Journal of Econometrics 184 (2015) 209–232

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Estimating a spatial autoregressive model with an endogenous spatial weight matrix[☆]

Xi Qu[a,*], Lung-fei Lee[b]

[a] Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, 200052, China

[b] Department of Economics, The Ohio State University, United States

ARTICLE INFO

Article history:
Received 22 October 2012
Received in revised form
5 August 2014
Accepted 26 August 2014
Available online 19 September 2014

JEL classification:
C31
C51

Keywords:
Spatial autoregressive model
Endogenous spatial weight matrix
2SIV
QMLE
GMM

ABSTRACT

The spatial autoregressive (SAR) model is a standard tool for analyzing data with spatial correlation. Conventional estimation methods rely on the key assumption that the spatial weight matrix is strictly exogenous, which would likely be violated in some empirical applications where spatial weights are determined by economic factors. This paper presents model specification and estimation of the SAR model with an endogenous spatial weight matrix. We provide three estimation methods: two-stage instrumental variable (2SIV) method, quasi-maximum likelihood estimation (QMLE) approach, and generalized method of moments (GMM). We establish the consistency and asymptotic normality of these estimators and investigate their finite sample properties by a Monte Carlo study.

1. Introduction

The spatial autoregressive (SAR) model is of great interest to economists because it has a game structure and can be interpreted as a reaction function. It is widely used in spatial econometrics and for modeling social networks. In spatial econometrics, the SAR model has been applied to cases where outcomes of a spatial unit at one location depend on those of its neighbors. The corresponding spatial weight matrix is a measure of connections among different locations. Consequently, the spatial dependence parameter provides a multiplier for the spillover effect. SAR models can also be used to model social networks. For example, a student's behavior (such as smoking or academic achievement) can be directly affected by his/her friends' behaviors. The weight matrix can then be constructed by using friendship relations, and the network (spatial) dependence parameter can be interpreted as the strength of peer effects. As measuring spillover and peer effects has strong policy implications, such as setting school policies, correct estimation of the spatial dependence parameter is important to both theory and practice.

Estimation methods for the SAR model with an exogenous spatial weight matrix has been well established in the literature: the maximum likelihood estimation (MLE) of Ord (1975) and Lee (2004); the instrumental variable (IV) methods of Anselin (1980) and Kelejian and Prucha (1998, 1999), and the generalized method of moments (GMM) of Lee (2007), Lee and Liu (2010), Lin and Lee (2010), and Liu et al. (2010). Consistency and asymptotic normality of these estimators are established under the assumption that the spatial weight matrix is strictly exogenous. This exogenous assumption may hold when spatial weights are constructed using predetermined geographic distances; for example, between different cities or countries. However, if "economic distance" such as the relative GDP or trade volume is used to construct the weight matrix, then it is very likely that these elements are correlated with the final outcome. Similarly, in the social network framework, some unobserved characteristics may affect both the friendship relationship and behavioral outcomes (Hsieh and Lee, 2011). Therefore, in many applications, the exogenous spatial weight assumption might be violated.

☆ We would like to thank the editor, Peter Robinson, the associate editor, and two anonymous referees for insightful and instructive comments. An earlier version of the paper was presented in seminars at the Ohio State University, City University of HK, Nanyang Technological University, Tsinghua University, UEST of China, and Shanghai Jiao Tong University. We appreciate comments from participants of those seminars, especially Robert de Jong and Xingbai Xu at the OSU. The usual disclaimer applies.

* Corresponding author.
E-mail address: xiqu@sjtu.edu.cn (X. Qu).

http://dx.doi.org/10.1016/j.jeconom.2014.08.008
0304-4076/

However, due to the technical complication in estimating spatial models with an endogenous spatial weight matrix, to the best of our knowledge, so far no estimation method has been proposed for this case. In Pinkse and Slade (2010), they pointed out future directions of spatial econometrics. Endogeneity of spatial weights was among several problems they emphasized. They concluded that "many of these are still waiting for good solutions" and the endogeneity problem "can admittedly be challenging".

