Economic Modelling 76 (2019) 14–30

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Hierarchically spatial autoregressive and moving average error model☆

Qianting Ye[a,b,c,*], Huajie Liang[a,b], Kuan-Pin Lin[d], Zhihe Long[e]

[a] School of Economics and Commerce, South China University of Technology, Guangzhou, 510006, China
[b] Bank of Dongguan Co. Ltd., Dongguan, 523000, China
[c] School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, 85281, USA
[d] Department of Economics, Portland State University, Box 751, Portland, OR, 97207, USA
[e] School of Business Administration, South China University of Technology, Guangzhou, 510641, China

ARTICLE INFO

JEL Classification:
C13
C23
C33

Keywords:
Hierarchically spatial autoregressive and moving average error model
Hierarchical data structure
GMM-FGLS estimation
Monte Carlo simulation

ABSTRACT

This paper considers a hierarchically spatial autoregressive and moving average error (HSEARMA) model. This model captures the spatially autoregressive and moving average error correlation, the county-level random effects, and the district-level random effects nested within each county. We propose optimal generalized method of moments (GMM) estimators for the spatial error correlation coefficient and the error components' variances terms, as well as a feasible generalized least squares (FGLS) estimator for the regression parameter vector. Further, we prove consistency of the GMM estimator and establish the asymptotic distribution of the FGLS estimator. A finite-scale Monte Carlo simulation is conducted to demonstrate the good finite sample performances of our GMM-FGLS estimators.

1. Introduction

Panel (longitudinal) data equipped spatial correlations and hierarchical structures are a common feature in most social and behavioral science studies. For the traditional panel data analysis methods, which reckon without the spatial correlation and hierarchical effects in the regression models, give fairly different estimates compared to the spatial models with the hierarchical data structure (Baltagi et al., 2014), implying that there may be a model-misspecification. Kreft and de Leeuw (1998) and Goldstein (2011) summarized that a hierarchy consists of individuals or units nested within different levels. For instance, when one analyzes the regional variation of house prices, this information might vary across both counties and districts; when one studies the output of firms in a country, firms can be grouped by states, as well as by industry. One can introduce the hierarchical data structure to control for the unobserved group and sub-group effects by using a nested error component model, which can prevent the loss of important information.

In order to recognize the unobserved spatial dependence and hierarchical effects, Corrado and Fingleton (2011, 2012) put forward the preliminary construction of the econometric model of hierarchically spatial data models. As for the spatial lag model, it captures the spatial spillover effects in the dependent variables among neighboring regions (Cliff and Ord, 1973, 1981; Anselin, 1988). For purpose of considering the hierarchical structure of the data, Baltagi et al. (2014) suggested a hierarchically spatial lag (HSLAG) model and used the instrumental variable and two stages least square (IV-2SLS) method for estimation. With regard to the spatial error model, it incorporates spatial correlation in the regression error term to capture the unobserved spillover effect. The spatial error model usually includes autoregressive or/and moving average error component. To simultaneously consider the spatial autoregressive error terms and the hierarchical effects, Ye and Long (2016) proposed the generalized method of moments based feasible generalized least square (GMM-FGLS) estimates for a hierarchically spatial autoregressive error (HSEAR) model. HSEAR model referred

☆ We acknowledge generous support from the Natural Science Foundation of Guangdong Province, China (Grant No.2015A030313216), the Special and Innovative Project of Guangdong Education Bureau, China (Grant No.2014WTSCX001), Funds for Basic Scientific Research Projects in Central Universities, China (XZD19) and China Scholarship Council. All remaining errors are our own.

* Corresponding author. School of Economics and Commerce, South China University of Technology, Guangzhou, 510006, China.
E-mail addresses: cynthia_yip@163.com (Q. Ye), huajie_liang@163.com (H. Liang), link@pdx.edu (K.-P. Lin), l-zh720@163.com (Z. Long).

https://doi.org/10.1016/j.econmod.2018.06.022
Received 27 July 2017; Received in revised form 28 April 2018; Accepted 21 June 2018
Available online 11 October 2018

to the method of Kapoor et al. (2007), while Fingleton et al. (2018) applied this approach to analyze the house price of England.

Actually, the HSEAR model considers the hierarchical effect and a global error shock-effect with feedback from the neighboring locations. However, if an HSEAR model contains moving average error component simultaneously, it can provide a further consideration on a local error shock-effect through the directly interacting locations. Therefore, this paper aims to calibrate a standard spatial econometric model, accommodating the global and local error shock-effect with the hierarchical data structure in nature. The contribution of this paper is two-fold. First, we establish a hierarchically spatial autoregressive and moving average error (HSEARMA) model, in which the autoregressive and moving average error is decomposed into two levels of random effects. Namely, these levels are the county-level random effect and the district-level nested within county-level random effect. Second, we propose a generalized method of moments estimator for the spatial error correlation coefficient and the error components' variances, and an FGLS estimator for the regression parameter vector.

The remainder of the paper proceeds as follows: Section 2 specifies a hierarchically spatial autoregressive and moving average error (HSEARMA) model. Section 3 considers the generalized method of moments (GMM) estimator for the spatial autoregressive and moving average error correlation coefficient and the error components' variances based on 18-moment condition elements and the FGLS estimator for the regression parameter vector. The asymptotic properties of the estimators are studied under regularity conditions. Section 4 conducts a finite-scale Monte Carlo simulation to demonstrate the good finite sample performances of the GMM-FGLS estimators. Concluding remarks are made in Section 5 with suggestions on future research topics. All the derivations, lemmas, and proofs of theorems are collected in the Appendices A, B, C, and the supplementary file.

We close this section by introducing some general notations. Let I_D denote the identity matrix of dimension $D \times D$, l_D denote a $D \times 1$ vector of ones, $J_D = l_D l_D'$ denote a $D \times D$ matrix of ones, $\bar{J}_D = J_D/D$, and $E_D = I_D - \bar{J}_D$. Let $\otimes$ denote the Kronecker product. For a matrix $A = [a_{ij}]$, let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be its smallest and largest eigenvalues, respectively. Let $||\cdot||$ denote the Euclidean norm for a vector. For matrices $A_1, \ldots, A_m$, let $d(A_m)$ be the block diagonal matrix formed by $A_1, \ldots, A_m$. For a sequence of matrices $\{A_S\}_{S=1}^{\infty}$, where $A_S = [a_{ij,S}]$ is of dimension $TS \times TS$ for some integer T, let $\|A_S\|_\infty = \max_i \sum_{j=1}^{TS} |a_{ij,S}|$ be the largest row sum of absolute values of its elements, and $\|A_S\|_1 = \max_j \sum_{i=1}^{TS} |a_{ij,S}|$ be the largest column sum of absolute values of its elements. We say that the row (column) sum of $\{A_S\}_{S=1}^{\infty}$ is uniformly bounded in absolute value if there exists a constant $0 < c < \infty$, independent of S, such that $\|A_S\|_\infty < c$ ($\|A_S\|_1 < c$) for any given S.

2. The model

In this section, we propose a hierarchically spatial autoregressive and moving average error (HSEARMA) model, which incorporates the nested random effects error component structure into a spatial econometric framework. Specifically, we consider the following form, for $i = 1, \ldots, N$, $j = 1, \ldots, m_i$, and $t = 1, \ldots, T$,

$$\begin{cases} y_{ijt} = x_{ijt}\beta + u_{ijt}, \\ u_{ijt} = \rho \sum_{g=1}^{N} \sum_{h=1}^{m_g} w_{ij,gh} u_{ght} + \theta \sum_{g=1}^{N} \sum_{h=1}^{m_g} w_{ij,gh} \epsilon_{ght} + \epsilon_{ijt}, \\ \epsilon_{ijt} = \alpha_i + \mu_{ij} + v_{ijt}, \end{cases} \tag{2.1}$$

where for district j within county i at time period t, y_{ijt} is the dependent variable, x_{ijt} is the $1 \times K$ nonstochastic covariates vector, and u_{ijt} is the unobserved disturbance term. In our model, ϵ_{ijt} is spatially correlated, with ϵ_{ijt} further decomposed into three components. These components are the unobserved county-level permanent unit-specific random effects α_i, district-level random effects μ_{ij} nested within county i, and the county, district, and time specific innovation term v_{ijt}. The regularity conditions for α_i, μ_{ij}, and v_{ijt} are listed in Assumption 1 in Section 3. β denotes the true value of the $K \times 1$ regression parameter vector. $w_{ij,gh}$ is an element in the spatial weights matrix capturing the neighboring relation between district j within county i and district h within county g. Typically, $w_{ij,gh}$ is specified by a first-order rook contiguity criterion, with $w_{ij,gh} = 1$ if district j within county i and district h within county g share a border and $w_{ij,gh} = 0$ otherwise. The weights matrix can be further row-standardized so that each row sums up to one. ρ is the true value of the spatial autoregressive error parameter that captures the strength of globally spatial error correlation. θ is the true value of the spatial moving average error parameter that captures the strength of locally spatial error correlation.

Let $S_N = \sum_{i=1}^{N} m_i$ be the total number of districts. We suppress the subscript "N" in the rest of this paper for notational simplicity. Model (2.1) can be rewritten in vector notation as

$$\begin{cases} y_t = X_t\beta + u_t, \\ u_t = \rho W_S u_t + \theta W_S \epsilon_t + \epsilon_t, \\ \epsilon_t = d(l_{m_i})\alpha + \mu + v_t, \end{cases} \tag{2.2}$$

where $y_t = (y_{11t}, \ldots, y_{1m_1t}, \ldots, y_{N1t}, \ldots, y_{Nm_Nt})'$ is an $S \times 1$ vector, X_t is the $S \times K$ nonstochastic covariates matrix whose first column is a vector of ones, and $u_t = (u_{11t}, \ldots, u_{1m_1t}, \ldots, u_{N1t}, \ldots, u_{Nm_Nt})'$ is the $S \times 1$ disturbance vector. $\epsilon_t = (\epsilon_{11t}, \ldots, \epsilon_{1m_1t}, \ldots, \epsilon_{N1t}, \ldots, \epsilon_{Nm_Nt})'$, $\alpha = (\alpha_1, \ldots, \alpha_N)'$, $\mu = (\mu_{11}, \ldots, \mu_{1m_1}, \ldots, \mu_{N1}, \ldots, \mu_{Nm_N})'$, and $v_t = (v_{11t}, \ldots, v_{1m_1t}, \ldots, v_{N1t}, \ldots, v_{Nm_Nt})'$. $W_S = [W_{ig}], i, g = 1, \ldots, N$, where $W_{ig} = [w_{ij,gh}], j = 1, \ldots, m_i, h = 1, \ldots, m_g$. In other words, W_S consists of N^2 submatrices, and the submatrix W_{ig} is of dimension $m_i \times m_g$.

By further stacking the observations in (2.2) along the time dimension, we obtain

$$\begin{cases} y = X\beta + u, \\ u = \rho W u + \theta W \epsilon + \epsilon, \\ \epsilon = [l_T \otimes d(l_{m_i})]\alpha + (l_T \otimes I_S)\mu + v, \end{cases} \tag{2.3}$$

where $y = (y_1', \ldots, y_T')'$, $X = (X_1', \ldots, X_T')'$, $u = (u_1', \ldots, u_T')'$, $\epsilon = (\epsilon_1', \ldots, \epsilon_T')'$, $v = (v_1', \ldots, v_T')'$, and $W = I_T \otimes W_S$.

By writing $u = (I_{TS} - \rho W)^{-1}(I_{TS} + \theta W)\epsilon$, it can be easily shown that $\mathrm{E}[u] = 0$ under Assumptions 1 (a), (b), and (c) in Section 3, and the variance-covariance matrix of u is

$$\Omega_u = \mathrm{E}[uu'] = (I_{TS} - \rho W)^{-1}(I_{TS} + \theta W)\Omega_\epsilon(I_{TS} + \theta W')(I_{TS} - \rho W')^{-1},$$

where

$$\Omega_\epsilon = \sigma_\alpha^2(J_T \otimes d(J_{m_i})) + (J_T\sigma_\mu^2 + I_T\sigma_v^2) \otimes d(I_{m_i}),$$

and $\sigma_\alpha^2, \sigma_\mu^2$, and σ_v^2 are the variances of α_i, μ_{ij}, and v_{ijt}, respectively. Following Baltagi et al. (2014), we obtain the following spectral decomposition

$$\Omega_\epsilon = \xi_1 Q_1 + \xi_2 Q_2 + (I_T \otimes d(\xi_{3i} I_{m_i}))Q_3,$$

where $\xi_1 = \sigma_v^2$, $\xi_2 = \sigma_\mu^2 T + \sigma_v^2$, $\xi_{3i} = \sigma_\alpha^2 T m_i + \sigma_\mu^2 T + \sigma_v^2$, $Q_1 = E_T \otimes I_S$, $Q_2 = \bar{J}_T \otimes d(E_{m_i})$, and $Q_3 = \bar{J}_T \otimes d(\bar{J}_{m_i})$. This decomposition is convenient in that the matrices Q_1, Q_2, and Q_3 are symmetric, idempotent, orthogonal to each other, and $Q_1 + Q_2 + Q_3 = I_{TS}$. It can be easily shown that $\mathrm{tr}(Q_1) = (T-1)S$, $\mathrm{tr}(Q_2) = S - N$, and $\mathrm{tr}(Q_3) = N$, where $\mathrm{tr}(\cdot)$ is the trace operator. By Lemma 2.2. in Magnus (1982), Ω_ϵ^{-1} is given by

$$\Omega_\epsilon^{-1} = \xi_1^{-1} Q_1 + \xi_2^{-1} Q_2 + (I_T \otimes d(\xi_{3i}^{-1} I_{m_i}))Q_3.$$

3. Estimation

In this paper, we establish asymptotic results when T is fixed and $N, S \to \infty$. In Section 3.1, we consider estimating the model parameters $(\beta', \rho, \theta, \sigma_\alpha^2, \sigma_\mu^2, \sigma_v^2)'$ by a combination of ordinary least squares and GMM methods. In Section 3.2, we consider the optimal GMM estimator for the spatial error correlation coefficient and error components' variances, and an FGLS estimator for the regression parameter vector. We impose the following regularity assumptions.

Assumption 1.

(a) $\alpha_i \sim$ i.i.d. $(0, \sigma_\alpha^2)$ across i, $\mu_{ij} \sim$ i.i.d. $(0, \sigma_\mu^2)$ across i,j, and $v_{ijt} \sim$ i.i.d. $(0, \sigma_v^2)$ across i,j,t;
(b) There exist $\overline{\sigma}_2$ and $\overline{\sigma}_4$ such that $0 < \sigma_\alpha^2, \sigma_\mu^2, \sigma_v^2 < \overline{\sigma}_2 < \infty$, and the fourth moments of $\alpha_i, \mu_{ij}, v_{ijt}$ are bounded above by $\overline{\sigma}_4$;
(c) The processes $\{v_{ijt}\}$, $\{\mu_{it}\}$ and $\{\alpha_t\}$ are independent;
(d) $\max_i m_i$ is bounded by a constant $\overline{m} < \infty$, which does not depend on N and S;
(e) $N/S \to \kappa \in (0,1)$ as $N, S \to \infty$.

Assumption 2.

(a) All the elements of W_S are finite constants, and the diagonal elements of W_S are zero;
(b) the true spatial error correlation coefficients ρ and θ satisfy $|\rho| < 1$ and $|\theta| < 1$, respectively;
(c) the matrix $I_S - \rho W_S$ and $I_S + \theta W_S$ are nonsingular;
(d) the row and column sums of the two sequences of matrices $\{W_S\}_{S=1}^\infty$, $\{(I_S - \rho W_S)^{-1}\}_{S=1}^\infty$ and $\{I_S + \theta W_S\}_{S=1}^\infty$ are uniformly bounded in absolute value.

Assumption 3.

(a) The elements of X are bounded in absolute value, uniform in N and S;
(b) $\lim_{S\to\infty} \frac{1}{ST} X'X$ and $\Sigma_{1,\infty} = \lim_{S\to\infty} \frac{1}{ST} X^{*\prime}(\rho, \theta)\Omega_\epsilon^{-1} X^*(\rho, \theta)$ are finite and nonsingular, where $X^*(\rho, \theta) = (I_{TS} + \theta W)^{-1}(I_{TS} - \rho W)X$;
(c) let $\Lambda_\alpha = \frac{\sigma_\alpha^2 X_\alpha^{*\prime} X_\alpha^*}{TS}$, $\Lambda_\mu = \frac{\sigma_\mu^2 X_\mu^{*\prime} X_\mu^*}{TS}$, and $\Lambda_v = \frac{\sigma_v^2 X_v^{*\prime} X_v^*}{TS}$, assume that $\Lambda_{\alpha,\infty} = \lim_{S\to\infty} \Lambda_\alpha$, $\Lambda_{\mu,\infty} = \lim_{S\to\infty} \Lambda_\mu$, and $\Lambda_{v,\infty} = \lim_{S\to\infty} \Lambda_v$ are finite and nonsingular, where the expressions for X_α^*, X_μ^*, and X_v^* are in (C.1) in Appendix C.

Remark 1. Assumptions 1 (a)-(c) are standard in the literature, while 1 (d) and (e) are new. 1 (d) restricts the number of districts in each county to be bounded, it rules out the case when the number of districts in each county can grow slowly to ∞ as $N, S \to \infty$. 1 (d) implies that the row and column sums of $d(l_{mi})$ are uniformly bounded in absolute value so that Lemma 2 in Appendix B holds. 1 (e) says that N goes to ∞ in a proportional manner as S, and it also does not allow the number of districts in each country to grow to ∞ as $N, S \to \infty$. 1 (e) enables us to control certain moments and apply the central limit Theorem in Pötscher and Prucha (2001). Assumptions 2 and 3 are standard in the literature, and similar assumptions are made in Kapoor et al. (2007).

3.1. OLS and initial GMM estimation

We first show that the OLS estimator for β is consistent in the following Theorem.

Theorem 1. Under Assumptions 1 (a)-(d), 2 (d), 3 (a), and (b), for the OLS estimator $\widetilde{\beta} = (X'X)^{-1}X'y$, we have $\widetilde{\beta} - \beta = o_P(1)$.

