly Journal of Economics*, 125(3), 1253–1296.
- Swoboda, A., Nega, T., & Timm, M. (2015). Hedonic analysis over time and space: The case of house price and traffic noise. *Journal of Regional Science*, 55, 644–670.
- Upton, G. J. G., & Fingleton, B. (1985). *Spatial data analysis by example* (Vol. 1). Chichester, UK: Wiley.
- Whittle, P. (1954). On stationary processes on the plane. *Biometrika*, 41, 434–449.
- Yang, C. F. (2020). Common factors and spatial dependence: An application to US house prices. *Econometric Reviews*. Advance online publication. <https://doi.org/10.1080/07474938.2020.1741785>
- Yu, J., de Jong, R. M., & Lee, L.-F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146, 118–134.
- Yu, J., de Jong, R., & Lee, L. F. (2012). Estimation for spatial dynamic panel data with fixed effects: The case of spatial cointegration. *Journal of Econometrics*, 167, 16–37.
- Zhou, J., Tu, Y., Chen, Y., & Wang, H. (2017). Estimating spatial autocorrelation with sampled network data. *Journal of Business and Economic Statistics*, 35(1), 130–138.

How to cite this article: Aquaro M, Bailey N, Pesaran MH. Estimation and inference for spatial models with heterogeneous coefficients: An application to US house prices. *J Appl Econ.* 2021;36:18–44. <https://doi.org/10.1002/jae.2792>