Recently, Kelejian and Piras (2014) try to find IVs for endogenous weights by regressing these spatial weights on exogenous variables. Their asymptotic results are based on high level assumptions. We attempt to tackle the issue of endogenous spatial weight matrix by exploring the model structure of spatial weights in this paper. Our asymptotic results are derived through basic regularity conditions. By modeling explicitly the source of endogeneity, we obtain two sets of equations—one is for the SAR outcome, and the other is for entries of the spatial weight matrix. The disturbances in the SAR outcome equation and the error terms in the entry equation are allowed to be correlated. When their correlation coefficient is nonzero, the spatial weight matrix becomes endogenous. We focus on estimation issues for this type of SAR model. By imposing assumptions of conditional mean independence and homoskedasticity, we can overcome the endogeneity problem using the control function method. By exploring the unobservable control variables for endogeneity in the outcome SAR equation, we propose three estimation methods. The first estimation method is a two-stage instrumental variables (2SIV) approach. In the first stage estimation, we consistently estimate the parameters of the entry equation. In the second stage, we replace the unobserved control variables in the outcome equation by the residuals of the entry equation, and then use the standard IV methods to estimate the SAR outcome equation. The second method we propose is the quasi-maximum likelihood estimation (QMLE), in which all the parameters can be jointly estimated via a normal likelihood function of the equation system even the disturbances in the model are not normally distributed. The third method is a GMM approach, in which an outcome equation with control variables for endogeneity provides additional quadratic moments for estimation.

The main aim of this paper is to show the consistency and asymptotic normality of aforementioned three estimators. The estimators involve statistics with linear–quadratic forms of disturbances, in which the quadratic matrix depends on the spatial weight matrix. As entries in the spatial weight matrix are nonlinear functions of disturbances, those statistics are not really of quadratic forms with nonstochastic quadratic matrices. Therefore, the standard asymptotic results for linear–quadratic forms do not directly apply to the situation here. Instead, we adopt the asymptotic inference under near-epoch dependence (NED) from Jenish and Prucha (2012).[1] Our key work is to show the NED properties of random variables and functions involved in our estimators. To do that, we assume either the spatial weight matrix is sparse or the upper bound of its elements decreases as a power function of the physical distance. Therefore, in our setting, the physical distance plays an important role to constrain the magnitude of the spatial weights.

The rest of this paper is organized as follows. In Section 2, we present the model specification of the outcome equation and the entries of its spatial weight matrix. In Section 3, we propose the 2SIV, QMLE and GMM estimation methods for this model. Consistency and asymptotic normality of estimates from these methods are derived in Section 4. Some extensions with a generalized control function are discussed in Section 5. In Section 6, Monte Carlo simulations are provided to investigate finite sample properties of our proposed estimators and compare their performances with those under the exogenous spatial weight matrix assumption. Related expressions of the log quasi-likelihood function are collected in Appendix A. Proofs of all the lemmas, propositions, and theorems are given in Appendix B.

2. The model

2.1. Model specification

Following Jenish and Prucha (2009, 2012), we consider spatial processes located on a (possibly) unevenly spaced lattice $D \subseteq R^{d_0},\ d_0 \geq 1$. Asymptotic methods we employ are increasing domain asymptotics: growth of the sample is ensured by an unbounded expansion of the sample region as in Jenish and Prucha (2012).[2]

Assumption 1. The lattice $D \subset R^{d_0},\ d_0 \geq 1$, is infinitely countable. The location $l : \{1, \dots, n\} \to D_n \subset D$ is a mapping of individual i to its location $l(i) \in D_n \subset R^{d_0}$. All elements in D are located at distances of at least $\rho_0 > 0$ from each other, i.e., $\forall l(i), l(j) \in D : \rho_{ij} \geq \rho_0$, where ρ_{ij} is the distance between individuals i and j (locations $l(i)$ and $l(j)$); w.l.o.g. we assume that $\rho_0 = 1$.

As our asymptotic analysis is based on inference under the spatial near-epoch dependence for increasing domain but not for infill asymptotics, physical distance plays an important role in keeping agents apart from each other. For the case of pure economic distance, if there were economic factors which keep agents apart, we might replace the "physical distance" in Assumption 1 by economic distance. In this regard, with Assumption 1, our model will be more relevant for regional economic studies rather than social network ones. In regional issues, physical distance would definitely play a role.