Proof. See Appendix C. □

Next, we construct the initial GMM estimator for $(\rho, \theta, \sigma_\alpha^2, \sigma_\mu^2, \sigma_v^2)$. The initial GMM estimator is defined by the sample version of eighteen-moment conditions discussed in the following. Previous works in the literature considered similar ideas in obtaining the GMM estimator. Fingleton and Le Gallo (2007) and Fingleton (2008) proposed a GMM estimator defined by twelve-moment conditions in a panel data SEMA model. In light of the framework of Kapoor et al. (2007), Ye and Long (2016) and Fingleton et al. (2018) introduced a GMM estimator defined by nine-moment conditions in a cross-sectional data SEAR model. To include an autoregressive and moving average process, we need to use an additional spatial lag of the model error structure: $\overline{\overline{\epsilon}} = (I_T \otimes W_S)(I_T \otimes W_S)\epsilon$.

$$\begin{cases} \overline{u} = (I_T \otimes W_S)u, \\ \overline{\epsilon} = (I_T \otimes W_S)\epsilon, \\ \overline{\overline{\epsilon}} = (I_T \otimes W_S)\overline{\epsilon}. \end{cases} \tag{3.1}$$

It follows from (2.3) and (3.1) that we have

$$\begin{cases} u - \rho\overline{u} = \theta\overline{\epsilon} + \epsilon, \\ \overline{u} - \rho\overline{\overline{u}} = \theta\overline{\overline{\epsilon}} + \overline{\epsilon}. \end{cases} \tag{3.2}$$

$\forall i = 1,2,3$, we premultiply (3.2) by Q_i and therefore we obtain the following quadratic forms:

$$\begin{cases} \sum_j [Q_i(u - \rho\overline{u})]_j^2 = \sum_j [Q_i(\theta\overline{\epsilon} + \epsilon)]_j^2, \\ \sum_j [Q_i(\overline{u} - \rho\overline{\overline{u}})]_j^2 = \sum_j [Q_i(\theta\overline{\overline{\epsilon}} + \epsilon)]_j^2, \\ \sum_j [(\overline{u} - \rho\overline{\overline{u}})Q_i(u - \rho\overline{u})]_j = \sum_j [(\theta\overline{\overline{\epsilon}} + \overline{\epsilon})Q_i(\theta\overline{\epsilon} + \epsilon)]_j. \end{cases} \tag{3.3}$$

The above quadratic can be converted to the left side of the expression with parameters, and the right side is the expression without parameters.

$$E[\epsilon' Q_i \epsilon] + 2\theta E[\overline{\epsilon}' Q_i \epsilon] + \theta^2 E[\overline{\epsilon}' Q_i \overline{\epsilon}] + 2\rho E[\overline{u}' Q_i u] - \rho^2 E[\overline{u}' Q_i \overline{u}] = E[u' Q_i u],$$

$$E[\overline{\epsilon}' Q_i \overline{\epsilon}] + 2\theta E[\overline{\overline{\epsilon}}' Q_i \overline{\epsilon}] + \theta^2 E[\overline{\overline{\epsilon}}' Q_i \overline{\overline{\epsilon}}] + 2\rho E[\overline{\overline{u}}' Q_i \overline{u}] - \rho^2 E[\overline{\overline{u}}^T Q_i \overline{\overline{u}}] = E[\overline{u}^T Q_i \overline{u}],$$

$$E[\overline{\epsilon}' Q_i \epsilon] + \theta E[\overline{\epsilon}' Q_i \overline{\epsilon}] + \theta E[\overline{\overline{\epsilon}}' Q_i \epsilon] + \theta^2 E[\overline{\overline{\epsilon}}' Q_i \overline{\epsilon}] + \rho E[\overline{u}' Q_i \overline{u}] + \rho E[\overline{\overline{u}}' Q_i u] - \rho^2 E[\overline{\overline{u}}' Q_i \overline{u}] = E[\overline{u}' Q_i u].$$

Under Assumption 1, for fixed $T \geq 2$, we consider the following moment condition elements in Table 1:

where $\xi_3 = \sigma_\alpha^2 \frac{TS}{N} + \xi_2$, $J^* = \frac{N}{S} diag(J_{m_i})$, $W_S^* = diag(\overline{J}_{m_i})W_S$, $R_1 = W_S W_S$, $R_2 = W_S^T - W_S^{*T}$, $R_3 = W_S^{*T} W_S$. For the detailed derivation of moment conditions, please see Appendix A.

Substituting (3.2) and the moment conditions into (3.3), a system of nine equation involves θ, ξ_1, ξ_2 and ξ_3 can be expressed as

$$\Gamma\varphi - \gamma = 0, \tag{3.4}$$

where $\varphi = [\rho, \rho^2, \xi_1, \xi_2, \xi_3, \theta\xi_1, \theta\xi_2, \theta\xi_3, \theta^2\xi_1, \theta^2\xi_2, \theta^2\xi_3]'$. As space is limited, the expressions of Γ and γ can be found in Appendix A.

Let $\widetilde{\beta}$ be an initial consistent estimator of β, and let $\widetilde{u} = y - X\widetilde{\beta}$, $\widetilde{\overline{u}} = W\widetilde{u}$. Let $\psi = (\rho, \theta, \xi_1, \xi_2, \xi_3)$, $\underline{\psi} = (\underline{\rho}, \underline{\theta}, \underline{\xi}_1, \underline{\xi}_2, \underline{\xi}_3)$, where $\underline{\rho}, \underline{\theta}, \underline{\xi}_1, \underline{\xi}_2, \underline{\xi}_3$ denote the lower bound of generic parameters $\rho, \theta, \xi_1, \xi_2, \xi_3$. Let Θ be the parameter space. Then, for the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, define $\varsigma : \Omega \times \Theta \to \mathbb{R}^9$ by

$$\varsigma(\underline{\psi}) = G\underline{\varphi} - g, \tag{3.5}$$

where $\varsigma(\underline{\psi})$ is a vector of residual, the elements in G and g are the same as the corresponding elements in Γ and γ except that the population moments are replaced by pseudo sample moments.

Theorem 2. Let $\widetilde{\beta}$ be a consistent estimator of β, and define

$$\widetilde{\psi} = \arg\min_{\underline{\psi}} \varsigma'(\underline{\psi})\varsigma(\underline{\psi}). \tag{3.6}$$

Under Assumptions 1, 2 (d), and 3 (a), $\widetilde{\psi} - \psi = o_P(1)$ and the consistency of $(\widetilde{\rho}, \widetilde{\theta}, \widetilde{\sigma}_\alpha^2, \widetilde{\sigma}_\mu^2, \widetilde{\sigma}_v^2)$ for $(\rho, \theta, \sigma_\alpha^2, \sigma_\mu^2, \sigma_v^2)$ follows immediately from the continuous mapping Theorem.

Table 1
18-Moment condition elements.

i	1	2	3
η	$\frac{1}{(T-1)S}$	$\frac{1}{S-N}$	$\frac{1}{N}$
$\eta E[\epsilon' Q_i \epsilon]$	ξ_1	ξ_2	ξ_3
$\eta E[\bar{\epsilon}' Q_i \bar{\epsilon}]$	$\frac{\xi_1}{S} tr[W_S' W_S]$	$\frac{\xi_3-\xi_2}{S-N} tr[R_2 W_S J^*] + \frac{\xi_2}{S-N} tr[W_S R_2]$	$\frac{\xi_3-\xi_2}{N} tr[R_3 J^*] + \frac{\xi_2}{N} tr[R_3]$
$\eta E[\bar{\epsilon}' Q_i \epsilon]$	0	$\frac{\xi_2}{S-N} tr[R_2]$	$\frac{\xi_3-\xi_2}{N} tr[W_S' J^*] + \frac{\xi_2}{N} tr[W_S^{*T}]$
$\eta E[\bar{\bar{\epsilon}}' Q_i \epsilon]$	$\frac{\xi_1}{S} tr[R_1]$	$\frac{\xi_2}{S-N} tr[W_S' R_2]$	$\frac{\xi_3-\xi_2}{N} tr[W_S' W_S' J^*] + \frac{\xi_2}{N} tr[W_S' W_S^{*T}]$
$\eta E[\bar{\bar{\epsilon}}' Q_i \bar{\epsilon}]$	$\frac{\xi_1}{S} tr[W_S' R_1]$	$\frac{\xi_3-\xi_2}{S-N} tr[W_S' R_2 W_S J^*] + \frac{\xi_2}{S-N} tr[W_S' R_2 W_S]$	$\frac{\xi_3-\xi_2}{N} tr[W_S' R_3 J^*] + \frac{\xi_2}{N} tr[W_S' R_3]$
$\eta E[\bar{\bar{\epsilon}}' Q_i \bar{\bar{\epsilon}}]$	$\frac{\xi_1}{S} tr[R_1' R_1]$	$\frac{\xi_3-\xi_2}{S-N} tr[W_S' R_2 R_1 J^*] + \frac{\xi_2}{S-N} tr[W_S' R_2 R_1]$	$\frac{\xi_3-\xi_2}{N} tr[W_S' R_3 W_S J^*] + \frac{\xi_2}{N} tr[W_S' R_3 W_S]$

Proof. See Appendix C.[1] □

3.2. Optimal GMM and FGLS estimation

In the generalized method of moments literature, it is well known that in order to achieve asymptotic efficiency, it is optimal to use the inverse of the sample moments' covariance matrix, evaluated at the true parameter, as the weights matrix (see Hansen (1982), Kapoor et al. (2007)). We provide the expression for the population optimal weights matrix Ξ in the following, and the derivation can be found in Appendix A and the supplementary file.

We further make the following Assumption.

Assumption 4. (a) Ξ is nonsingular; (b) There exist $\underline{\lambda}, \overline{\lambda}$ such that $0 < \underline{\lambda} \le \lambda_{\min}(\Gamma'\Gamma)$, $0 < \underline{\lambda} \le \lambda_{\min}(\Gamma'\Xi^{-1}\Gamma)$, and $0 < \underline{\lambda} \le \lambda_{\min}(\Xi^{-1}) \le \lambda_{\max}(\Xi^{-1}) \le \overline{\lambda} < \infty$, where $\underline{\lambda}, \overline{\lambda}$ might depend on $\rho, \theta, \sigma_\alpha^2, \sigma_\mu^2, \sigma_\nu^2$, but do not depend on N, S.

Remark 2. Assumption 4 is similar to Assumption 5 in Kapoor et al. (2007). 4 (a) enables us to apply the continuous mapping Theorem to $\widetilde{\Xi} - \Xi = o_P(1)$ and obtain $\widetilde{\Xi}^{-1} - \Xi^{-1} = o_P(1)$, and 4 (b) guarantees the identification uniqueness of ρ and θ.

Let $\widetilde{\Xi}$ be defined by plugging the initial GMM estimator $\widetilde{\psi}$ into the expression of Ξ, under Assumptions 1 (b), (d), (e), and 2 (d), $\widetilde{\Xi} - \Xi = o_P(1)$. Moreover, under Assumption 4 (a), we have $\widetilde{\Xi}^{-1} - \Xi^{-1} = o_P(1)$. Now, we define the optimal GMM estimator of ψ as

$$\widehat{\psi} = \arg\min_{\psi} \varsigma'(\underline{\psi})\, \widetilde{\Xi}^{-1} \varsigma(\underline{\psi}), \tag{3.7}$$

Theorem 3. Given consistent estimators $\widetilde{\beta}$ and $\widetilde{\Xi}$ of β and Ξ, under Assumptions 1, 2 (d), 3 (a), and 4,

$$\widehat{\psi} - \psi = o_P(1),$$

and the consistency of $(\widehat{\rho}, \widehat{\theta}, \widehat{\sigma}_\alpha^2, \widehat{\sigma}_\mu^2, \widehat{\sigma}_\nu^2)$ for $(\rho, \theta, \sigma_\alpha^2, \sigma_\mu^2, \sigma_\nu^2)$ follows immediately from the continuous mapping Theorem.[2]

Proof. See Appendix C. □

To obtain an efficient estimator of β, we can define the generalized least square (GLS) estimator as

$$\begin{aligned}\widehat{\beta}_{\mathrm{GLS}} &= (X'\Omega_u^{-1}X)^{-1}(X'\Omega_u^{-1}y) \\ &= [X^*(\rho,\theta)'\Omega_\epsilon^{-1}X^*(\rho,\theta)]^{-1}[X^*(\rho,\theta)'\Omega_\epsilon^{-1}y^*(\rho,\theta)],\end{aligned} \tag{3.8}$$

where $y^*(\rho,\theta) = (I_{TS} + \theta W)^{-1}(I_{TS} - \rho W)y$ and $X^*(\rho,\theta) = (I_{TS} + \theta W)^{-1}(I_{TS} - \rho W)X$. The variables $y^*(\rho,\theta)$ and $X^*(\rho,\theta)$ can be viewed as the result of a spatial Cochrane-Orcutt type transformation of the original model. The estimator $\widehat{\beta}_{\mathrm{GLS}}$ is not feasible since Ω_u (or Ω_ϵ) depends on the unknown true quantities $\rho, \theta, \xi_1, \xi_2$, and ξ_3.

Let $\widehat{\rho}$, $\widehat{\theta}$, $\widehat{\xi}_1$, $\widehat{\xi}_2$, and $\widehat{\xi}_3$ be the optimal GMM estimators from Theorem 3, and let $\widehat{\Omega}_\epsilon$ be the plug-in estimator of Ω_ϵ. The corresponding FGLS estimator of β, denoted as $\widehat{\beta}_{\mathrm{FGLS}}$, is then obtained by replacing ρ, θ, ξ_1, ξ_2, and ξ_3 by the corresponding estimators in the expression of the GLS estimator. Specifically,

$$\widehat{\beta}_{\mathrm{FGLS}} = [X^*(\widehat{\rho},\widehat{\theta})'\widehat{\Omega}_\epsilon^{-1}X^*(\widehat{\rho},\widehat{\theta})]^{-1}[X^*(\widehat{\rho},\widehat{\theta})'\widehat{\Omega}_\epsilon^{-1}y^*(\widehat{\rho},\widehat{\theta})]. \tag{3.9}$$

Theorem 4. Under Assumptions 1, 2 (d), 3, and 4, the following hold:

(a) $(TS)^{\frac{1}{2}}(\widehat{\beta}_{\mathrm{GLS}} - \beta) \xrightarrow{d} N(0, \Lambda_\infty^*)$, where $\Lambda_\infty^* = \Sigma_{1,\infty}^{-1}\Lambda_\infty(\Sigma_{1,\infty}^{-1})'$, $\Lambda_\infty = \Lambda_{\alpha,\infty} + \Lambda_{\mu,\infty} + \Lambda_{\nu,\infty}$, and X_α^*, X_μ^*, and X_ν^* are defined in (C.1) in Appendix C;
(b) $(TS)^{\frac{1}{2}}(\widehat{\beta}_{\mathrm{FGLS}} - \widehat{\beta}_{\mathrm{GLS}}) = o_P(1)$;
(c) $\widehat{\Lambda}^* - \Lambda_\infty^* = o_P(1)$, where $\widehat{\Lambda}^* = \widehat{\Sigma}_1^{-1}\widehat{\Lambda}(\widehat{\Sigma}_1^{-1})'$, $\widehat{\Sigma}_1$ and $\widehat{\Lambda}$ are plug-in estimators of Σ_1 and Λ, respectively.

Proof. See Appendix C. □

4. Monte Carlo simulation

This section conducts a finite-scale Monte Carlo simulation to evaluate the estimates' finite sample performances of HSEARMA model.

[1] One can also define another initial GMM estimator by using Kapoor et al. (2007)'s idea to partition the matrices Γ, G and the vectors γ, g. In this way, one can estimate (ρ, θ, ξ_1) and (ξ_2, ξ_3) separately. We point out this possibility and do not pursue it further in this paper.

[2] Let Γ^* and γ^* be identical to Γ and γ except that the expectation operator is dropped. Under Assumptions 1, 2 (d) and Lemma 2, $\Gamma = O(1)$, $\gamma = O(1)$, and $(\Gamma^*, \gamma^*) - (\Gamma, \gamma) = o_P(1)$, which ensure the identifiably uniqueness of the parameters $(\rho, \theta, \sigma_\alpha^2, \sigma_\mu^2, \sigma_\nu^2)$ and the elements of Ξ^{-1} are $O(1)$.

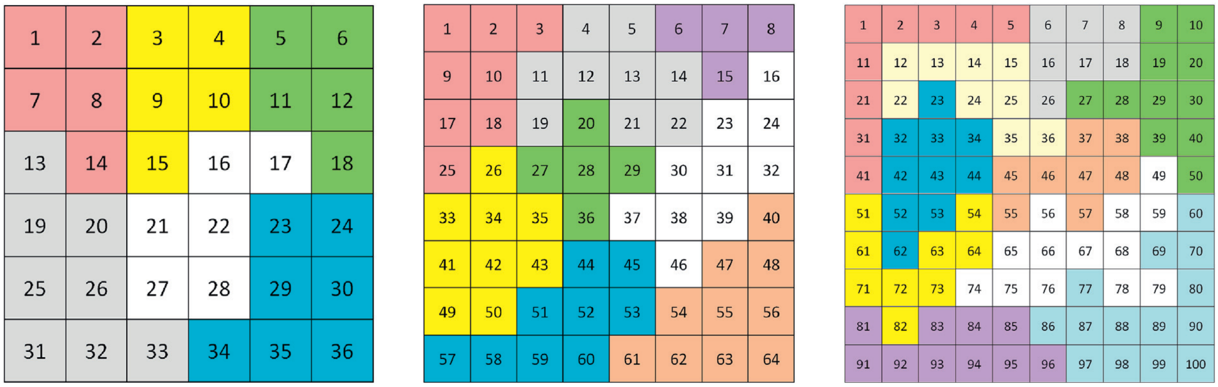

Fig. 1. Geography designs for S = 36, 64, 100.

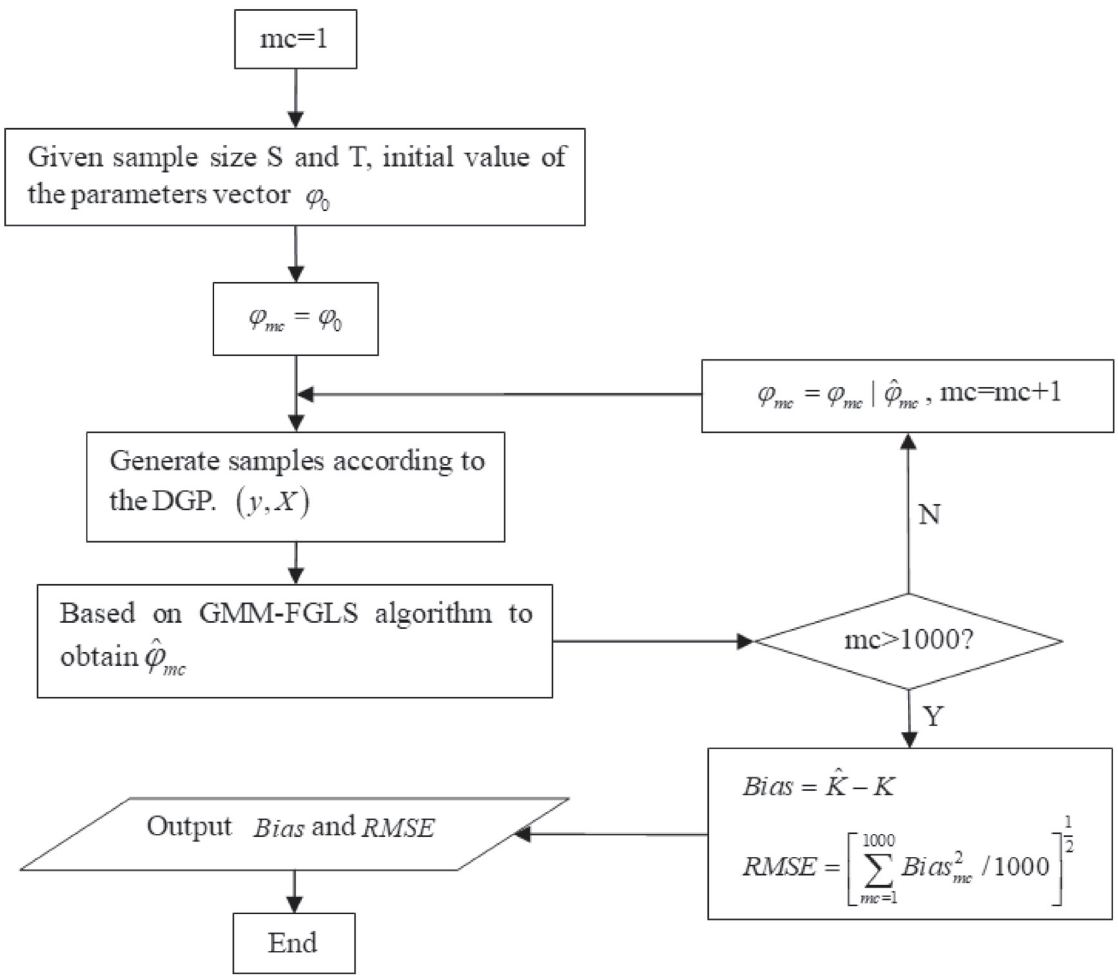

Fig. 2. Monte Carlo simulation algorithm flow chart.