Let $\{(\varepsilon_{l(i),n}, v_{l(i),n});\ l(i) \in D_n,\ n \in N\}$ be a triangular double array of real random variables defined on a probability space $(\Omega; F; P)$, where the index set $D_n \subset D$ is a finite set and D satisfies Assumption 1. For simplicity, we use the notation $\varepsilon_{i,n}$ and $v_{i,n}$ to indicate $\varepsilon_{l(i),n}$ and $v_{l(i),n}$. Same for the other variables. Let

$$Z_n = X_{2n}\Gamma + \varepsilon_n, \tag{2.1}$$

where X_{2n} is an $n \times k_2$ matrix with its elements $\{x_{2,in};\ l(i) \in D_n,\ n \in N\}$ being deterministic and bounded in absolute value for all i and n, Γ is a $k_2 \times p_2$ matrix of coefficients, $\varepsilon_n = (\varepsilon_{1,n}, \dots, \varepsilon_{n,n})'$ is an $n \times p_2$ matrix of disturbances with $\varepsilon_{i,n} = (\varepsilon_{1,in}, \dots, \varepsilon_{p_2,in})'$ being p_2 dimensional column vectors, and $Z_n = (z_{1,n}, \dots, z_{n,n})'$ is an $n \times p_2$ matrix with $z_{i,n} = (z_{1,in}, \dots, z_{p_2,in})'$. $W_n = (w_{ij,n})$[3] is an $n \times n$ nonnegative matrix with zero diagonals and its elements constructed by $Z_n : w_{ij,n} = h_{ij}(Z_n, \rho_{ij})$ for $i, j = 1, \dots, n;\ i \neq j$, where $h(\cdot)$ is a bounded function.[4] $Y_n = (y_{1,n}, \dots, y_{n,n})'$ is an $n \times 1$ vector from a cross-sectional SAR model specified as

$$Y_n = \lambda W_n Y_n + X_{1n}\beta + V_n, \tag{2.2}$$

[1] In our earlier version, we explore finite neighbor's dependence which would be similar to m-dependence in time series analysis. But the NED is more general as we have found in this version.

[2] Infill asymptotics have not been developed for a NED process in the literature.

[3] Here we simplify the notation by regarding the subscripts i and j as integer values to indicate entries in a vector or matrix even though i and j refer formally in Assumption 1 to locations in the lattice D contained in the d_0-dimensional Euclidean space R^d.

[4] In the example that W_n is constructed by $w_{ij,n} = 1/|z_{i,n} - z_{j,n}|$, for the boundedness, we actually need to have a trimming on it such that $w_{ij,n} = c_{e0}$ if $|z_{i,n} - z_{j,n}| < d_{e0}$, where c_{e0} and d_{e0} are constants. This seems sensible, otherwise, units with similar values of z would have extremely strong influence on each other.

where X_{1n} is an $n \times k_1$ matrix with its elements $\{x_{1,in};\ l(i) \in D_n,\ n \in N\}$ being deterministic and bounded in absolute value for all i and n, $V_n = (v_{1,n}, \ldots, v_{n,n})'$, λ is a scalar, and $\beta = (\beta_1, \ldots, \beta_{k_1})'$ is a $k_1 \times 1$ vector of coefficients.

2.2. Model interpretation

We consider n agents in an area where each agent i is endowed with a predetermined location $l(i)$. Any two agents are separated away by a distance of at least 1. Due to some competition or spillover effects, each agent i has an outcome $y_{i,n}$ directly affected by its neighbors' outcomes $y'_{j,n}$s. The outcome equation is $y_{i,n} = \lambda \sum_{j \neq i} w_{ij,n} y_{j,n} + x'_{1,in}\beta + v_{i,n}$, where the spatial weight $w_{ij,n}$ is a measure of the relative strength of linkage between agents i and j, and the spatial coefficient λ provides a multiplier for the spillover effects. However, the spatial weight $w_{ij,n}$ is not predetermined but depends on some observable random variable Z_n. We can think of $z_{i,n}$ as some economic variables at location $l(i)$ such as GDP, consumption and economic growth rate which influence strength of links across units.

This specification has been used in the literature, and it may introduce endogeneity into the spatial weight matrix. For example, Anselin and Bera (1997) provided several examples in economic applications on the use of weights based on "economic" distance. In Case et al. (1993), weights (before row normalization) of the form $w_{ij,n} = 1/|z_{i,n} - z_{j,n}|$ were specifically suggested, where $z_{i,n}$ and $z_{j,n}$ are observations on "meaningful" socioeconomic characteristics. In Conway and Rork (2004), they used migration flow data to construct a spatial weight matrix. Another example is in Crabbé and Vandenbussche (2008), where in addition to the physical distance, spatial weight matrices were constructed by inverse trade share and inverse distance between GDP per capita.

2.3. Source of endogeneity

We have the following moment assumption.

Assumption 2. The error terms $v_{i,n}$ and $\varepsilon_{i,n}$, have a joint distribution: $(v_{i,n}, \varepsilon'_{i,n})' \sim$ i.i.d.$(0, \Sigma_{v\varepsilon})$, where $\Sigma_{v\varepsilon} = \begin{pmatrix} \sigma_v^2 & \sigma'_{v\varepsilon} \\ \sigma_{v\varepsilon} & \Sigma_\varepsilon \end{pmatrix}$ is positive definite, σ_v^2 is a scalar variance, covariance $\sigma_{v\varepsilon} = (\sigma_{v\varepsilon_1}, \ldots, \sigma_{v\varepsilon_{p_2}})'$ is a p_2 dimensional vector, and Σ_ε is a $p_2 \times p_2$ matrix. The $\sup_{i,n} E|v_{i,n}|^{4+\delta_\varepsilon}$ and $\sup_{i,n} E\|\varepsilon_{i,n}\|^{4+\delta_\varepsilon}$ exist for some $\delta_\varepsilon > 0$. Furthermore, $E(v_{i,n}|\varepsilon_{i,n}) = \varepsilon'_{i,n}\delta$ and $\mathrm{Var}(v_{i,n}|\varepsilon_{i,n}) = \sigma_\xi^2$.[5]

The endogeneity of W_n comes from the correlation between $v_{i,n}$ and $\varepsilon_{i,n}$. If $\sigma_{v\varepsilon}$ is zero, the spatial weight matrix W_n might be treated as strictly exogenous and we can apply conventional methodology of SAR models for estimation. However, if $\sigma_{v\varepsilon}$ is not zero, W_n becomes an endogenous spatial weights matrix.

From the two conditional moments assumptions in Assumption 2, we have the p_2 dimensional column vector $\delta = \Sigma_\varepsilon^{-1}\sigma_{v\varepsilon}$ and the scalar $\sigma_\xi^2 = \sigma_v^2 - \sigma'_{v\varepsilon}\Sigma_\varepsilon^{-1}\sigma_{v\varepsilon}$. Denote $\xi_n = V_n - \varepsilon_n\delta$, then its mean conditional on ε_n is zero and its conditional variance matrix is $\sigma_\xi^2 I_n$. In particular, ξ_n are uncorrelated with the terms of ε_n and the variance of ξ_n is $\sigma_\xi^2 I_n$. The outcome equation (2.2) becomes

$$Y_n = \lambda W_n Y_n + X_{1n}\beta + (Z_n - X_{2n}\Gamma)\delta + \xi_n, \tag{2.3}$$

with $E(\xi_{i,n}|\varepsilon_{i,n}) = 0$ and $E(\xi_{i,n}^2|\varepsilon_{i,n}) = \sigma_\xi^2$; and $\xi_{i,n}$'s are i.i.d. across i. Our subsequent asymptotic analysis will mainly rely on Eq. (2.3), where $(Z_n - X_{2n}\Gamma)$ are control variables to control the endogeneity of W_n. Assumption 2 is relatively general without imposing a specific distribution on disturbances as it is based on only conditional moments restrictions.

In the special case that $(v_{i,n}, \varepsilon'_{i,n})'$ has a jointly normal distribution, then $v_{i,n}|\varepsilon_{i,n} \sim N(\sigma'_{v\varepsilon}\Sigma_\varepsilon^{-1}\varepsilon_{i,n}, \sigma_v^2 - \sigma'_{v\varepsilon}\Sigma_\varepsilon^{-1}\sigma_{v\varepsilon})$ and ξ_n is independent of ε_n in Eq. (2.1).