4.1. Simulation design

The key points of Monte Carlo simulation design are as follows:

(1) Initial Value and Spatial Weights Matrix

Throughout the simulation, we set $(\beta_0, \beta_1) = (1,1)$, $\sigma_v^2 = 1$. For σ_α^2 and σ_μ^2, we define $r_1 = \frac{\xi_1}{\xi_3}$, $r_2 = \frac{\xi_1}{\xi_2}$, and set (r_1, r_2) to be (0.2, 0.5) .[3] The accuracy of the estimates replicate the initial spatial correlation coefficients, ρ_0 and θ_0 vary in {-0.9,-0.75,-0.5,-0.25,0,0.25,0.5,0.75,0.9}, for each of the 1000 times Monte Carlo simulations. The spatial weights matrix, W_S, is constructed according to the geographic designs in Fig. 1

[3] For the sake of compactness, we provide simulation results only for the case when $(r_1, r_2) = (0.2, 0.5)$. We also conducted simulations for the following parameter specifications: $r_1 = \{0.1, 0.2\}$, when $r_2 = 0.25$; $r_1 = \{0.1, 0.2, 0.3, 0.4\}$, when $r_2 = 0.5$; $r_1 = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$, when $r_2 = 0.75$. The results are similar and available upon request.

Table 2
Biases and RMSEs of the estimators for HSEARMA model ($\theta_0 = 0.5, T = 5$).

ρ_0	Bias							RMSE						
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\rho}$	$\hat{\theta}$	$\hat{\sigma}_\alpha$	$\hat{\sigma}_\mu$	$\hat{\sigma}_\nu$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\rho}$	$\hat{\theta}$	$\hat{\sigma}_\alpha$	$\hat{\sigma}_\mu$	$\hat{\sigma}_\nu$
(36.5)														
−0.9	−0.002	0.027	0.221	−0.397	0.013	−0.012	0.092	0.050	0.509	0.323	0.550	0.170	0.142	0.162
−0.75	−0.001	0.009	0.470	−0.558	0.008	−0.006	0.065	0.049	0.509	0.636	0.736	0.172	0.127	0.116
−0.5	0.000	0.033	0.555	−0.568	−0.041	−0.018	−0.011	0.048	1.105	0.799	0.801	0.145	0.124	0.066
−0.25	0.000	−0.003	0.153	−0.211	−0.044	−0.027	−0.027	0.044	0.492	0.518	0.492	0.148	0.123	0.071
0	−0.001	−0.005	−0.059	−0.047	−0.033	−0.024	−0.012	0.042	0.503	0.378	0.308	0.150	0.128	0.076
0.25	0.000	−0.017	−0.095	−0.021	−0.021	−0.028	0.003	0.038	0.522	0.345	0.262	0.155	0.128	0.087
0.5	0.000	0.028	−0.095	−0.012	−0.014	−0.025	0.014	0.036	0.644	0.281	0.220	0.167	0.128	0.098
0.75	0.000	−0.013	−0.073	0.004	−0.003	−0.017	0.028	0.032	1.107	0.183	0.167	0.171	0.128	0.113
0.9	0.001	0.006	−0.023	0.004	−0.027	−0.013	0.011	0.031	2.800	0.066	0.134	0.159	0.128	0.087
(64,5)														
−0.9	0.000	0.006	0.145	−0.254	−0.026	0.014	0.049	0.036	0.381	0.243	0.405	0.134	0.096	0.105
−0.75	0.001	−0.014	0.412	−0.482	−0.019	0.017	0.057	0.034	0.361	0.587	0.678	0.139	0.103	0.099
−0.5	0.000	−0.001	0.464	−0.472	−0.055	0.006	−0.008	0.035	0.388	0.738	0.730	0.128	0.095	0.050
−0.25	0.001	−0.002	0.059	−0.104	−0.051	0.003	−0.017	0.032	0.365	0.440	0.401	0.127	0.090	0.055
0	0.001	−0.009	−0.069	0.002	−0.050	0.002	−0.002	0.031	0.382	0.311	0.243	0.128	0.092	0.058
0.25	−0.001	0.018	−0.078	0.009	−0.040	0.004	0.007	0.029	0.410	0.263	0.200	0.129	0.097	0.067
0.5	0.001	−0.036	−0.072	0.006	−0.036	0.004	0.013	0.026	0.512	0.205	0.164	0.136	0.097	0.075
0.75	0.001	−0.035	−0.034	−0.004	−0.037	0.003	0.007	0.025	0.840	0.111	0.127	0.130	0.096	0.074
0.9	0.001	−0.056	−0.018	0.013	−0.039	−0.004	0.011	0.023	2.078	0.047	0.097	0.127	0.095	0.063
(100,5)														
−0.9	−0.001	0.007	0.073	−0.129	−0.053	0.009	0.021	0.028	0.284	0.151	0.262	0.110	0.074	0.064
−0.75	0.001	−0.013	0.257	−0.298	−0.050	0.013	0.033	0.029	0.297	0.442	0.517	0.106	0.078	0.076
−0.5	0.000	−0.006	0.470	−0.464	−0.056	0.005	−0.008	0.027	0.284	0.767	0.751	0.106	0.071	0.038
−0.25	−0.001	0.016	0.098	−0.120	−0.057	0.002	−0.014	0.026	0.286	0.433	0.393	0.109	0.071	0.048
0	−0.001	0.006	0.008	−0.036	−0.048	0.002	−0.008	0.024	0.287	0.285	0.242	0.109	0.074	0.051
0.25	0.000	−0.002	−0.011	−0.020	−0.056	0.003	−0.002	0.022	0.304	0.210	0.183	0.108	0.072	0.055
0.5	0.000	0.007	−0.020	−0.008	−0.049	0.008	0.004	0.021	0.374	0.157	0.146	0.110	0.074	0.062
0.75	0.000	0.029	−0.009	−0.010	−0.044	0.006	0.000	0.018	0.675	0.066	0.096	0.111	0.074	0.051
0.9	0.000	0.022	−0.013	0.016	−0.058	0.005	0.010	0.019	1.515	0.033	0.081	0.110	0.072	0.048

in which different colors represent different counties. Each cell represents a district, and two cells are considered as neighbors by the first-order rook contiguity criterion. Moreover, W_S is row-standardized so that each row sums up to 1.

(2) Data Generating Process (DGP)

The DGP for the independent and dependent variable x_{ijt} and y_{ijt} are carried out based on Eq. (4.1).

$$\begin{cases} y_{ijt} = \beta_0 + x_{ijt}\beta_1 + u_{ijt}, \\ u_{ijt} = \rho \sum_{g=1}^{N} \sum_{h=1}^{m_g} w_{ij,gh} u_{ght} + \theta \sum_{g=1}^{N} \sum_{h=1}^{m_g} w_{ij,gh} \epsilon_{ght} + \epsilon_{ijt}, \\ \epsilon_{ijt} = \alpha_i + \mu_{ij} + v_{ijt}, \end{cases} \tag{4.1}$$

The explanatory variable $x_{ijt} = 5\delta_{ijt} + 5 + 0.5t + \upsilon_{ijt}$ with $\delta_{ijt} \sim$ i.i.d. $U[0,1]$ is respectively random variable uniformly distributed over the interval [0, 1], t according to the initial setting, and $\upsilon_{ijt} \sim$ i.i.d. $N(0,1)$ is respectively random variable normally distributed over the interval [0, 1]. The item $5\delta_{ijt} + 5$ makes x_{ijt} belong to a stable value interval from 5 to 10, $0.5t$ lets the variable vary from time and the random generated υ_{ijt} ensures x_{ijt} a stochastic variable. α_i, μ_{ij} and v_{ijt} satisfy Assumptions 1 (a), (c), and are normally distributed.

(3) Sample Size

According to the literature review, the numbers of sample sizes on Monte Carlo simulation are generally used to specify the observations S equal to less than 50 and the time periods T equal to less than 10. In this part, with two pairs of samples for the sample sizes $S \in \{36, 64, 100\}$, $T \in \{5, 8, 10\}$, the simulation results show the performance of the estimate of parameters $\beta_0, \beta_1, \rho, \theta, \sigma_\alpha, \sigma_\mu, \sigma_\nu$ for HSEARMA model.

The Monte Carlo simulation algorithm flow chart displays in Fig. 2. It follows the steps below:

(1) Given the initial number of simulation $mc = 1$, the sample sizes (S, T) and spatial weights matrices W_S are given according to the simulation design.
(2) Set the initial value of $\beta_0, \beta_1, \rho_0, \theta_0, \sigma_\nu$ and define $r_1 = 0.2, r_2 = 0.5$ (to obtain the initial value of σ_α and σ_μ), we have the initial value vector of the parameter $\varphi_0 = (\beta_0, \beta_1, \rho_0, \theta_0, \sigma_\alpha, \sigma_\mu, \sigma_\nu)$.
(3) By the mc times simulation, according to DGP, the independent variable matrix X, the dependent variable y, and the corresponding error vector u of the HSEARMA model are obtained. After these steps, we can obtain the sample (y, X).
(4) Based on the GMM-FGLS estimation algorithm of HSEARMA model, the estimate of parameters vector $\hat{\varphi}_{mc} = (\hat{\beta}_{0_{mc}}, \hat{\beta}_{1_{mc}}, \hat{\rho}_{mc}, \hat{\theta}_{mc}, \hat{\sigma}_{\alpha_{mc}}, \hat{\sigma}_{\mu_{mc}}, \hat{\sigma}_{\nu_{mc}})$ is obtained, where $mc = 1, \ldots, 100$.
(5) Repeat steps (3) and (4) 1000 times and finally get 1000 estimates for each parameter.
(6) Calculate the Bias and RMSE by the corresponding formulae in Fig. 2, and analyze their performances.

4.2. Monte Carlo simulation results

To summarize, the simulation design above contains 729 cases,[4] which results from 9 sections of ρ_0, 9 sections of θ_0, 9 pairs of the sample sizes. Each table corresponds to 1 spatial moving average error

[4] For all the results, please see the supplementary.

Table 3
Biases and RMSEs of the estimators for HSEARMA model ($\theta_0 = 0.5, T = 8$).

ρ_0	Bias							RMSE						
	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\rho}$	$\widehat{\theta}$	$\widehat{\sigma}_\alpha$	$\widehat{\sigma}_\mu$	$\widehat{\sigma}_\nu$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\rho}$	$\widehat{\theta}$	$\widehat{\sigma}_\alpha$	$\widehat{\sigma}_\mu$	$\widehat{\sigma}_\nu$
(36.5)														
−0.9	0.001	−0.017	0.170	−0.311	0.012	−0.005	0.072	0.038	0.453	0.268	0.476	0.136	0.101	0.133
−0.75	0.001	−0.018	0.441	−0.525	0.009	−0.010	0.065	0.037	0.448	0.609	0.714	0.139	0.103	0.110
−0.5	−0.001	0.016	0.549	−0.557	−0.030	−0.015	−0.007	0.037	0.578	0.797	0.792	0.124	0.098	0.048
−0.25	−0.002	0.024	0.130	−0.169	−0.038	−0.018	−0.019	0.033	0.418	0.483	0.456	0.119	0.097	0.059
0	0.000	−0.005	−0.007	−0.059	−0.025	−0.018	−0.012	0.031	0.419	0.344	0.299	0.118	0.090	0.062
0.25	0.002	0.005	−0.059	−0.022	−0.026	−0.024	0.003	0.029	0.443	0.275	0.232	0.121	0.101	0.072
0.5	−0.001	0.004	−0.066	−0.014	−0.017	−0.014	0.012	0.025	0.531	0.234	0.201	0.126	0.096	0.087
0.75	−0.001	0.034	−0.038	−0.006	−0.017	−0.014	0.014	0.022	0.899	0.131	0.144	0.127	0.096	0.084
0.9	0.000	0.024	−0.015	0.001	−0.024	−0.018	0.010	0.020	2.132	0.052	0.114	0.119	0.099	0.071
(64,5)														
−0.9	0.001	−0.010	0.095	−0.174	−0.025	0.007	0.033	0.026	0.313	0.175	0.309	0.105	0.078	0.078
−0.75	0.000	−0.007	0.324	−0.386	−0.017	0.011	0.046	0.027	0.318	0.502	0.594	0.104	0.077	0.086
−0.5	0.000	−0.008	0.460	−0.472	−0.037	0.002	−0.007	0.026	0.311	0.728	0.726	0.102	0.072	0.037
−0.25	0.002	−0.028	0.008	−0.046	−0.041	−0.001	−0.009	0.025	0.310	0.381	0.334	0.100	0.072	0.043
0	0.000	0.005	−0.055	0.005	−0.037	0.003	−0.001	0.023	0.308	0.272	0.214	0.099	0.073	0.046
0.25	0.000	−0.011	−0.058	0.017	−0.037	0.002	0.009	0.021	0.330	0.200	0.165	0.104	0.074	0.053
0.5	0.001	−0.012	−0.057	0.011	−0.031	0.002	0.014	0.020	0.403	0.168	0.146	0.108	0.072	0.064
0.75	−0.001	0.005	−0.015	−0.011	−0.037	−0.003	0.002	0.018	0.671	0.073	0.101	0.101	0.072	0.053
0.9	0.000	0.055	−0.010	0.010	−0.035	0.001	0.008	0.017	1.685	0.029	0.078	0.101	0.072	0.045
(100,5)														
−0.9	0.001	−0.016	0.047	−0.088	−0.037	0.007	0.014	0.022	0.257	0.105	0.193	0.088	0.057	0.046
−0.75	0.001	−0.015	0.155	−0.184	−0.040	0.005	0.018	0.022	0.256	0.344	0.408	0.087	0.058	0.057
−0.5	0.001	−0.006	0.484	−0.483	−0.042	0.006	−0.007	0.021	0.259	0.783	0.768	0.086	0.057	0.032
−0.25	0.000	0.000	0.066	−0.079	−0.043	0.004	−0.010	0.020	0.250	0.366	0.330	0.086	0.057	0.038
0	0.000	0.004	0.022	−0.040	−0.045	0.003	−0.006	0.018	0.251	0.250	0.222	0.089	0.059	0.042
0.25	0.000	0.002	−0.007	−0.014	−0.035	0.003	−0.002	0.016	0.254	0.171	0.157	0.087	0.059	0.045
0.5	0.000	0.005	−0.010	−0.010	−0.039	0.005	0.001	0.015	0.313	0.124	0.127	0.089	0.057	0.049
0.75	0.000	0.009	−0.004	−0.010	−0.032	0.004	−0.002	0.014	0.526	0.052	0.080	0.086	0.056	0.039
0.9	0.000	0.021	−0.007	0.011	−0.046	0.003	0.004	0.014	1.211	0.024	0.062	0.089	0.058	0.035

Table 4
Biases and RMSEs of the estimators for HSEARMA model ($\theta_0 = 0.5, T = 10$).

ρ_0	Bias							RMSE						
	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\rho}$	$\widehat{\theta}$	$\widehat{\sigma}_\alpha$	$\widehat{\sigma}_\mu$	$\widehat{\sigma}_\nu$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\rho}$	$\widehat{\theta}$	$\widehat{\sigma}_\alpha$	$\widehat{\sigma}_\mu$	$\widehat{\sigma}_\nu$
(36.5)														
−0.9	0.002	−0.023	0.156	−0.288	0.001	−0.006	0.064	0.031	0.396	0.255	0.447	0.125	0.094	0.122
−0.75	0.000	0.008	0.405	−0.480	0.003	−0.004	0.062	0.031	0.403	0.575	0.674	0.118	0.090	0.106
−0.5	0.000	−0.004	0.523	−0.533	−0.028	−0.011	−0.007	0.031	0.408	0.781	0.776	0.102	0.085	0.044
−0.25	0.000	0.004	0.092	−0.129	−0.029	−0.014	−0.017	0.028	0.388	0.448	0.416	0.105	0.087	0.052
0	0.001	−0.003	−0.022	−0.039	−0.023	−0.014	−0.006	0.026	0.379	0.315	0.271	0.106	0.085	0.058
0.25	0.001	−0.020	−0.036	−0.030	−0.027	−0.018	−0.001	0.024	0.407	0.263	0.231	0.104	0.089	0.067
0.5	0.001	−0.013	−0.053	−0.020	−0.021	−0.018	0.011	0.023	0.473	0.218	0.195	0.112	0.089	0.081
0.75	0.001	−0.016	−0.029	−0.013	−0.013	−0.015	0.009	0.020	0.809	0.123	0.138	0.114	0.092	0.081
0.9	0.000	−0.034	−0.009	−0.005	−0.026	−0.013	0.006	0.019	1.899	0.042	0.103	0.105	0.086	0.064
(64,5)														
−0.9	−0.001	0.004	0.088	−0.164	−0.022	0.011	0.029	0.024	0.305	0.168	0.297	0.099	0.071	0.073
−0.75	−0.001	0.011	0.291	−0.351	−0.024	0.005	0.040	0.024	0.308	0.478	0.568	0.094	0.067	0.080
−0.5	0.000	0.010	0.430	−0.437	−0.035	0.003	−0.006	0.023	0.374	0.724	0.718	0.089	0.065	0.033
−0.25	0.001	−0.010	0.025	−0.056	−0.035	−0.003	−0.009	0.021	0.291	0.359	0.319	0.093	0.065	0.038
0	0.000	0.001	−0.062	0.016	−0.028	−0.001	0.005	0.021	0.289	0.247	0.199	0.093	0.064	0.044
0.25	0.000	−0.004	−0.051	0.014	−0.032	0.004	0.007	0.019	0.292	0.192	0.167	0.090	0.064	0.051
0.5	0.000	0.002	−0.049	0.005	−0.028	0.002	0.011	0.018	0.368	0.162	0.141	0.088	0.067	0.058
0.75	−0.001	0.016	−0.012	−0.010	−0.032	0.006	0.003	0.016	0.602	0.070	0.095	0.091	0.064	0.049
0.9	0.000	−0.025	−0.006	0.002	−0.038	0.003	0.002	0.015	1.448	0.025	0.072	0.092	0.064	0.040
(100,5)														
−0.9	−0.001	0.013	0.036	−0.067	−0.037	0.006	0.010	0.019	0.244	0.101	0.183	0.076	0.050	0.042
−0.75	0.001	−0.008	0.128	−0.150	−0.035	0.008	0.015	0.019	0.241	0.313	0.372	0.078	0.052	0.052
−0.5	0.000	0.003	0.448	−0.447	−0.037	0.006	−0.006	0.019	0.249	0.742	0.730	0.078	0.052	0.029
−0.25	0.001	−0.007	0.063	−0.077	−0.034	0.002	−0.008	0.018	0.244	0.360	0.329	0.077	0.051	0.036
0	0.001	−0.010	0.002	−0.016	−0.035	0.003	−0.002	0.016	0.232	0.212	0.185	0.078	0.048	0.037
0.25	0.000	0.004	−0.008	−0.008	−0.034	0.005	0.000	0.016	0.251	0.155	0.142	0.078	0.053	0.041
0.5	0.000	0.004	0.000	−0.017	−0.034	0.002	−0.003	0.014	0.292	0.109	0.116	0.076	0.051	0.044
0.75	0.000	0.006	−0.001	−0.013	−0.032	0.003	−0.002	0.013	0.475	0.045	0.071	0.081	0.050	0.034
0.9	0.000	−0.050	−0.005	0.006	−0.038	0.004	0.004	0.012	1.131	0.020	0.056	0.081	0.050	0.032

coefficient θ_0, 9 spatial autoregressive error coefficient ρ_0, 3 specifications S of observations and 1 time period T. Tables 2–4 including 81 cases report bias and RMSEs when the spatial moving average error coefficient $\theta_0 = 0.5$, the performance of GMM-FGLS estimates vary across the initial value of spatial autoregressive error coefficient $\rho_0 = \{-0.9, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 0.9\}$ for the sample sizes $S \in \{36, 64, 100\}$ and $T = 5, T = 8, T = 10$, respectively.