3. Estimation methods

3.1. The two-stage IV estimation

In the first stage, we estimate $Z_n = X_{2n}\Gamma + \varepsilon_n$ by the ordinary least squares (OLS) method, so $\widehat{\Gamma} = (X'_{2n}X_{2n})^{-1}X'_{2n}Z_n$. Then, in the second stage by substituting $\widehat{\Gamma}$ for Γ in (2.3), we have

$$Y_n = \lambda W_n Y_n + X_{1n}\beta + (Z_n - X_{2n}\widehat{\Gamma})\delta + \widehat{\xi}_n, \tag{3.1}$$

where $\widehat{\xi}_n = \xi_n + X_{2n}(\widehat{\Gamma} - \Gamma)\delta = \xi_n + P_n\varepsilon_n\delta$ with $P_n = X_{2n}(X'_{2n}X_{2n})^{-1}X'_{2n}$. Since $Z_n - X_{2n}\widehat{\Gamma} = P_n^{\perp}Z_n = P_n^{\perp}\varepsilon_n$ with $P_n^{\perp} = I_n - P_n$, (3.1) can be explicitly rewritten as

$$Y_n = (W_nY_n, X_{1n}, P_n^{\perp}Z_n)\kappa + (\xi_n + P_n\varepsilon_n\delta), \tag{3.2}$$

where $\kappa = \begin{pmatrix}\lambda & \beta' & \delta'\end{pmatrix}'$. For estimation, with the control variables $(Z_n - X_{2n}\Gamma)$ added in (2.3) or $P_n^{\perp}Z_n$ in (3.2), W_n can be treated as predetermined or exogenous. However, W_nY_n remains endogenous in (2.3) and (3.2). So for an IV estimation, we need instruments for W_nY_n. Let Q_n be an $n \times m$ matrix of IVs, then a 2SIV estimator of κ with Q_n will be

$$\begin{aligned}\widehat{\kappa} &= [(W_nY_n, X_{1n}, P_n^{\perp}Z_n)'Q_n(Q_n'Q_n)^{-1} \\ &\quad \times Q_n'(W_nY_n, X_{1n}, P_n^{\perp}Z_n)]^{-1}(W_nY_n, X_{1n}, P_n^{\perp}Z_n)' \\ &\quad \times Q_n(Q_n'Q_n)^{-1}Q_n'Y_n.\end{aligned}$$

As the composite error $(\xi_n + P_n\varepsilon_n\delta)$ is not homogeneous as its variance matrix is $\Pi_n = \sigma_{\xi0}^2 I_n + \delta_0'\Sigma_{\varepsilon0}\delta_0 P_n$, we may also consider a generalized 2SIV (G2SIV), which is

$$\begin{aligned}\widehat{\kappa}_G &= [(W_nY_n, X_{1n}, P_n^{\perp}Z_n)'\Pi_n^{-1}Q_n(Q_n'\Pi_n^{-1}Q_n)^{-1} \\ &\quad \times Q_n'\Pi_n^{-1}(W_nY_n, X_{1n}, P_n^{\perp}Z_n)]^{-1} \\ &\quad \cdot (W_nY_n, X_{1n}, P_n^{\perp}Z_n)'\Pi_n^{-1}Q_n(Q_n'\Pi_n^{-1}Q_n)^{-1}Q_n'\Pi_n^{-1}Y_n.\end{aligned}$$

In practice, as Π_n involves unknown parameters, they need to be consistently estimated by some initial estimates so as to have a consistent $\widehat{\Pi}_n$, and a feasible G2SIV. The details of such a construction are in Section 4.4.

3.2. The quasi-maximum likelihood estimation

As in White (1982), based on the i.i.d. disturbances $(v_{i,n}, \varepsilon'_{i,n})' \sim (0, \Sigma_{v\varepsilon})$ with $\Sigma_{v\varepsilon} = \begin{pmatrix} \sigma_v^2 & \sigma'_{v\varepsilon} \\ \sigma_{v\varepsilon} & \Sigma_\varepsilon \end{pmatrix}$, we can directly write down the log quasi-likelihood function under a normal distributional specification as:

$$\begin{aligned}\ln L_n &= -n\ln(2\pi) - \frac{n}{2}\ln|\Sigma_{v\varepsilon}| + \ln|S_n(\lambda)| \\ &\quad - \frac{1}{2}[(S_n(\lambda)Y_n - X_{1n}\beta), (\mathrm{vec}(Z_n - X_{2n}\Gamma))'] \\ &\quad \times (\Sigma_{v\varepsilon}^{-1} \otimes I_n)\begin{pmatrix} S_n(\lambda)Y_n - X_{1n}\beta \\ \mathrm{vec}(Z_n - X_{2n}\Gamma)\end{pmatrix},\end{aligned} \tag{3.3}$$