In these tables, the main column headings distinguish the column between parameter values, bias and RMSE. The second rows of the tables designate the GMM-FGLS estimators for HSEARMA model. $\hat{\rho}, \hat{\theta}, \hat{\sigma}_\alpha, \hat{\sigma}_\mu, \hat{\sigma}_\nu$ represent the optimal GMM estimators; $\hat{\beta}_0, \hat{\beta}_1$ represent the feasible GLS estimator based on the optimal GMM estimators. The bias and RMSE for each case can be found from the third row.

Firstly, the bias and RMSE of the regression parameter estimators are investigated. For all the estimators across all the sample sizes, the biases of the intercept coefficient $\hat{\beta}_0$ and the slope coefficient $\hat{\beta}_1$ are very close to zero, signifying that the distributions of our GMM-FGLS estimators are centered around the true value of β_0 and β_1. On average, the RMSEs of $\hat{\beta}_0$ are less than 0.05 performing much better than the $\hat{\beta}_1$'s. β_1 is not estimated as precisely as the other parameters, especially when ρ takes large positive values. Given that the bias of $\hat{\beta}_1$ is reasonable, we attribute it to a larger finite sample variance of $\hat{\beta}_1$ than estimators for the other parameters. Similar results have been reported in Kapoor et al. (2007). Secondly, the bias and RMSE of spatial error correlation coefficients estimators are considered. Though the spatial autoregressive error coefficient $\hat{\rho}$'s and spatial moving average error coefficient $\hat{\theta}$'s bias and RMSE are not fairly tiny value than the other parameters, they are still reasonable and not interfered by the large initial value of spatial error correlation coefficients. Finally, the performances of the standard errors of county-level permanent unit-specific random effect $\hat{\sigma}_\alpha$, district nested within county random effect $\hat{\sigma}_\mu$, and the other innovation $\hat{\sigma}_\nu$, are taken into account. The biases of $\hat{\sigma}_\alpha, \hat{\sigma}_\mu, \hat{\sigma}_\nu$ are very close to 0, implying that they are fairly close to the true value of the parameters. Based on the RMSEs of $\hat{\sigma}_\alpha, \hat{\sigma}_\mu, \hat{\sigma}_\nu$, $\sigma_\alpha, \sigma_\mu, \sigma_\nu$ are estimated quite precisely. Obviously, the precision of $\hat{\beta}_0, \hat{\beta}_1, \hat{\rho}, \hat{\theta}, \hat{\sigma}_\alpha, \hat{\sigma}_\mu, \hat{\sigma}_\nu$ improves as either N or T increases. Take the RMSEs of $\hat{\sigma}_\mu$ for example, when $T = 8$ and $\rho_0 = 0$, the RMSEs are 0.096, 0.071, 0.057 for $N = 36,64,100$, respectively; when $N = 64$ and $\rho_0 = 0$, the RMSEs are 0.094, 0.071, 0.065 for $T = 5,8,10$, respectively.

5. Concluding remarks and further research

This paper considers a hierarchically spatial autoregressive and moving average error (HSEARMA) model. This model incorporates an autoregressive and moving average error and the nested random effects structure at both the county and district levels. We develop an optimal GMM estimator for the spatial error correlation coefficient and the error components' variances, and a feasible generalized least squares estimator for the regression parameter vector. We study their asymptotic properties and conduct a finite-scale Monte Carlo simulation to evaluate their finite sample performances.

For future research, while we considered the nested random effects model, it would be interesting to consider a hierarchically spatial econometric model with nested fixed effects. In addition, a model that incorporates both the spatial autoregressive and moving average error, the spatially lagged dependent variables and geographic coordinates would be of great interest. Moreover, one can attempt to take the dynamic effects into account to extend the hierarchically spatial model. We leave these for future research.

Appendices

A. Moment Conditions and the optimal weights matrix

Moment conditions

To derive the moment conditions, we make use of the following formula: for a column random vector ϵ and a conformable matrix A, $E[\epsilon' A\epsilon] = tr[A\mathrm{Var}(\epsilon)] + E[\epsilon]' A E[\epsilon]$. Using this formula and the fact that $\bar{\epsilon} = W\epsilon$, we get Derivation of moment conditions

$$E[\epsilon' Q_1 \epsilon] = E[\nu' Q_1 \nu] = \sigma_\nu^2 tr(Q_1) = \sigma_\nu^2 (T-1)S = \xi_1 (T-1)S$$

$$E[\bar{\epsilon}' Q_1 \bar{\epsilon}] = E[\nu'(I_T \otimes W_S')Q_1 Q_1 (I_T \otimes W_S)\nu] = \sigma_\nu^2 tr[(I_T \otimes W_S')(E_T \otimes diag(I_{m_i}))(I_T \otimes W_S)] = \xi_1 (T-1) tr[W_S' W_S]$$

$$E[\bar{\epsilon}' Q_1 \epsilon] = E[\nu'(I_T \otimes W_S')Q_1 Q_1 \nu] = \sigma_\nu^2 tr[(I_T \otimes W_S')(E_T \otimes diag(I_{m_i}))] = \sigma_\nu^2 (T-1) tr[W_S'] = 0$$

$$E[\bar{\bar{\epsilon}}' Q_1 \epsilon] = E[\nu'(I_T \otimes W_S' W_S')Q_1 Q_1 \nu] = \sigma_\nu^2 tr[(I_T \otimes W_S' W_S')(E_T \otimes diag(I_{m_i}))] = \xi_1 (T-1) tr[W_S W_S]$$

$$E[\bar{\bar{\epsilon}}' Q_1 \bar{\epsilon}] = E[\nu'(I_T \otimes W_S' W_S')Q_1 Q_1 (I_T \otimes W_S)\nu] = \sigma_\nu^2 tr[(I_T \otimes W_S' W_S')(E_T \otimes diag(I_{m_i}))(I_T \otimes W_S)] = \xi_1 (T-1) tr[W_S W_S W_S']$$

$$E[\bar{\bar{\epsilon}}' Q_1 \bar{\bar{\epsilon}}] = E[\nu'(I_T \otimes W_S' W_S')Q_1 Q_1 (I_T \otimes W_S W_S)\nu] = \sigma_\nu^2 tr[(I_T \otimes W_S' W_S')(E_T \otimes diag(I_{m_i}))(I_T \otimes W_S W_S)] = \xi_1 (T-1) tr[W_S W_S W_S' W_S'].$$

$$\begin{aligned}
E[\epsilon' Q_2 \epsilon] &= E[\mu'(l_T' \otimes diag(E_{m_i}'))(l_T \otimes diag(E_{m_i}))\mu] + E[\nu' Q_2 \nu] \\
&= \sigma_\mu^2 tr(l_T' l_T \otimes diag(E_{m_i}' E_{m_i})) + \sigma_\nu^2 tr(Q_2) = \sigma_\mu^2 T tr(diag(E_{m_i})) + \sigma_\nu^2 tr(Q_2) \\
&= \sigma_\mu^2 T(S-N) + \sigma_\nu^2 (S-N) = \xi_2 (S-N)
\end{aligned}$$

$$\begin{aligned}
E[\bar{\epsilon}' Q_2 \bar{\epsilon}] &= E[\epsilon'(I_T \otimes W_S')Q_2 Q_2 (I_T \otimes W_S)\epsilon] \\
&= E[\alpha'(l_T' \otimes diag(l_{m_i}')W_S' diag(E_{m_i}'))(l_T \otimes diag(E_{m_i})W_S diag(l_{m_i}))\alpha] + E[\mu'(l_T' \otimes W_S' diag(E_{m_i}'))(l_T \otimes diag(E_{m_i})W_S)\mu] + E[\nu'(I_T \otimes W_S')Q_2 Q_2 (I_T \otimes W_S)\nu] \\
&= \sigma_\alpha^2 tr[l_T' l_T \otimes diag(l_{m_i}')W_S' diag(E_{m_i})W_S diag(l_{m_i})] + \sigma_\mu^2 tr[l_T' l_T \otimes W_S' diag(E_{m_i})W_S] + \sigma_\nu^2 tr[\bar{J}_T \otimes W_S' diag(E_{m_i})W_S] \\
&= \sigma_\alpha^2 T tr[W_S' diag(E_{m_i})W_S diag(l_{m_i}) diag(l_{m_i}')] + \sigma_\mu^2 T tr[W_S' diag(E_{m_i})W_S] + \sigma_\nu^2 tr[W_S' diag(E_{m_i})W_S] \\
&= \sigma_\alpha^2 T tr[W_S' diag(E_{m_i})W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S' diag(E_{m_i})] \\
&= \frac{(\xi_3 - \xi_2)N}{S} tr[W_S' diag(E_{m_i})W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S' diag(E_{m_i})]
\end{aligned}$$

$$
\begin{aligned}
E[\overline{\epsilon}'Q_2\epsilon] &= E[\epsilon'(I_T \otimes W_S')Q_2Q_2\epsilon] \\
&= E[\alpha'(l_T' \otimes diag(l_{m_i}')W_S'diag(E_{m_i}'))(l_T \otimes diag(E_{m_i}l_{m_i}))\alpha] + E[\mu'(l_T' \otimes W_S'diag(E_{m_i}'))(l_T \otimes diag(E_{m_i}))\mu] + E[\nu'(I_T \otimes W_S')Q_2Q_2\nu] \\
&= \sigma_\mu^2 tr[l_T'l_T \otimes W_S'diag(E_{m_i}')diag(E_{m_i})] + \sigma_\nu^2 tr[(I_T \otimes W_S')(\overline{J}_T \otimes diag(E_{m_i}))] \\
&= \sigma_\mu^2 Ttr[W_S'diag(E_{m_i})] + \sigma_\nu^2 tr[W_S'diag(E_{m_i})] \\
&= \xi_2 tr[W_S'diag(E_{m_i})]
\end{aligned}
$$

$$
\begin{aligned}
E[\overline{\overline{\epsilon}}'Q_2\epsilon] &= E[\epsilon'(I_T \otimes W_S')(I_T \otimes W_S')Q_2Q_2\epsilon] \\
&= E[\alpha'(l_T' \otimes diag(l_{m_i}')W_S'W_S'diag(E_{m_i}'))(l_T \otimes diag(E_{m_i}l_{m_i}))\alpha] + E[\mu'(l_T' \otimes W_S'W_S'diag(E_{m_i}'))(l_T \otimes diag(E_{m_i}))\mu] + E[\nu'(I_T \otimes W_S'W_S')Q_2Q_2\nu] \\
&= \sigma_\mu^2 tr[l_T'l_T \otimes W_S'W_S'diag(E_{m_i}')diag(E_{m_i})] + \sigma_\nu^2 tr[(I_T \otimes W_S'W_S')(\overline{J}_T \otimes diag(E_{m_i}))] \\
&= \sigma_\mu^2 Ttr[W_S'W_S'diag(E_{m_i})] + \sigma_\nu^2 tr[W_S'W_S'diag(E_{m_i})] = \xi_2 tr[W_S'W_S'diag(E_{m_i})]
\end{aligned}
$$

$$
\begin{aligned}
E[\overline{\overline{\epsilon}}'Q_2\overline{\epsilon}] &= E[\epsilon'(I_T \otimes W_S')(I_T \otimes W_S')Q_2Q_2(I_T \otimes W_S)\epsilon] \\
&= E[\alpha'(l_T' \otimes diag(l_{m_i}')W_S'W_S'diag(E_{m_i}'))(l_T \otimes diag(E_{m_i})W_S diag(l_{m_i}))\alpha] \\
&\quad + E[\mu'(l_T' \otimes W_S'W_S'diag(E_{m_i}'))(l_T \otimes diag(E_{m_i})W_S)\mu] + E[\nu'(I_T \otimes W_S'W_S')Q_2Q_2(I_T \otimes W_S)\nu] \\
&= \sigma_\alpha^2 tr[l_T'l_T \otimes diag(l_{m_i}')W_S'W_S'diag(E_{m_i})W_S diag(l_{m_i})] + \sigma_\mu^2 tr[l_T'l_T \otimes W_S'W_S'diag(E_{m_i})W_S] + \sigma_\nu^2 tr[\overline{J}_T \otimes W_S'W_S'diag(E_{m_i})W_S] \\
&= \sigma_\alpha^2 Ttr[W_S'W_S'diag(E_{m_i})W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S'W_S'diag(E_{m_i})] \\
&= \frac{(\xi_3-\xi_2)N}{S} tr[W_S'W_S'diag(E_{m_i})W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S'W_S'diag(E_{m_i})]
\end{aligned}
$$

$$
\begin{aligned}
E[\overline{\overline{\epsilon}}'Q_2\overline{\overline{\epsilon}}] &= E[\epsilon'(I_T \otimes W_S')(I_T \otimes W_S')Q_2Q_2(I_T \otimes W_S)(I_T \otimes W_S)\epsilon] \\
&= E[\alpha'(l_T' \otimes diag(l_{m_i}')W_S'W_S'diag(E_{m_i}'))(l_T \otimes diag(E_{m_i})W_S W_S diag(l_{m_i}))\alpha] \\
&\quad + E[\mu'(l_T' \otimes W_S'W_S'diag(E_{m_i}'))(l_T \otimes diag(E_{m_i})W_S W_S)\mu] + E[\nu'(I_T \otimes W_S'W_S')Q_2Q_2(I_T \otimes W_S W_S)\nu] \\
&= \sigma_\alpha^2 tr[l_T'l_T \otimes diag(l_{m_i}')W_S'W_S'diag(E_{m_i})W_S W_S diag(l_{m_i})] + \sigma_\mu^2 tr[l_T'l_T \otimes W_S'W_S'diag(E_{m_i})W_S W_S] + \sigma_\nu^2 tr[\overline{J}_T \otimes W_S'W_S'diag(E_{m_i})W_S W_S] \\
&= \sigma_\alpha^2 Ttr[W_S'W_S'diag(E_{m_i})W_S W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S W_S'W_S'diag(E_{m_i})] \\
&= \frac{(\xi_3-\xi_2)N}{S} tr[W_S'W_S'diag(E_{m_i})W_S W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S W_S'W_S'diag(E_{m_i})]
\end{aligned}
$$

$$
\begin{aligned}
E[\epsilon'Q_3\epsilon] &= E[\alpha'(l_T' \otimes diag(l_{m_i}'))(l_T \otimes diag(l_{m_i}))\alpha] + E[\mu'(l_T' \otimes diag(\overline{J}_{m_i}'))(l_T \otimes diag(\overline{J}_{m_i}))\mu] + E[\nu'Q_3\nu] \\
&= \sigma_\alpha^2 tr[l_T'l_T \otimes diag(l_{m_i}'l_{m_i})] + \sigma_\mu^2 tr[l_T'l_T \otimes diag(\overline{J}_{m_i}'\overline{J}_{m_i})] + \sigma_\nu^2 tr(Q_3) \\
&= \sigma_\alpha^2 Ttr(diag(m_i)) + \sigma_\mu^2 Ttr(diag(\overline{J}_{m_i})) + \sigma_\nu^2 N = \sigma_\alpha^2 TS + (\sigma_\mu^2 T + \sigma_\nu^2)N = \sigma_\alpha^2 TS + \xi_2 N = \xi_3 N \\
&\Rightarrow \sigma_\alpha^2 = \frac{(\xi_3-\xi_2)N}{TS}
\end{aligned}
$$

$$
\begin{aligned}
E[\overline{\epsilon}'Q_3\overline{\epsilon}] &= E[\epsilon'(I_T \otimes W_S')Q_3(I_T \otimes W_S)\epsilon] \\
&= E[\alpha'(l_T' \otimes diag(l_{m_i}')W_S'diag(\overline{J}_{m_i}'))(l_T \otimes diag(\overline{J}_{m_i})W_S diag(l_{m_i}))\alpha] + E[\mu'(l_T' \otimes W_S'diag(\overline{J}_{m_i}'))(l_T \otimes diag(\overline{J}_{m_i})W_S)\mu] + E[\nu'(I_T \otimes W_S')Q_3Q_3(I_T \otimes W_S)\nu] \\
&= \sigma_\alpha^2 Ttr[W_S'diag(\overline{J}_{m_i})W_S diag(l_{m_i})diag(l_{m_i}')] + \sigma_\mu^2 Ttr[W_S'diag(\overline{J}_{m_i})W_S] + \sigma_\nu^2 tr[W_S'diag(\overline{J}_{m_i})W_S] \\
&= \sigma_\alpha^2 Ttr[W_S'diag(\overline{J}_{m_i})W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S'diag(\overline{J}_{m_i})] \\
&= \frac{(\xi_3-\xi_2)N}{S} tr[W_S'diag(\overline{J}_{m_i})W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S'diag(\overline{J}_{m_i})]
\end{aligned}
$$

$$
\begin{aligned}
E[\overline{\epsilon}'Q_3\epsilon] &= E[\epsilon'(I_T \otimes W_S')Q_3\epsilon] = E[\alpha'(l_T' \otimes diag(l_{m_i}')W_S'diag(\overline{J}_{m_i}'))(l_T \otimes diag(l_{m_i}))\alpha] + E[\mu'(l_T' \otimes W_S'diag(\overline{J}_{m_i}'))(l_T \otimes diag(\overline{J}_{m_i}))\mu] + E[\nu'(I_T \otimes W_S')Q_3Q_3\nu] \\
&= \sigma_\alpha^2 Ttr[diag(l_{m_i}')W_S'diag(\overline{J}_{m_i}')diag(l_{m_i})] + \sigma_\mu^2 Ttr[W_S'diag(\overline{J}_{m_i})] + \sigma_\nu^2 tr[W_S'diag(\overline{J}_{m_i})] \\
&= \sigma_\alpha^2 Ttr[W_S'diag(J_{m_i})] + \xi_2 tr[W_S'diag(\overline{J}_{m_i})] = \frac{(\xi_3-\xi_2)N}{S} tr[W_S'diag(J_{m_i})] + \xi_2 tr[W_S'diag(\overline{J}_{m_i})]
\end{aligned}
$$

$$
\begin{aligned}
E[\bar{\bar{\epsilon}}'Q_3\epsilon] &= E[\epsilon'(I_T \otimes W_S')(I_T \otimes W_S')Q_3\epsilon] \\
&= E[\alpha'(l_T' \otimes diag(l_{m_i}')W_S'W_S'diag(\bar{J}_{m_i}'))(l_T \otimes diag(l_{m_i}))\alpha] + E[\mu'(l_T' \otimes W_S'W_S'diag(\bar{J}_{m_i}'))(l_T \otimes diag(\bar{J}_{m_i}))\mu] + E[\nu'(I_T \otimes W_S'W_S')Q_3Q_3\nu] \\
&= \sigma_\alpha^2 Ttr[diag(l_{m_i}')W_S'W_S'diag(\bar{J}_{m_i}')diag(l_{m_i})] + \sigma_\mu^2 Ttr[W_S'W_S'diag(\bar{J}_{m_i})] + \sigma_\nu^2 tr[W_S'W_S'diag(\bar{J}_{m_i})] \\
&= \sigma_\alpha^2 Ttr[W_S'W_S'diag(J_{m_i})] + \xi_2 tr[W_S'W_S'diag(\bar{J}_{m_i})] \\
&= \frac{(\xi_3-\xi_2)N}{S} tr[W_S'W_S'diag(J_{m_i})] + \xi_2 tr[W_S'W_S'diag(\bar{J}_{m_i})]
\end{aligned}
$$

$$
\begin{aligned}
E[\bar{\bar{\epsilon}}'Q_3\bar{\epsilon}] &= E[\epsilon'(I_T \otimes W_S'W_S')Q_3(I_T \otimes W_S)\epsilon] \\
&= E[\alpha'(l_T' \otimes diag(l_{m_i}')W_S'W_S'diag(\bar{J}_{m_i}'))(l_T \otimes diag(\bar{J}_{m_i})W_S diag(l_{m_i}))\alpha] \\
&\quad + E[\mu'(l_T' \otimes W_S'W_S'diag(\bar{J}_{m_i}'))(l_T \otimes diag(\bar{J}_{m_i})W_S)\mu] + E[\nu'(I_T \otimes W_S'W_S')Q_3Q_3(I_T \otimes W_S)\nu] \\
&= \sigma_\alpha^2 Ttr[diag(l_{m_i}')W_S'W_S'diag(\bar{J}_{m_i})W_S diag(l_{m_i})] + \sigma_\mu^2 Ttr[W_S'W_S'diag(\bar{J}_{m_i})W_S] + \sigma_\nu^2 tr[W_S'W_S'diag(\bar{J}_{m_i})W_S] \\
&= \sigma_\alpha^2 Ttr[W_S'W_S'diag(\bar{J}_{m_i})W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S'W_S'diag(\bar{J}_{m_i})] \\
&= \frac{(\xi_3-\xi_2)N}{S} tr[W_S'W_S'diag(\bar{J}_{m_i})W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S'W_S'diag(\bar{J}_{m_i})]
\end{aligned}
$$

$$
\begin{aligned}
E[\bar{\bar{\epsilon}}'Q_3\bar{\bar{\epsilon}}] &= E[\epsilon'(I_T \otimes W_S'W_S')Q_3(I_T \otimes W_S W_S)\epsilon] \\
&= E[\alpha'(l_T' \otimes diag(l_{m_i}')W_S'W_S'diag(\bar{J}_{m_i}'))(l_T \otimes diag(\bar{J}_{m_i})W_S W_S diag(l_{m_i}))\alpha] \\
&\quad + E[\mu'(l_T' \otimes W_S'W_S'diag(\bar{J}_{m_i}'))(l_T \otimes diag(\bar{J}_{m_i})W_S W_S)\mu] + E[\nu'(I_T \otimes W_S'W_S')Q_3Q_3(I_T \otimes W_S W_S)\nu] \\
&= \sigma_\alpha^2 Ttr[diag(l_{m_i}')W_S'W_S'diag(\bar{J}_{m_i})W_S W_S diag(l_{m_i})] + \sigma_\mu^2 Ttr[W_S'W_S'diag(\bar{J}_{m_i})W_S W_S] + \sigma_\nu^2 tr[W_S'W_S'diag(\bar{J}_{m_i})W_S W_S] \\
&= \sigma_\alpha^2 Ttr[W_S'W_S'diag(\bar{J}_{m_i})W_S W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S W_S'W_S'diag(\bar{J}_{m_i})] \\
&= \frac{(\xi_3-\xi_2)N}{S} tr[W_S'W_S'diag(\bar{J}_{m_i})W_S W_S diag(J_{m_i})] + \xi_2 tr[W_S W_S W_S'W_S'diag(\bar{J}_{m_i})]
\end{aligned}
$$

with $Q_1 = E_T \otimes I_S = (I_T - \bar{J}_T) \otimes I_S$, $Q_2 = \bar{J}_T \otimes diag(E_{m_i}) = \bar{J}_T \otimes [I_S - diag(\bar{J}_{m_i})]$, $Q_3 = \bar{J}_T \otimes diag(\bar{J}_{m_i})$, $\xi_1 = \frac{1}{(T-1)S}\epsilon'Q_1\epsilon$, $\xi_2 = \frac{1}{S-N}\epsilon'Q_2\epsilon$, $\xi_3 = \frac{1}{N}\epsilon'Q_3\epsilon$where $\bar{J}_T l_T = l_T$, $E_T l_T = 0$, $diag(l_{m_i})diag(l_{m_i}') = diag(J_{m_i})$, $diag(\bar{J}_{m_i}')diag(J_{m_i}) = diag(J_{m_i})$.

Substituting (3.2) and the moment condition elements into (3.3), a system of nine equation involves ρ, θ, ξ_1, ξ_2 and ξ_3 can be expressed as

$$\Gamma\varphi - \gamma = 0, \tag{A.1}$$

where $\varphi = [\rho, \rho^2, \xi_1, \xi_2, \xi_3, \theta\xi_1, \theta\xi_2, \theta\xi_3, \theta^2\xi_1, \theta^2\xi_2, \theta^2\xi_3]'$,

$$
\Gamma = \begin{bmatrix}
\gamma_{11}^1 & \gamma_{12}^1 & 1 & 0 & 0 & 0 & 0 & 0 & \gamma_{15}^1 & 0 & 0 \\
\gamma_{21}^1 & \gamma_{22}^1 & \gamma_{23}^1 & 0 & 0 & \gamma_{24}^1 & 0 & 0 & \gamma_{25}^1 & 0 & 0 \\
\gamma_{31}^1 & \gamma_{32}^1 & 0 & 0 & 0 & \gamma_{34}^1 & 0 & 0 & \gamma_{35}^1 & 0 & 0 \\
\gamma_{11}^2 & \gamma_{12}^2 & 0 & 1 & 0 & 0 & \gamma_{14}^2 & 0 & 0 & \gamma_{15}^2 & \gamma_{45}^2 \\
\gamma_{21}^2 & \gamma_{22}^2 & 0 & \gamma_{23}^2 & \gamma_{53}^2 & 0 & \gamma_{24}^2 & \gamma_{54}^2 & 0 & \gamma_{25}^2 & \gamma_{55}^2 \\
\gamma_{31}^2 & \gamma_{32}^2 & 0 & \gamma_{33}^2 & 0 & 0 & \gamma_{34}^2 & \gamma_{64}^2 & 0 & \gamma_{35}^2 & \gamma_{65}^2 \\
\gamma_{11}^3 & \gamma_{12}^3 & 0 & 0 & 1 & 0 & \gamma_{44}^3 & \gamma_{14}^3 & 0 & \gamma_{45}^3 & \gamma_{15}^3 \\
\gamma_{21}^3 & \gamma_{22}^3 & 0 & \gamma_{53}^3 & \gamma_{23}^3 & 0 & \gamma_{54}^3 & \gamma_{24}^3 & 0 & \gamma_{55}^3 & \gamma_{25}^3 \\
\gamma_{31}^3 & \gamma_{32}^3 & 0 & \gamma_{63}^3 & \gamma_{33}^3 & 0 & \gamma_{64}^3 & \gamma_{34}^3 & 0 & \gamma_{65}^3 & \gamma_{35}^3
\end{bmatrix}, \gamma = \begin{bmatrix}
\gamma_1^1 \\ \gamma_2^1 \\ \gamma_3^1 \\ \gamma_1^2 \\ \gamma_2^2 \\ \gamma_3^2 \\ \gamma_1^3 \\ \gamma_2^3 \\ \gamma_3^3
\end{bmatrix}
$$

In the expressions of Γ and γ, for $s = 1,2,3$, $\gamma_{11}^s = \phi_s\tau_{11}^s$, $\gamma_{21}^s = \phi_s\tau_{21}^s$, $\gamma_{31}^s = \phi_s\tau_{31}^s$, $\gamma_{12}^s = \phi_s\tau_{12}^s$, $\gamma_{22}^s = \phi_s\tau_{22}^s$, $\gamma_{32}^s = \phi_s\tau_{32}^s$, $\gamma_1^s = \phi_s\tau_1^s, \gamma_2^s = \phi_s\tau_2^s$, $\gamma_3^s = \phi_s\tau_3^s$, where $\tau_{11}^s = 2\mathrm{E}[\bar{u}'Q_s u]$, $\tau_{21}^s = 2\mathrm{E}[\bar{\bar{u}}'Q_s\bar{u}]$, $\tau_{31}^s = \mathrm{E}[\bar{u}'Q_s\bar{u} + \bar{\bar{u}}'Q_s u]$, $\tau_{12}^s = -\mathrm{E}[\bar{u}'Q_s\bar{u}]$, $\tau_{22}^s = -\mathrm{E}[\bar{\bar{u}}'Q_s\bar{\bar{u}}]$, $\tau_{32}^s = -\mathrm{E}[\bar{\bar{u}}'Q_s\bar{u}]$, $\phi_1 = \frac{1}{(T-1)S}, \phi_2 = \frac{1}{S-N}, \phi_3 = \frac{1}{N}, \tau_1^s = E[u^T Q_s u], \tau_2^s = E[\bar{u}^T Q_s\bar{u}], \tau_3^s = E[\bar{u}^T Q_s u]$.

Moreover,

$\gamma_{23}^1 = \gamma_{15}^1 = \frac{1}{S}tr[W_S'W_S]$, $\gamma_{24}^1 = \frac{2}{S}tr[W_S'R_1]$, $\gamma_{25}^1 = \frac{1}{S}tr[R_1'R_1]$, $\gamma_{34}^1 = \frac{1}{S}(tr[W_S'W_S] + tr[R_1])$, $\gamma_{35}^1 = \frac{1}{S}tr[W_S'R_1]$, $\gamma_{23}^2 = \gamma_{15}^2 = -\phi_2(tr[R_2W_SJ^*] - tr[R_2W_S])$, $\gamma_{33}^2 = \phi_2 tr[R_2]$, $\gamma_{53}^2 = \gamma_{45}^2 = \phi_2 tr[R_2W_SJ^*]$, $\gamma_{14}^2 = 2\phi_2 tr[R_2]$, $\gamma_{24}^2 = -2\phi_2(tr[W_S'R_2W_SJ^*] - tr[W_S'R_2W_S])$, $\gamma_{34}^2 = -\phi_2(tr[R_2W_SJ^*] - tr[R_2W_S] - tr[W_S'R_2])$, $\gamma_{54}^2 = 2\phi_2 tr[W_S'R_2W_SJ^*]$, $\gamma_{64}^2 = \phi_2 tr[R_2W_SJ^*]$, $\gamma_{25}^2 = -\phi_2(tr[W_S'R_2R_1J^*] - tr[W_S'R_2R_1])$, $\gamma_{35}^2 = -\phi_2(tr[W_S'R_2W_SJ^*] - tr[W_S'R_2W_S])$, $\gamma_{55}^2 = \phi_2 tr[W_S'R_2R_1J^*]$, $\gamma_{65}^2 = \phi_2 tr[W_S'R_2W_SJ^*]$, $\gamma_{23}^3 = \gamma_{15}^3 = \phi_3 tr[R_3J^*]$, $\gamma_{33}^3 = \phi_3 tr[W_S'J^*]$, $\gamma_{53}^3 = \gamma_{45}^3 = -\phi_3(tr[R_3J^*] - tr[R_3])$, $\gamma_{63}^3 = -\phi_3(tr[W_S'J^*] - tr[W_S^{*\prime}])$, $\gamma_{14}^3 = 2\phi_3 tr[W_S'J^*]$, $\gamma_{24}^3 = 2\phi_3 tr[W_S'R_3J^*]$, $\gamma_{34}^3 = \phi_3(tr[R_3J^*] + tr[R_1'J^*])$, $\gamma_{44}^3 = -2\phi_3(tr[W_S'J^*] - tr[W_S^{*\prime}])$, $\gamma_{54}^3 = -2\phi_3(tr[W_S'R_3J^*] - tr[W_S'R_3])$, $\gamma_{64}^3 = -\phi_3(tr[R_3J^*] - tr[R_3] + tr[R_1'J^*] - tr[W_S'W_S^{*\prime}])$, $\gamma_{25}^3 = \phi_3 tr[W_S'R_3W_SJ^*]$, $\gamma_{35}^3 = \phi_3 tr[W_S'R_3J^*]$, $\gamma_{55}^3 = -\phi_3(tr[W_S'R_3W_SJ^*] - tr[W_S'R_3W_S])$, $\gamma_{65}^3 = -\phi_3(tr[W_S'R_3J^*] - tr[W_S'R_3])$.

Optimal weights matrix

To derive elements in the optimal weights matrix, we make use of the following fact. Let $\epsilon \sim N(0, \Omega_\epsilon)$, A_S and B_S be nonnegative definite symmetric matrices. By the Lemma in Amemiya (1971), we obtain $\mathrm{Cov}(\epsilon' A_S \epsilon, \epsilon' B_S \epsilon) = 2\mathrm{tr}[A_S \Omega_\epsilon B_S \Omega_\epsilon]$. If either A_S or B_S is not a symmetric matrix, we use $(A_S + A_S')/2$ or $(B_S + B_S')/2$ to replace A_S and B_S, respectively, and obtain

$$\mathrm{Cov}(\epsilon' A_S \epsilon, \epsilon' B_S \epsilon) = \mathrm{Cov}(\epsilon' \frac{A_S' + A_S}{2}\epsilon, \epsilon' \frac{B_S' + B_S}{2}\epsilon) = \frac{1}{4}\mathrm{tr}[(A_S' + A_S)\Omega_\epsilon (B_S' + B_S)\Omega_\epsilon] = \frac{1}{2}\mathrm{tr}[(A_S + A_S')\Omega_\epsilon B_S \Omega_\epsilon].$$

Here we compute the term Ξ_{12} as an illustration. Derivations for other terms of Ξ are analogous and can be found in the supplementary file. Recall that $\overline{\epsilon} = W\epsilon$ and $\Omega_\epsilon = \xi_1 Q_1 + \xi_2 Q_2 + (I_T \otimes d(\xi_{3i} I_{m_i}))Q_3$, then it is not difficult to see that

$$\Xi_{12} = S\mathrm{Cov}(\frac{\epsilon' Q_1 \epsilon}{(T-1)S}, \frac{\overline{\epsilon}' Q_1 \overline{\epsilon}}{(T-1)S}) = \frac{2}{(T-1)^2 S}\mathrm{tr}[Q_1 \Omega_\epsilon (I_T \otimes W_S') Q_1 (I_T \otimes W_S)\Omega_\epsilon]$$
$$= \frac{2}{(T-1)^2 S}\mathrm{tr}[\xi_1^2 Q_1 (I_T \otimes (W_S' W_S))] = \frac{2\xi_1^2}{(T-1)S}\mathrm{tr}[W_S' W_S].$$

B Lemmas

Lemma 1.

(a) If the row (column) sum of $\{A_S\}_{S=1}^\infty$ is uniformly bounded in absolute value, and B_T is a finite constant matrix of fixed dimension $T \times T$, then the row (column) sum of $\{B_T \otimes A_S\}_{S=1}^\infty$ is uniformly bounded in absolute value.
(b) If the row (column) sum of $\{A_S\}_{S=1}^\infty$ and $\{B_S\}_{S=1}^\infty$ are both uniformly bounded in absolute value, so is $\{A_S B_S\}_{S=1}^\infty$ and $\{A_S + B_S\}_{S=1}^\infty$.
(c) If the row (column) sum of $\{A_S\}_{S=1}^\infty$ is uniformly bounded in absolute value, and the elements of $\{B_S\}_{S=1}^\infty$ (B_S of dimension $K \times S$) and $\{C_S\}_{S=1}^\infty$ (C_S of dimension $S \times K$) are uniformly bounded in absolute value, then the elements of $\{A_S C_S\}_{S=1}^\infty$ and $\{\frac{B_S A_S C_S}{S}\}_{S=1}^\infty$ are also uniformly bounded in absolute value.

Lemma 2. Let B_{1T} be a $T \times T$ constant matrix with T fixed and B_{2S} be an $S \times S$ constant matrix, such that the row and column sums of B_{1T} and B_{2S} are uniformly bounded in absolute value. Under Assumption 1, for $\epsilon = [l_T \otimes d(l_{m_i})]\alpha + (l_T \otimes I_S)\mu + \nu$ and

$$Z = \frac{1}{\varphi_{N,S}}\epsilon'(B_{1T} \otimes B_{2S})\epsilon,$$

we have $\mathrm{E}[Z] = O(1)$, $\mathrm{Var}(Z) = o(1)$, and consequently $Z - \mathrm{E}[Z] = o_P(1)$ as $N, S \to \infty$, where $\varphi_{N,S}$ can either be N, $S - N$, or S.

Proof. By writing

$\epsilon = [l_T \otimes d(l_{m_i}), l_T \otimes I_S, I_{TS}]\zeta$,

where $\zeta = (\alpha', \mu', \nu')'$, we get $Z = \frac{1}{\varphi_{N,S}}\epsilon'(B_{1T} \otimes B_{2S})\epsilon = \frac{1}{\varphi_{N,S}}\zeta' \Lambda \zeta$, where

$$\Lambda = \begin{pmatrix} (l_T' B_{1T} l_T) \otimes [d'(l_{m_i}) B_{2S} d(l_{m_i})] & (l_T' B_{1T} l_T) \otimes [d'(l_{m_i}) B_{2S}] & (l_T' B_{1T}) \otimes [d'(l_{m_i}) B_{2S}] \\ (l_T' B_{1T} l_T) \otimes [B_{2S} d(l_{m_i})] & (l_T' B_{1T} l_T) \otimes B_{2S} & (l_T' B_{1T}) \otimes B_{2S} \\ (B_{1T} l_T) \otimes [B_{2S} d(l_{m_i})] & (B_{1T} l_T) \otimes B_{2S} & B_{1T} \otimes B_{2S} \end{pmatrix}.$$

Under Assumptions 1 (a) and (c), $\mathrm{E}[\zeta] = 0$, and

$$\Omega_\zeta \equiv \mathrm{E}[\zeta\zeta'] = \begin{pmatrix} \sigma_\alpha^2 I_N & 0 & 0 \\ 0 & \sigma_\mu^2 I_S & 0 \\ 0 & 0 & \sigma_\nu^2 I_{TS} \end{pmatrix}.$$

For $i, j = 1, 2, \ldots, N + S + TS$, let the elements of Λ and Ω_ζ be denoted by Λ_{ij} and $\Omega_{\zeta,ij}$, respectively. Since the row and column sums of B_{1T} and B_{2S} are uniformly bounded in absolute value by hypothesis, and the row and column sums of $d(l_{m_i})$ are uniformly bounded in absolute value under Assumption 1 (d), then by Lemma 1 (a) and (b), the row and column sums of each submatrix of Λ are uniformly bounded in absolute value, and thus the row and column sums of Λ are uniformly bounded in absolute value. Consequently, there exists $c_\Lambda > 0$ such that $|\Lambda_{ij}| < c_\Lambda$ for all i, j, uniformly over N and S. Therefore, under Assumptions 1 (b) and (e),

$$|\mathrm{E}[Z]| = \left|\frac{\mathrm{tr}[\Lambda \mathrm{E}(\zeta\zeta')]}{\varphi_{N,S}}\right| = \left|\frac{1}{\varphi_{N,S}}\sum_{i=1}^{N+S+TS}\Lambda_{ii}\Omega_{\zeta,ii}\right| \le \frac{1}{\varphi_{N,S}}\sum_{i=1}^{N+S+TS} c_\Lambda \overline{\sigma}_2 = \frac{N+S+TS}{\varphi_{N,S}} c_\Lambda \overline{\sigma}_2 = O(1).$$

Next, for the variance of Z, using the result for variance of quadratic forms given in Kelejian and Prucha (2001), we get

$$\mathrm{Var}[Z] = \frac{1}{2\varphi_{N,S}^2}\{\mathrm{tr}[(\Lambda + \Lambda')\Omega_\zeta(\Lambda + \Lambda')\Omega_\zeta] + \sum_{i=1}^{N+S+TS}\Lambda_{ii}^2[\mathrm{E}(\zeta_i^4) - 3\mathrm{Var}^2(\zeta_i)]\}.$$

Let $\overline{\sigma} = \max\{\overline{\sigma}_4, \overline{\sigma}_2, \overline{\sigma}_2^2\}$. By Lemma 1 (b), the row and column sums of the sequence of matrices $(\Lambda + \Lambda')\Omega_\zeta(\Lambda + \Lambda')\Omega_\zeta$ are bounded uniformly in absolute value, say, by a constant $c < \infty$, and thus each of its elements is bounded in absolute value by c. By the triangle inequality, we obtain

$$\mathrm{Var}[Z] \le \frac{1}{2\varphi_{N,S}^2}\{|\mathrm{tr}[(\Lambda + \Lambda')\Omega_\zeta(\Lambda + \Lambda')\Omega_\zeta]| + \sum_{i=1}^{N+S+TS}\Lambda_{ii}^2[\mathrm{E}(\zeta_i^4) + 3\mathrm{Var}^2(\zeta_i)]\}$$
$$\le \frac{N+S+TS}{2\varphi_{N,S}^2}(c + 4c_\Lambda^2\overline{\sigma}) = o(1), \text{ as } N, S \to \infty,$$

under Assumption 1 (e). Finally, $Z - \mathrm{E}[Z] = o_P(1)$ follows from Chebyshev's inequality. ■

Lemma 3. Let Γ^* and γ^* be identical to Γ and γ except that the expectation operator is dropped. Under Assumptions 1 and 2 (d), $\Gamma = O(1)$, $\gamma = O(1)$, and $(\Gamma^*, \gamma^*) - (\Gamma, \gamma) = o_P(1)$.

Proof. First, we study the terms in Γ and γ that do not depend on the moments of quadratic forms. It is easy to see from the expressions of Γ and γ that the terms γ_{23}^1, γ_{24}^1, γ_{34}^1, γ_{15}^1, γ_{25}^1, γ_{35}^1, γ_{23}^2, γ_{33}^2, γ_{53}^2, γ_{14}^2, γ_{24}^2, γ_{34}^2, γ_{54}^2, γ_{64}^2, γ_{15}^2, γ_{25}^2, γ_{35}^2, γ_{45}^2, γ_{55}^2, γ_{65}^2, γ_{23}^3, γ_{33}^3, γ_{53}^3, γ_{63}^3, γ_{14}^3, γ_{24}^3, γ_{34}^3, γ_{44}^3, γ_{54}^3, γ_{64}^3, γ_{15}^3, γ_{25}^3, γ_{35}^3, γ_{45}^3, γ_{55}^3, and γ_{65}^3 have such property. As an illustration, we show that $\gamma_{64}^2 = O(1)$, and other terms follow similar arguments. $\gamma_{64}^2 = \frac{1}{S-N}\mathrm{tr}[(W_S' - W_S^{*\prime})W_S J^*]$, under Assumption 2 (d), the row and column sums of $(W_S' - W_S^{*\prime})$ and J^* are uniformly bounded in absolute value. By Lemma 1 (b), the row and column sums of $(W_S' - W_S^{*\prime})W_S J^*$ are uniformly bounded in absolute value, say, by a constant $c < \infty$, and the absolute value of each of its elements bounded by c, then $|\gamma_{64}^2| \le \frac{Sc}{S-N} = O(1)$ under Assumption 1 (e). Moreover, the claim $(\Gamma^*, \gamma^*) - (\Gamma, \gamma) = o_P(1)$ for these terms is trivial.

Second, we study other terms in Γ and γ that depend on moments of quadratic forms. Let $A_S = \frac{I_S + \theta W_S}{I_S - \rho W_S}$. Recall that $u = (I_T \otimes A_S)\epsilon$, $\overline{u} = (I_T \otimes W_S)u = [I_T \otimes (W_S A_S)]\epsilon$, and $\overline{\overline{u}} = (I_T \otimes W_S)\overline{u} = [I_T \otimes (W_S^2 A_S^{-1})]\epsilon$, then it is clear that for $s = 1,2,3$,

$$Z_{1s} \equiv u' Q_s u = \epsilon'(I_T \otimes A_S')Q_s(I_T \otimes A_S)\epsilon,$$

$$Z_{2s} \equiv \overline{u}' Q_s u = \epsilon'[I_T \otimes (W_S A_S)']Q_s(I_T \otimes A_S)\epsilon,$$

$$Z_{3s} \equiv \overline{\overline{u}}' Q_s u = \epsilon'[I_T \otimes (W_S^2 A_S)']Q_s(I_T \otimes A_S)\epsilon,$$

$$Z_{4s} \equiv \overline{u}' Q_s \overline{u} = \epsilon'[I_T \otimes (W_S A_S)']Q_s[I_T \otimes (W_S A_S)]\epsilon,$$

$$Z_{5s} \equiv \overline{\overline{u}}' Q_s \overline{u} = \epsilon'[I_T \otimes (W_S^2 A_S)']Q_s[I_T \otimes (W_S A_S)]\epsilon,$$

$$Z_{6s} \equiv \overline{\overline{u}}' Q_s \overline{\overline{u}} = \epsilon'[I_T \otimes (W_S^2 A_S)']Q_s[I_T \otimes (W_S^2 A_S)]\epsilon.$$

All the above terms are of the form $\epsilon'(B_{1T} \otimes B_{2S})\epsilon$, which falls into the framework of Lemma 2. We show it in detail for the term γ_3^1, and other terms follow a similar argument. Recall that $\gamma_3^1 = \frac{1}{(T-1)S}\mathrm{E}[\overline{u}' Q_1 \overline{u}]$, and write

$$\gamma_3^{1*} \equiv \frac{1}{(T-1)S}(\overline{u}' Q_1 \overline{u}) = \frac{1}{(T-1)S}\epsilon'\{E_T \otimes (W_S A_S)' A_S\}\epsilon.$$

Since E_T is a constant matrix with a fixed dimension $T \times T$, and the row and column sums of $(W_S A_S)' A_S$ are uniformly bounded in absolute value due to Assumption 2 (d) and Lemma 1 (b). Then by taking $\varphi_{N,S} = S$ and ignoring the constant $T-1$, Lemma 2 applies and we obtain that $\gamma_3^1 = O(1)$ and $\gamma_3^{1*} - \gamma_3^1 = o_P(1)$. ■

Lemma 4. Under Assumptions 1, 2 (d), and 3 (a), $(\widehat{\Gamma}, \widehat{\gamma}) - (\Gamma^*, \gamma^*) = o_P(1)$.

Proof. First, for those terms in Γ and γ that do not depend on the moments of quadratic forms, the above claim is trivial. Second, for those terms in Γ and γ that depend on the moments of quadratic forms, we first have that $\overline{u} = Wu$, $\overline{\overline{u}} = W^2 u$, $\widetilde{u} = y - X\widetilde{\beta} = X(\beta - \widetilde{\beta}) + u$, $\widetilde{\overline{u}} = W[X(\beta - \widetilde{\beta}) + u]$, and $\widetilde{\overline{\overline{u}}} = W^2[X(\beta - \widetilde{\beta}) + u]$. Then for $s = 1,2,3$, we can write

$$\frac{1}{\varphi_{N,S}}(\widetilde{u}' Q_s \widetilde{u} - u' Q_s u) = (\widetilde{\beta} - \beta)' \frac{X' Q_s X}{\varphi_{N,S}}(\widetilde{\beta} - \beta) - \frac{u'(Q_s + Q_s)X}{\varphi_{N,S}}(\widetilde{\beta} - \beta),$$

$$\frac{1}{\varphi_{N,S}}(\widetilde{\overline{u}}' Q_s \widetilde{u} - \overline{u}' Q_s u) = (\widetilde{\beta} - \beta)' \frac{X' W' Q_s X}{\varphi_{N,S}}(\widetilde{\beta} - \beta) - \frac{u'(Q_s W + W' Q_s)X}{\varphi_{N,S}}(\widetilde{\beta} - \beta),$$

$$\frac{1}{\varphi_{N,S}}(\widetilde{\overline{\overline{u}}}' Q_s \widetilde{u} - \overline{\overline{u}}' Q_s u) = (\widetilde{\beta} - \beta)' \frac{X' W^{2\prime} Q_s X}{\varphi_{N,S}}(\widetilde{\beta} - \beta) - \frac{u'(Q_s W^2 + W^{2\prime} Q_s)X}{\varphi_{N,S}}(\widetilde{\beta} - \beta),$$

$$\frac{1}{\varphi_{N,S}}(\widetilde{\overline{u}}' Q_s \widetilde{\overline{u}} - \overline{u}' Q_s \overline{u}) = (\widetilde{\beta} - \beta)' \frac{X' W' Q_s W X}{\varphi_{N,S}}(\widetilde{\beta} - \beta) - \frac{u'(W' Q_s W + W^{2\prime} Q_s W)X}{\varphi_{N,S}}(\widetilde{\beta} - \beta),$$

$$\frac{1}{\varphi_{N,S}}(\widetilde{\overline{\overline{u}}}' Q_s \widetilde{\overline{u}} - \overline{\overline{u}}' Q_s \overline{u}) = (\widetilde{\beta} - \beta)' \frac{X' W^{2\prime} Q_s W X}{\varphi_{N,S}}(\widetilde{\beta} - \beta) - \frac{u'(W' Q_s W^2 + W^{2\prime} Q_s W)X}{\varphi_{N,S}}(\widetilde{\beta} - \beta),$$

$$\frac{1}{\varphi_{N,S}}(\widetilde{\overline{\overline{u}}}' Q_s \widetilde{\overline{\overline{u}}} - \overline{\overline{u}}' Q_s \overline{\overline{u}}) = (\widetilde{\beta} - \beta)' \frac{X' W^{2\prime} Q_s W^2 X}{\varphi_{N,S}}(\widetilde{\beta} - \beta) - \frac{u'(W^{2\prime} Q_s W^2 + W^{2\prime} Q_s W^2)X}{\varphi_{N,S}}(\widetilde{\beta} - \beta).$$

All of the above terms are in the form of

$$(\widetilde{\beta} - \beta)' \frac{X' H X}{\varphi_{N,S}}(\widetilde{\beta} - \beta) - \frac{u'(H + H')X}{\varphi_{N,S}}(\widetilde{\beta} - \beta),$$

where H is a $TS \times TS$ matrix of constants. Given $\widetilde{\beta} - \beta = o_P(1)$, to prove that the above term is $o_P(1)$, it suffices to show (i) $\frac{X'HX}{\varphi_{N,S}} = O(1)$; (ii) $\frac{u'(H+H')X}{\varphi_{N,S}} = o_P(1)$. For (i), it is obvious that the row and column sums of Q_1, Q_2, and Q_3 are bounded in absolute value by $2(1-\frac{1}{T})$, $\max_i 2(1-\frac{1}{m_i})$, and 1, respectively. Then $\frac{X'HX}{\varphi_{N,S}} = O(1)$ follows from Assumptions 1 (d), 2 (d), 3 (a) and repeated use of Lemma 1.

For (ii), we have

$$\mathrm{Var}[\frac{u'(H+H')X}{\varphi_{N,S}}] \overset{(1)}{=} \frac{S}{\varphi_{N,S}^2}\frac{X'(H+H')\Omega_u(H+H')X}{S} \overset{(2)}{=} \frac{S}{\varphi_{N,S}^2}O(1) \overset{(3)}{=} o(1), \text{ as } N,S \to \infty.$$

In the above, (1) is from Assumption 1 (a). (2) is from that the row and column sums of $H+H'$ and Ω_u are uniformly bounded in absolute value, Assumption 3 (a), and Lemma 1 (c), where (i) the row and column sums of $H+H'$ are uniformly bounded in absolute value by Assumptions 1 (d), 2 (d), 3 (a), and Lemma 1; (ii) the row and column sums of Ω_u are uniformly bounded in absolute value due to the fact that the row and column sums of Ω_ϵ are uniformly bounded in absolute value by $\max_i \xi_{3i}$, Assumptions 1 (b)-(d), 2 (d), and Lemma 1 (a). (3) is from Assumption 1 (e). Therefore, as $\mathrm{E}[\frac{u'(H+H')X}{\varphi_{N,S}}] = 0$ under Assumption 1 (a), then $\frac{u'(H+H')X}{\varphi_{N,S}} = o_P(1)$ by Chebyshev's inequality. Finally,

$$(\widetilde{\beta}-\beta)'\frac{X'HX}{\varphi_{N,S}}(\widetilde{\beta}-\beta) - \frac{u'(H+H')X}{\varphi_{N,S}}(\widetilde{\beta}-\beta) = o_P(1)O(1)o_P(1) - o_P(1)o_P(1) = o_P(1).$$

Next, we apply the above discussion to the term $\hat{\gamma}_3^1 - \gamma_3^{1*}$, where

$$\hat{\gamma}_3^1 - \gamma_3^{1*} = \frac{1}{(T-1)S}[\widetilde{\overline{u}}'Q_1\widetilde{u} - \overline{u}'Q_1 u] = (\widetilde{\beta}-\beta)'\frac{X'W'Q_sX}{\varphi_{N,S}}(\widetilde{\beta}-\beta) - \frac{u'(Q_sW+W'Q_s)X}{\varphi_{N,S}}(\widetilde{\beta}-\beta) = o_P(1),$$

and the results for other terms follow a similar argument. ■

Lemma 5. Let $D = \frac{ABC}{TS}$, where B is either $I_T \otimes d((\hat{\xi}_{3i}^{-1} - \xi_{3i}^{-1})I_{m_i})$ or $I_T \otimes d((\hat{\xi}_{3i}^{-2} - \xi_{3i}^{-2})I_{m_i})$, both of which are of dimension $TS \times TS$, A is of dimension $K \times TS$, and C is of dimension $TS \times K$. Elements of A and C are uniformly bounded in absolute value by constants a and c, respectively. Let the (i,j) -th element of D be D_{ij}, then $|D_{ij}| = o_P(1)$.

Proof. Let the (i,j) -th element of D be D_{ij}, and similarly apply for A, B, and C. Then for $i,j = 1,\dots,K$,

$$\begin{aligned}|D_{ij}| &= \frac{1}{TS}\left|\sum_{k=1}^{TS}\sum_{l=1}^{TS}A_{ik}B_{kl}C_{lj}\right|\\ &\overset{(1)}{\le} \frac{1}{TS}\sum_{k=1}^{TS}|A_{ik}||B_{kk}||C_{kj}|\\ &\overset{(2)}{\le} \frac{ac}{TS}\sum_{k=1}^{TS}|B_{kk}|\\ &\overset{(3)}{=} \frac{ac}{TS}\sum_{k=1}^{TS}O_P((TS)^{-\frac{1}{2}}) = acO_P((TS)^{-\frac{1}{2}}) = o_P(1),\end{aligned}$$

where (1) is by the triangle inequality and that B is a diagonal matrix, (2) is by the uniform boundedness of elements of A and C, and (3) is by the standard parametric rate that $\hat{\xi}_{3i}^{-1} - \xi_{3i}^{-1} = O_P((TS)^{-\frac{1}{2}})$. The fact that $|D_{ij}| = o_P(1)$ for $i,j = 1,\dots,K$ if $B = I_T \otimes d((\hat{\xi}_{3i}^{-2} - \xi_{3i}^{-2})I_{m_i})$ follows exactly the same argument. ■

C Proofs of theorems

Proof (Theorem 1) For $\widetilde{\beta} = (X'X)^{-1}(X'y)$, it is easy to see that
$\mathrm{Var}(\widetilde{\beta}) = \frac{1}{S}(\frac{X'X}{S})^{-1}\frac{X'\Omega_uX}{S}(\frac{X'X}{S})^{-1} \overset{(1)}{=} \frac{1}{S}O(1)O(1)O(1) = o(1)$, as $N,S \to \infty$,
where (1) follows from Assumptions 1 (a)-(d), 2 (d), 3 (a), (b), and Lemma 1. Moreover, $\mathrm{E}[\widetilde{\beta}] = \beta$ under Assumptions 1 (a) and (c). Then the element-wise application of Chebyshev's inequality to $\widetilde{\beta} - \mathrm{E}[\widetilde{\beta}]$ yields $\widetilde{\beta} - \beta = o_P(1)$. ■

Proof (Theorem 2) It follows the same logic as the Proof of Theorem 3 after replacing both Ξ^{-1} and $\widetilde{\Xi}^{-1}$ in Theorem 3 by the identity matrix.■

Proof (Theorem 3) We now let the true parameter vector be $\varphi = (\rho, \rho^2, \xi_1, \xi_2, \xi_3, \theta\xi_1, \theta\xi_2, \theta\xi_3,\ \theta^2\xi_1, \theta^2\xi_2, \theta^2\xi_3)'$, and the generic parameter vector be $\underline{\varphi} = (\underline{\rho}, \underline{\rho}^2, \underline{\xi}_1, \underline{\xi}_2, \underline{\xi}_3, \underline{\theta\xi}_1, \underline{\theta\xi}_2, \underline{\theta\xi}_3, \underline{\theta}^2\underline{\xi}_1,\ \underline{\theta}^2\underline{\xi}_2, \underline{\theta}^2\underline{\xi}_3)'$, with the parameter space $\Theta = [-a,a] \times [0,a^2] \times [0,c] \times [0,d] \times [0,e] \times [-bc,bc] \times [-bd,bd] \times [-be,be] \times [0,b^2c] \times [0,b^2d] \times [0,b^2e]$, where $0 < a,b,c,d,e < \infty$. Let the population and sample criterion functions for the optimal GMM estimator be

$$Q(\underline{\psi}) = (\Gamma\underline{\varphi} - \gamma)'\Xi^{-1}(\Gamma\underline{\varphi} - \gamma),$$

$$\overline{Q}(\underline{\psi}) = (G\underline{\varphi} - g)'\widetilde{\Xi}^{-1}(G\underline{\varphi} - g).$$

To prove consistency, it suffices to show (i) ψ is the unique zero of $Q(\underline{\psi}) = 0$ in Δ; (ii) $\sup_{\underline{\psi}\in\Delta}|\overline{Q}(\underline{\psi}) - Q(\underline{\psi})| = o_P(1)$.

Step (*i*): First, $Q(\psi) = 0$ by definition. Second, given any $\underline{\varphi} \in \Theta$ such that $\underline{\varphi} \ne \varphi$, let $\underline{\varphi} - \varphi = (\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\xi}_1 - \xi_1, \underline{\xi}_2 - \xi_2, \underline{\xi}_3 - \xi_3, \underline{\theta\xi}_1 - \theta\xi_1, \underline{\theta\xi}_2 - \theta\xi_2, \underline{\theta\xi}_3 - \theta\xi_3, \underline{\theta}^2\underline{\xi}_1 - \theta^2\xi_1, \underline{\theta}^2\underline{\xi}_2 - \theta^2\xi_2, \underline{\theta}^2\underline{\xi}_3 - \theta^2\xi_3)'$, we have

$Q(\underline{\psi}) - Q(\psi) = (\underline{\varphi} - \varphi)'\Gamma'\Xi^{-1}\Gamma(\underline{\varphi} - \varphi) \geq \lambda_{\min}(\Gamma'\Xi^{-1}\Gamma)\|\underline{\varphi} - \varphi\|^2 \geq \underline{\lambda}\|\underline{\varphi} - \varphi\|^2,$

where the last inequality is by Assumption 4 (b). Therefore, 0 is a well-separated minimum of $Q(\cdot)$, attained at ψ.

Step (ii): Let $\widehat{\Phi} = [G, -g]$ and $\Phi = [\Gamma, -\gamma]$, we can write

$$Q(\underline{\psi}) = (\underline{\varphi}', 1)\Phi'\Xi^{-1}\Phi(\underline{\varphi}', 1)',$$

$$\widehat{Q}(\underline{\psi}) = (\underline{\varphi}', 1)\widehat{\Phi}'\widetilde{\Xi}^{-1}\widehat{\Phi}(\underline{\varphi}', 1)'.$$

Then, the following inequality

$$\begin{aligned}|\overline{Q}(\underline{\psi}) - Q(\underline{\psi})| &= |(\underline{\varphi}', 1)(\widehat{\Phi}'\widetilde{\Xi}^{-1}\widehat{\Phi} - \Phi'\Xi^{-1}\Phi)(\underline{\varphi}', 1)'| \\ &\overset{(1)}{\leq} \|\widehat{\Phi}'\widetilde{\Xi}^{-1}\widehat{\Phi} - \Phi'\Xi^{-1}\Phi\|_2\|(\underline{\varphi}', 1)\|^2 \\ &\leq \|\widehat{\Phi}'\widetilde{\Xi}^{-1}\widehat{\Phi} - \Phi'\Xi^{-1}\Phi\|_2[1 + a^2 + a^4 + (1 + b^2 + b^4)(c^2 + d^2 + e^2)]\end{aligned}$$

holds for any $\underline{\psi} \in \Delta$, where (1) is based on the Cauchy-Schwarz inequality and the definition of matrix 2-norm.

We need to prove $\widehat{\Phi}'\ \widetilde{\Xi}^{-1}\widehat{\Phi} - \Phi'\Xi^{-1}\Phi = o_P(1)$. Due to the facts that (1) $\widetilde{\Xi}^{-1} - \Xi^{-1} = o_P(1)$ by hypothesis; (2) $\widehat{\Phi} - \Phi = o_P(1)$ by Assumptions 1, 2 (d), 3 (a), and Lemma 2; (3)

$$\widehat{\Phi}'\widetilde{\Xi}^{-1}\widehat{\Phi} - \Phi'\Xi^{-1}\Phi = (\widehat{\Phi} - \Phi)'(\widetilde{\Xi}^{-1}\widehat{\Phi} - \Xi^{-1}\Phi) + \Phi'(\widetilde{\Xi}^{-1}\widehat{\Phi} - \Xi^{-1}\Phi) + (\widehat{\Phi} - \Phi)'\Xi^{-1}\Phi,$$

$$\widetilde{\Xi}^{-1}\widehat{\Phi} - \Xi^{-1}\Phi = (\widetilde{\Xi}^{-1} - \Xi^{-1})(\widehat{\Phi} - \Phi) + \Xi^{-1}(\widehat{\Phi} - \Phi) + (\widetilde{\Xi}^{-1} - \Xi^{-1})\Phi,$$

it suffices to show that $\Phi = O(1)$, $\Xi^{-1} = O(1)$, and consequently, $\sup_{\underline{\psi}\in\Delta} |\overline{Q}(\underline{\psi}) - Q(\underline{\psi})| \leq \|\widehat{\Phi}'\ \widetilde{\Xi}^{-1}\widehat{\Phi} - \Phi'\Xi^{-1}\Phi\|_2[1 + a^2 + a^4 + (1 + b^2 + b^4)(c^2 + d^2 + e^2)] = o_P(1)$. Putting Steps (i) and (ii) together, the claim of the Theorem follows from Lemma 3.1 in Pötscher and Prucha (1997), and the consistency of original parameters follows from the continuous mapping theorem.

Lastly, $\Phi = O(1)$ is by Assumptions 1, 2 (d), 3 (a), and Lemma 2, and we show that $\Xi^{-1} = O(1)$ holds under Assumption 4 (b). For $i, j = 1, \ldots, 9$, let Ξ_{ij}^{-1} denote the (i,j)-th element of Ξ^{-1}. To show that $\Xi^{-1} = O(1)$, it suffices to demonstrate that $|\Xi_{ij}^{-1}|$ is bounded above for all $i, j = 1, \ldots, 9$, uniformly in N, S. From Rao (1973, p. 62), we have

$$\lambda_{\min}(\Xi^{-1}) = \inf_{x\neq 0}\frac{x'\Xi^{-1}x}{x'x}, \lambda_{\max}(\Xi^{-1}) = \sup_{x\neq 0}\frac{x'\Xi^{-1}x}{x'x}.$$

It follows from Assumption 4 (b) that for any $x \in \mathbb{R}^9$ and $x \neq 0$, $0 < \underline{\lambda} \leq \frac{x'\Xi^{-1}x}{x'x} \leq \overline{\lambda} < \infty$. In particular, let x be a vector with ones at the i-th and j-th positions and zeros elsewhere, we get

$$0 < \Xi_{ii}^{-1} \leq \overline{\lambda} < \infty, i = j,$$

$$0 < (\Xi_{ii}^{-1} + \Xi_{jj}^{-1} + 2\Xi_{ij}^{-1})/2 \leq \overline{\lambda} < \infty, i \neq j.$$

The above two inequalities easily imply that $|\Xi_{ij}^{-1}| \leq \overline{\lambda}$ for all $i, j = 1, \ldots, 9$, uniformly in N, S. ■

Proof (Theorem 4)

(a) Let $\widehat{\beta}_{\mathrm{GLS}} = (X'\Omega_u^{-1}X)^{-1}(X'\Omega_u^{-1}y)$, then

$$\begin{aligned}(TS)^{\frac{1}{2}}(\widehat{\beta}_{\mathrm{GLS}} - \beta) &= (TS)^{\frac{1}{2}}[(X'\Omega_u^{-1}X)^{-1}(X'\Omega_u^{-1}y) - \beta] \\ &= (TS)^{\frac{1}{2}}(X'\Omega_u^{-1}X)^{-1}(X'\Omega_u^{-1}u) \\ &= \Sigma_1^{-1}(TS)^{-\frac{1}{2}}[X^{*\prime}(\rho,\theta)\Omega_\epsilon^{-1}\epsilon],\end{aligned}$$

where $\Sigma_1 = \frac{X^{*\prime}(\rho,\theta)\Omega_\epsilon^{-1}X^*(\rho,\theta)}{TS}$ is of dimension $K \times K$, and it follows from Assumption 3 (b) that $\Sigma_1^{-1} - \Sigma_{1,\infty}^{-1} = o(1)$, as $N, S \to \infty$.

The second term can be written as

$$X^{*\prime}(\rho,\theta)\Omega_\epsilon^{-1}\epsilon = X_\alpha^{*\prime}\alpha + X_\mu^{*\prime}\mu + X_\nu^{*\prime}\nu,$$

where

$$\begin{aligned}X_\alpha^{*\prime} &= X^{*\prime}(\rho,\theta)(l_T \otimes (d(\xi_{3i}^{-1}l_{m_i})), X_\nu^{*\prime} = X^{*\prime}(\rho,\theta)\Omega_\epsilon^{-1}, \\ X_\mu^{*\prime} &= X^{*\prime}(\rho,\theta)[\xi_2^{-1}(l_T \otimes d(E_{m_i})) + (l_T \otimes d(\xi_{3i}^{-1}\overline{J}_{m_i}))].\end{aligned} \tag{C.1}$$

Elements in $X^*(\rho,\theta)$ are uniformly bounded in absolute value by Assumptions 2 (d), 3 (a), and Lemma 1. Similarly, it becomes clear that the elements of $X_\alpha^{*\prime}$, $X_\mu^{*\prime}$, and $X_\nu^{*\prime}$ are uniformly bounded in absolute value by Assumptions 1 (a)-(d) and Lemma 1.

Let $\Lambda_\alpha = \frac{\sigma_\alpha^2 X_\alpha^{*\prime}X_\alpha^*}{TS}$, $\Lambda_\mu = \frac{\sigma_\mu^2 X_\mu^{*\prime}X_\mu^*}{TS}$, and $\Lambda_\nu = \frac{\sigma_\nu^2 X_\nu^{*\prime}X_\nu^*}{TS}$. Under Assumptions 1 (a), (b), and 3, applying Theorem 30 in Pötscher and Prucha (2001) yields

$$(TS)^{-\frac{1}{2}}X_\alpha^{*\prime}\alpha \xrightarrow{d} N(0, \Lambda_{\alpha,\infty}), \Lambda_{\alpha,\infty} = \lim_{S\to\infty}\Lambda_\alpha,$$

$$(TS)^{-\frac{1}{2}} X_{\mu}^{*\prime} \mu \xrightarrow{d} N(0, \Lambda_{\mu,\infty}), \Lambda_{\mu,\infty} = \lim_{S\to\infty} \Lambda_{\mu},$$

$$(TS)^{-\frac{1}{2}} X_{\nu}^{*\prime} \nu \xrightarrow{d} N(0, \Lambda_{\nu,\infty}), \Lambda_{\nu,\infty} = \lim_{S\to\infty} \Lambda_{\nu},$$

where the convergence results for $(TS)^{-\frac{1}{2}} X_{\mu}^{*\prime}\mu$ and $(TS)^{-\frac{1}{2}} X_{\nu}^{*\prime}\nu$ are immediate. Meanwhile, the convergence result for $(TS)^{-\frac{1}{2}} X_{\alpha}^{*\prime}\alpha$ is due to Assumption 1 (e) and the fact that $N^{-\frac{1}{2}} X_{\alpha}^{*\prime}\alpha \xrightarrow{d} N(0, \lim_{N\to\infty} \frac{\sigma_{\alpha}^2 X_{\alpha}^{*\prime} X_{\alpha}^{*}}{N})$.

Therefore, under Assumption 1 (c),

$$(TS)^{-\frac{1}{2}} X^{*\prime}(\rho,\theta)\Omega_{\epsilon}^{-1}\epsilon \xrightarrow{d} N(0, \Lambda_{\infty}),$$

where $\Lambda_{\infty} = \Lambda_{\alpha,\infty} + \Lambda_{\mu,\infty} + \Lambda_{\nu,\infty}$. Finally, by the Slutsky's Theorem,

$$(TS)^{\frac{1}{2}} (\hat{\beta}_{\mathrm{GLS}} - \beta) \xrightarrow{d} N(0, \Lambda_{\infty}^{*}),$$

where $\Lambda_{\infty}^{*} = \Sigma_{1,\infty}^{-1} \Lambda_{\infty} (\Sigma_{1,\infty}^{-1})'$.

(b) To show $(TS)^{\frac{1}{2}}(\hat{\beta}_{\mathrm{FGLS}} - \hat{\beta}_{\mathrm{GLS}}) = o_P(1)$, it is conspicuous that

$$(TS)^{\frac{1}{2}}(\hat{\beta}_{\mathrm{FGLS}} - \hat{\beta}_{\mathrm{GLS}}) = \hat{\Sigma}_1^{-1}\hat{\Sigma}_2 - \Sigma_1^{-1}\Sigma_2,$$

where $\Sigma_1 = \frac{X^{*\prime}(\rho,\theta)\Omega_{\epsilon}^{-1}X^{*}(\rho,\theta)}{TS}$, $\hat{\Sigma}_1 = \frac{X^{*\prime}(\hat{\rho},\hat{\theta})\hat{\Omega}_{\epsilon}^{-1}X^{*}(\hat{\rho},\hat{\theta})}{TS}$, $\Sigma_2 = \frac{X^{*\prime}(\rho,\theta)\Omega_{\epsilon}^{-1}u^{*}(\rho,\theta)}{(TS)^{1/2}}$, $\hat{\Sigma}_2 = \frac{X^{*\prime}(\hat{\rho},\hat{\theta})\hat{\Omega}_{\epsilon}^{-1}u^{*}(\hat{\rho},\hat{\theta})}{(TS)^{1/2}}$, $\hat{\Omega}_{\epsilon}^{-1} = \hat{\xi}_1^{-1}Q_1 + \hat{\xi}_2^{-1}Q_2 + (I_T \otimes d(\hat{\xi}_{3i}^{-1} I_{m_i}))Q_3$, and $u^{*}(\rho,\theta) = (I_{TS} + \theta W)^{-1}(I_{TS} - \rho W)u$.

Due to the following decomposition

$$\hat{\Sigma}_1^{-1}\hat{\Sigma}_2 - \Sigma_1^{-1}\Sigma_2 = (\hat{\Sigma}_1^{-1} - \Sigma_1^{-1})(\hat{\Sigma}_2 - \Sigma_2) + (\hat{\Sigma}_1^{-1} - \Sigma_1^{-1})\Sigma_2 + (\Sigma_1^{-1} - \Sigma_{1,\infty}^{-1})(\hat{\Sigma}_2 - \Sigma_2) + \Sigma_{1,\infty}^{-1}(\hat{\Sigma}_2 - \Sigma_2), \tag{C.2}$$

and the fact that $\Sigma_1^{-1} - \Sigma_{1,\infty}^{-1} = o(1)$, it suffices to show $\Sigma_{1,\infty}^{-1} = O(1)$, $\Sigma_2 = O_P(1)$, $\hat{\Sigma}_1 - \Sigma_1 = o_P(1)$, and $\hat{\Sigma}_2 - \Sigma_2 = o_P(1)$.

$\Sigma_{1,\infty}^{-1} = O(1)$ follows from Assumption 3 (b), and $\Sigma_2 = O_P(1)$ follows from $\mathrm{E}[\Sigma_2] = 0$ and $\mathrm{Var}[\Sigma_2] = (X'\Omega_u^{-1}X)/(TS) = O(1)$ under Assumptions 1 (a)-(d), 2 (d), 3 (a), and by Lemma 1.

Now we show $\hat{\Sigma}_1 - \Sigma_1 = o_P(1)$, and $\hat{\Sigma}_2 - \Sigma_2 = o_P(1)$. Let $\vartheta = \frac{I_{TS} - \rho W}{I_{TS} + \theta W}$, $\hat{\vartheta} = \frac{I_{TS} - \hat{\rho} W}{I_{TS} + \hat{\theta} W}$. Obviously,

$$X^{*}(\hat{\rho},\hat{\theta}) - X^{*}(\rho,\theta) = \frac{(I_{TS} - \hat{\rho}W)X}{I_{TS} + \hat{\theta}W} - \frac{(I_{TS} - \rho W)X}{I_{TS} + \theta W} = (\hat{\vartheta} - \vartheta)X,$$
$$u^{*}(\hat{\rho},\hat{\theta}) - u^{*}(\rho,\theta) = \frac{(I_{TS} - \hat{\rho}W)u}{I_{TS} + \hat{\theta}W} - \frac{(I_{TS} - \rho W)u}{I_{TS} + \theta W} = (\hat{\vartheta} - \vartheta)u,$$

$$\hat{\Omega}_{\epsilon}^{-1} - \Omega_{\epsilon}^{-1} = (\hat{\xi}_1^{-1} - \xi_1^{-1})Q_1 + (\hat{\xi}_2^{-1} - \xi_2^{-1})Q_2 + [I_T \otimes d((\hat{\xi}_{3i}^{-1} - \xi_{3i}^{-1})I_{m_i})]Q_3 \equiv \hat{\Delta}.$$

First,

$$\begin{aligned}
\hat{\Sigma}_1 - \Sigma_1 &= \frac{X^{*\prime}(\hat{\rho},\hat{\theta})\hat{\Omega}_{\epsilon}^{-1}X^{*}(\hat{\rho},\hat{\theta})}{TS} - \frac{X^{*\prime}(\rho,\theta)\Omega_{\epsilon}^{-1}X^{*}(\rho,\theta)}{TS} \\
&= \frac{[(\hat{\vartheta} - \vartheta)X + X^{*}(\rho,\theta)]'(\Omega_{\epsilon}^{-1} + \hat{\Delta})[(\hat{\vartheta} - \vartheta)X + X^{*}(\rho,\theta)]}{TS} - \frac{X^{*\prime}(\rho,\theta)\Omega_{\epsilon}^{-1}X^{*}(\rho,\theta)}{TS} \\
&= \left[\frac{(\hat{\vartheta} - \vartheta)^2 X'\Omega_{\epsilon}^{-1}X}{TS} + \frac{(\hat{\vartheta} - \vartheta)X'\Omega_{\epsilon}^{-1}X^{*}(\rho,\theta)}{TS} + \frac{(\hat{\vartheta} - \vartheta)X^{*\prime}(\rho,\theta)\Omega_{\epsilon}^{-1}X}{TS}\right] \\
&\quad + \left[\frac{(\hat{\vartheta} - \vartheta)^2 X'\hat{\Delta}X}{TS} + \frac{(\hat{\vartheta} - \vartheta)X'\hat{\Delta}X^{*}(\rho,\theta)}{TS} + \frac{(\hat{\vartheta} - \vartheta)X^{*\prime}(\rho,\theta)\hat{\Delta}X}{TS} + \frac{X^{*\prime}(\theta)\hat{\Delta}X^{*}(\rho,\theta)}{TS}\right] \\
&\equiv \left[(\hat{\vartheta} - \vartheta)^2 \mathcal{I}_{11} + (\hat{\vartheta} - \vartheta)(\mathcal{I}_{12} + \mathcal{I}_{12}')] + [(\hat{\vartheta} - \vartheta)^2 \mathcal{I}_{13} + (\hat{\vartheta} - \vartheta)(\mathcal{I}_{14} + \mathcal{I}_{14}') + \mathcal{I}_{15}\right].
\end{aligned}$$

$\mathcal{I}_{11} = o_P(1)$ by Assumptions 1 (a)-(d), 2 (d), 3 (a), and Lemma 1. A similar argument shows that $\mathcal{I}_{12} = o_P(1)$. For $\mathcal{I}_{13}$, we have

$$\begin{aligned}
\mathcal{I}_{13} &= (\hat{\xi}_1^{-1} - \xi_1^{-1})\frac{X'Q_1X}{TS} + (\hat{\xi}_2^{-1} - \xi_2^{-1})\frac{X'Q_2X}{TS} + \frac{X'[I_T \otimes d((\hat{\xi}_{3i}^{-1} - \xi_{3i}^{-1})I_{m_i})]Q_3X}{TS} \\
&= o_P(1) + o_P(1) + o_P(1) = o_P(1).
\end{aligned}$$

In the above, under Assumptions 2 (d), 3 (a), and Lemma 1, the first two terms' being $o_P(1)$ is trivial, and the third term is $o_P(1)$ by the fact that the elements of WX and Q_3WX are uniformly bounded in absolute value and by Lemma 5. $\mathcal{I}_{14}$ and $\mathcal{I}_{15}$ are $o_P(1)$ by a similar argument, and thus

$\widehat{\Sigma}_1 - \Sigma_1 = o_P(1)$.

Second,

$$\begin{aligned}
\widehat{\Sigma}_2 - \Sigma_2 &= \frac{X^{*\prime}(\widehat{\rho},\widehat{\theta})\widehat{\Omega}_\epsilon^{-1}u^*(\widehat{\rho},\widehat{\theta})}{(TS)^{1/2}} - \frac{X^{*\prime}(\rho,\theta)\Omega_\epsilon^{-1}u^*(\rho,\theta)}{(TS)^{1/2}} \\
&= \frac{[(\widehat{\vartheta}-\vartheta)X + X^*(\rho,\theta)]'(\Omega_\epsilon^{-1}+\widehat{\Delta})[(\widehat{\vartheta}-\vartheta)u + u^*(\rho,\theta)]}{(TS)^{1/2}} - \frac{X^{*\prime}(\rho,\theta)\Omega_\epsilon^{-1}u^*(\rho,\theta)}{(TS)^{1/2}} \\
&= \left[\frac{(\widehat{\vartheta}-\vartheta)^2X'\Omega_\epsilon^{-1}u}{(TS)^{1/2}} + \frac{(\widehat{\vartheta}-\vartheta)X'\Omega_\epsilon^{-1}u^*(\rho,\theta)}{(TS)^{1/2}} + \frac{(\widehat{\vartheta}-\vartheta)X^{*\prime}(\rho,\theta)\Omega_\epsilon^{-1}u}{(TS)^{1/2}}\right] \\
&\quad + \left[\frac{(\widehat{\vartheta}-\vartheta)^2X'\widehat{\Delta}u}{(TS)^{1/2}} + \frac{(\widehat{\vartheta}-\vartheta)X'\widehat{\Delta}u^*(\rho,\theta)}{(TS)^{1/2}} + \frac{(\widehat{\vartheta}-\vartheta)X^{*\prime}(\rho,\theta)\widehat{\Delta}u}{(TS)^{1/2}} + \frac{X^{*\prime}(\rho,\theta)\widehat{\Delta}u^*(\rho,\theta)}{(TS)^{1/2}}\right] \\
&\equiv \left[(\widehat{\vartheta}-\vartheta)^2\mathcal{I}_{21} + (\widehat{\vartheta}-\vartheta)\mathcal{I}_{22} + (\widehat{\vartheta}-\vartheta)\mathcal{I}_{23}] + [(\widehat{\vartheta}-\vartheta)^2\mathcal{I}_{24} + (\widehat{\vartheta}-\vartheta)\mathcal{I}_{25} + (\widehat{\vartheta}-\vartheta)\mathcal{I}_{26} + \mathcal{I}_{27}\right].
\end{aligned}$$

For the first term, we have $\mathrm{E}[\mathcal{I}_{21}] = 0$ and

$$\mathrm{Var}[\mathcal{I}_{21}] = \frac{1}{T}\frac{X'\Omega_\epsilon^{-1}\Omega_u\Omega_\epsilon^{-1}X}{S} = \frac{1}{T}O(1),\ N,S \to \infty,$$

under Assumptions 1 (a)-(d), 2 (d), 3 (a), and by Lemma 1. This implies that $\mathcal{I}_{21} = O_P(1)$. Similar arguments lead to $\mathcal{I}_{22} = O_P(1)$ and $\mathcal{I}_{23} = O_P(1)$.

For $\mathcal{I}_{24}$, we can write

$$\begin{aligned}
\mathcal{I}_{24} &= \frac{X'\widehat{\Delta}u}{(TS)^{\frac{1}{2}}} = (\widehat{\xi}_1^{-1} - \xi_1^{-1})\frac{X'Q_1u}{(TS)^{\frac{1}{2}}} + (\widehat{\xi}_2^{-1} - \xi_2^{-1})\frac{X'Q_2u}{(TS)^{\frac{1}{2}}} + \frac{X'[I_T \otimes d((\widehat{\xi}_{3i}^{-1} - \xi_{3i}^{-1})I_{m_i})]Q_3u}{(TS)^{\frac{1}{2}}} \overset{(1)}{=} (\widehat{\xi}_1^{-1} - \xi_1^{-1})O_P(1) + (\widehat{\xi}_2^{-1} - \xi_2^{-1})O_P(1) + O_P((TS)^{-\frac{1}{2}})\frac{X'Q_3u}{(TS)^{\frac{1}{2}}} \\
&= o_P(1) + O_P((TS)^{-\frac{1}{2}})O_P(1) = o_P(1),
\end{aligned}$$

where (1) is by an argument similar to that for $\mathcal{I}_{21}$, and by the standard parametric rate $\widehat{\xi}_{3i}^{-1} - \xi_{3i}^{-1} = O_P((TS)^{-\frac{1}{2}})$. Similar arguments lead to the results that $\mathcal{I}_{25}$, $\mathcal{I}_{26}$, and $\mathcal{I}_{27}$ are all $o_P(1)$. Therefore, $\widehat{\Sigma}_2 - \Sigma_2 = o_P(1)$ and the conclusion $(TS)^{\frac{1}{2}}(\widehat{\beta}_{\mathrm{FGLS}} - \widehat{\beta}_{\mathrm{GLS}}) = o_P(1)$ follows.

(c) We show that $\widehat{\Lambda}^* - \Lambda_\infty^* = o_P(1)$, where $\widehat{\Lambda}^* = \widehat{\Sigma}_1^{-1}\widehat{\Lambda}(\widehat{\Sigma}_1^{-1})'$, $\widehat{\Lambda} = \widehat{\Lambda}_\alpha + \widehat{\Lambda}_\mu + \widehat{\Lambda}_\nu$, and

$$\widehat{\Lambda}_\alpha = \frac{\widehat{\sigma}_\alpha^2\widehat{X}_\alpha^{*\prime}\widehat{X}_\alpha^*}{TS}, \widehat{X}_\alpha^{*\prime} = X^{*\prime}(\widehat{\rho},\widehat{\theta})(l_T \otimes (d(\widehat{\xi}_{3i}^{-1}l_{m_i}))),$$

$$\widehat{\Lambda}_\mu = \frac{\widehat{\sigma}_\mu^2\widehat{X}_\mu^{*\prime}\widehat{X}_\mu^*}{TS}, \widehat{X}_\mu^{*\prime} = X^{*\prime}(\widehat{\rho},\widehat{\theta})[\widehat{\xi}_2^{-1}(l_T \otimes d(E_{m_i})) + l_T \otimes d(\widehat{\xi}_{3i}^{-1}\overline{J}_{m_i})],$$

$$\widehat{\Lambda}_\nu = \frac{\widehat{\sigma}_\nu^2\widehat{X}_\nu^{*\prime}\widehat{X}_\nu^*}{TS}, \widehat{X}_\nu^{*\prime} = X^{*\prime}(\widehat{\rho},\widehat{\theta})\widehat{\Omega}_\epsilon^{-1}.$$

By a decomposition of $\widehat{\Sigma}_1^{-1}\widehat{\Lambda}(\widehat{\Sigma}_1^{-1})' - \Sigma_{1,\infty}^{-1}\Lambda_\infty(\Sigma_{1,\infty}^{-1})'$ similar to that in (C.2), it can be seen that $\Sigma_{1,\infty}$ is nonsingular, $\Sigma_{1,\infty}^{-1} = O(1)$, $\Lambda_\infty = O(1)$, $\widehat{\Sigma}_1 - \Sigma_{1,\infty} = o_P(1)$, and $\widehat{\Lambda} - \Lambda_\infty = o_p(1)$.

$\Sigma_{1,\infty}$'s being nonsingular and $\Sigma_{1,\infty}^{-1} = O(1)$ are by Assumption 3 (b), and $\Lambda_\infty = O(1)$ is by Assumption 3 (c). Then it follows that $\widehat{\Sigma}_1 - \Sigma_{1,\infty} = \widehat{\Sigma}_1 - \Sigma_1 + \Sigma_1 - \Sigma_{1,\infty} = o_P(1) + o(1) = o_P(1)$.

For $\widehat{\Lambda} - \Lambda_\infty$, we can write

$$\begin{aligned}
\widehat{\Lambda} - \Lambda_\infty &= (\widehat{\Lambda}_\alpha - \Lambda_\alpha + \Lambda_\alpha - \Lambda_{\alpha,\infty}) + (\widehat{\Lambda}_\mu - \Lambda_\mu + \Lambda_\mu - \Lambda_{\mu,\infty}) + (\widehat{\Lambda}_\nu - \Lambda_\nu + \Lambda_\nu - \Lambda_{\nu,\infty}) \\
&= (\widehat{\Lambda}_\alpha - \Lambda_\alpha) + (\widehat{\Lambda}_\mu - \Lambda_\mu) + (\widehat{\Lambda}_\nu - \Lambda_\nu) + o(1).
\end{aligned}$$

We have

$$\begin{aligned}
\widehat{\Lambda}_\alpha - \Lambda_\alpha &= \frac{\widehat{\sigma}_\alpha^2X^{*'}(\widehat{\rho},\widehat{\theta})(J_T \otimes d(\widehat{\xi}_{3i}^{-2}J_{m_i}))X^*(\widehat{\rho},\widehat{\theta})}{TS} - \frac{\sigma_\alpha^2X^{*'}(\rho,\theta)(J_T \otimes d(\xi_{3i}^{-2}J_{m_i}))X^*(\rho,\theta)}{TS} \\
&= \frac{(\sigma_\alpha^2 + op(1))[(\widehat{\vartheta}-\vartheta)X + X^*(\rho,\theta)]'(J_T \otimes d((\widehat{\xi}_{3i}^{-2} - \xi_{3i}^{-2} + \xi_{3i}^{-2})J_{m_i}))[(\widehat{\vartheta}-\vartheta)X + X^*(\rho,\theta)]}{TS} - \frac{\sigma_\alpha^2X^{*'}(\rho,\theta)(J_T \otimes d(\xi_{3i}^{-2}J_{m_i}))X^*(\rho,\theta)}{TS} \\
&= \frac{\sigma_\alpha^2(\widehat{\vartheta}-\vartheta)^2X'(J_T \otimes d((\widehat{\xi}_{3i}^{-2} - \xi_{3i}^{-2})J_{m_i}))X}{TS} + \frac{\sigma_\alpha^2(\widehat{\vartheta}-\vartheta)X'(J_T \otimes d((\widehat{\xi}_{3i}^{-2} - \xi_{3i}^{-2})J_{m_i}))X^*(\rho,\theta)}{TS} \\
&\quad + \frac{\sigma_\alpha^2(\widehat{\vartheta}-\vartheta)^2X'(J_T \otimes d(\xi_{3i}^{-2}J_{m_i}))X}{TS} + \frac{\sigma_\alpha^2(\widehat{\vartheta}-\vartheta)X'(J_T \otimes d(\xi_{3i}^{-2}J_{m_i}))X^*(\rho,\theta)}{TS} \\
&\quad + \frac{\sigma_\alpha^2(\widehat{\vartheta}-\vartheta)X^{*'}(\rho,\theta)(J_T \otimes d((\widehat{\xi}_{3i}^{-2} - \xi_{3i}^{-2})J_{m_i}))X}{TS} + \frac{\sigma_\alpha^2X^{*'}(\rho,\theta)(J_T \otimes d((\widehat{\xi}_{3i}^{-2} - \xi_{3i}^{-2})J_{m_i}))X^*(\rho,\theta)}{TS} \\
&\quad + \frac{\sigma_\alpha^2(\widehat{\vartheta}-\vartheta)X^{*'}(\rho,\theta)(J_T \otimes d(\xi_{3i}^{-2}J_{m_i}))X}{TS} + \frac{op(1)X^{*'}(\rho,\theta)(J_T \otimes d(\xi_{3i}^{-2}J_{m_i}))X^*(\rho,\theta)}{TS} + s.o.
\end{aligned}$$

Let the elements of matrices A and C be uniformly bounded in absolute value. Then, up to $o_P(1)$ terms, the above expression involves terms either of the form $\frac{A[(J_T \otimes d((\widehat{\xi}_{3i}^{-2} - \xi_{3i}^{-2})J_{m_i}))]C}{TS}$ or of the form $\frac{A[J_T \otimes (d(\xi_{3i}^{-2}J_{m_i}))]C}{TS}$.

For terms of the first form, we can write

$$\frac{A[(J_T \otimes d((\hat{\xi}_{3i}^{-2} - \xi_{3i}^{-2})J_{m_i}))]C}{TS} = \frac{A[(I_T \otimes d((\hat{\xi}_{3i}^{-2} - \xi_{3i}^{-2})I_{m_i}))][J_T \otimes d(J_{m_i})]C}{TS}.$$

Since the elements of A and $[J_T \otimes d(J_{m_i})]C$ are uniformly bounded in absolute value by hypothesis, Assumption 1 (d), Lemma 1 (a) and (c), and Lemma 5 applies. From these circumstances, we can conclude that $\frac{A[(J_T \otimes d((\hat{\xi}_{3i}^{-2} - \xi_{3i}^{-2})J_{m_i}))]C}{TS} = o_P(1)$. ■

For terms of the second form, it is clear that the row and column sums of $J_T \otimes (d(\xi_{3i}^{-2} J_{m_i}))$ are bounded in absolute value, uniformly in N, S. Then by Lemma 1 (c), we have $\frac{A[J_T \otimes (d(\xi_{3i}^{-2})J_{m_i})]C}{TS} = O(1)$. Therefore, $\hat{\Lambda}_\alpha - \Lambda_\alpha = o_P(1)$. The facts that $\hat{\Lambda}_\mu - \Lambda_\mu = o_P(1)$ and $\hat{\Lambda}_\nu - \Lambda_\nu = o_P(1)$ followed by exactly the same arguments. As a result, $\hat{\Lambda} - \Lambda_\infty = o_P(1)$.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.econmod.2018.06.022.

References

Amemiya, T., 1971. The estimation of the variances in a variance-components model. Int. Econ. Rev. 12, 1–13, https://doi.org/10.2307/2525492.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht, https://doi.org/10.1007/978-94-015-7799-1.

Baltagi, B.H., Fingleton, B., Pirotte, A., 2014. Spatial lag models with nested random effects: an instrumental variable procedure with an application to English house prices. J. Urban Econ. 80, 76–86, https://doi.org/10.1016/j.jue.2013.10.006.

Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London. ISBN: 0850860369.

Cliff, A.D., Ord, J.K., 1981. Spatial Processes: Models and Applications. Pion, London. ISBN: 0850860814.

Corrado, L., Fingleton, B., 2011. Multilevel Modeling with Spatial Effects. Working Paper. University of Cambridge. URI: http://hdl.handle.net/10943/255.

Corrado, L., Fingleton, B., 2012. Where is the economics in spatial econometrics? J. Reg. Sci. 52, 210–239, https://doi.org/10.1111/j.1467-9787.2011.00726.x.

Fingleton, B., Le Gallo, J., 2007. Finite sample properties of estimators of spatial models with autoregressive, or moving average, disturbances and system feedback. Annales d'conomie et de Statistique 87/88, 39–62, https://doi.org/10.2307/27650041.

Fingleton, B., 2008. A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. Spatial Econ. Anal. 3 (1), 27–44, https://doi.org/10.1080/17421770701774922.

Fingleton, B., Gallo, J., Pirotte, A., 2018. Panel data models with spatially dependent nested random effects. J. Reg. Sci. 58, 63–80, https://doi.org/10.1111/jors.12327.

Goldstein, H., 2011. Multilevel Statistical Models. John Wiley & Sons. ISBN: 9780470748657.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054, https://doi.org/10.2307/1912775.

Kapoor, M., Kelejian, H.H., Prucha, I.R., 2007. Panel data models with spatially correlated error components. J. Econom. 140 (1), 97–130, https://doi.org/10.1016/j.jeconom.2006.09.004.

Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the moran I test statistic with applications. J. Econom. 104, 219–257, https://doi.org/10.1016/S0304-4076(01)00064-1.

Kreft, I.G.G., De Leeuw, J., 1998. Introducing Multilevel Modeling. SAGE, London. ISBN: ISBN: 9780761951414.

Magnus, J.R., 1982. Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. J. Econom. 19, 239–285, https://doi.org/10.1016/0304-4076(82)90005-7.

Pötscher, B.M., Prucha, I.R., 1997. Dynamic Nonlinear Econometric Models, Asymptotic Theory. Springer, New York, https://doi.org/10.1007/978-3-662-03486-6.

Pötscher, B.M., Prucha, I.R., 2001. Basic elements of asymptotic theory. In: Baltagi, B.H. (Ed.), A Companion to Theoretical Econometrics. Blackwell, Oxford, pp. 201–229, https://doi.org/10.1002/9780470996249.ch11.

Rao, C.R., 1973. Linear Statistical Inference and its Applications. Wiley, New York. ISBN: 0471708232.

Ye, Q., Long, Z., 2016. Estimation on hierarchical data model with spatial error autoregressive. J. Quant. Tech. Econ. 33 (5), 143–161, https://doi.org/10.13653/j.cnki.jqte.2016.05.009.