where $S_n(\lambda) = I_n - \lambda W_n$. Alternatively, by the partitioned quadratic formulation that

$$\begin{aligned}&(v_{i,n}, \varepsilon'_{i,n})\Sigma_{v\varepsilon}^{-1}(v_{i,n}, \varepsilon'_{i,n})' \\ &= (v_{i,n} - \sigma'_{v\varepsilon}\Sigma_\varepsilon^{-1}\varepsilon_{i,n})'(\sigma_v^2 - \sigma'_{v\varepsilon}\Sigma_\varepsilon^{-1}\sigma_{v\varepsilon})^{-1} \\ &\quad \times (v_{i,n} - \sigma'_{v\varepsilon}\Sigma_\varepsilon^{-1}\varepsilon_{i,n}) + \varepsilon'_{i,n}\Sigma_\varepsilon^{-1}\varepsilon_{i,n},\end{aligned}$$

[5] This conditional homoskedasticity condition is required for the QMLE theory. For the IV or GMM estimations, we can relax this assumption.

the log quasi-likelihood function can also be written as

$$\ln L_n(\theta) = -n\ln(2\pi) - \frac{n}{2}\ln\sigma_\xi^2 + \ln|S_n(\lambda)| - \frac{n}{2}\ln|\Sigma_\varepsilon| - \frac{1}{2}\sum_{i=1}^{n}(z'_{i,n} - x'_{2..in}\Gamma)\Sigma_\varepsilon^{-1}(z_{i,n} - \Gamma' x_{2,in}) - \frac{1}{2\sigma_\xi^2}[S_n(\lambda)Y_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta]' \times [S_n(\lambda)Y_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta] \quad (3.4)$$

where $\theta = (\lambda, \beta', \mathrm{vec}(\Gamma)', \sigma_\xi^2, \alpha', \delta')'$ with α being a J-dimensional column vector of distinct elements in Σ_ε, $\delta = \Sigma_\varepsilon^{-1}\sigma_{v\varepsilon}$, and $\sigma_\xi^2 = \sigma_v^2 - \sigma'_{v\varepsilon}\Sigma_\varepsilon^{-1}\sigma_{v\varepsilon}$. The QMLE $\widehat{\theta} = \arg\max_{\theta\in\Theta}\ln L_n(\theta)$. A necessary condition is $\frac{\partial \ln L_n(\widehat{\theta})}{\partial\theta} = 0$, where the first order derivatives of the log quasi-likelihood function are listed in Appendix A.

3.3. The generalized method of moments estimation

Let X_n collect different column vectors in X_{1n} and X_{2n}. For the GMM estimation, as X_n is strictly exogenous and $E(\xi_{i,n}|\varepsilon_{i,n}) = 0$, a possible set of linear moments for estimation can be

$E(X'_n\varepsilon_n) = 0$, $\quad E((M_nX_n)'\xi_n) = 0$, $\quad$ and $E((M_nZ_n)'\xi_n) = 0$,

where M_n is an $n \times n$ matrix which can be constructed from I_n and W_n. For example, a finite number of matrices M_n can be either I_n, $W_n^{m_1}W_n'^{m_2}$, G_n, G'_n, or G'_nG_n, where $G_n = W_n(I_n - \lambda_0 W_n)^{-1}$, for some nonnegative integers m_1 and m_2. In addition to linear moments, we have quadratic moments $E[\xi'_n(M_n - \mathrm{tr}(M_n)I_n/n)\xi_n] = 0$ from the assumption $E(\xi_{i,n}^2|\varepsilon_{i,n}) = \sigma_\xi^2$.[6]

Let Q_n be an $n \times m^*$ matrix with elements of M_nZ_n and M_nX_n, then the corresponding empirical moments can be:

(1) $X'_n(Z_n - X_{2n}\Gamma)$;
(2) $Q'_n[Y_n - \lambda W_nY_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta]$; and
(3) $[Y_n - \lambda W_nY_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta]'[M_n - \mathrm{tr}(M_n)I_n/n]'[Y_n - \lambda W_nY_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta]$.

With several constructed M_{jn} matrices, $j = 1, \ldots, m$, in place of a single M_n matrix, denote the matrices $P_{jn} = M_{jn} - \mathrm{tr}(M_{jn})I_n/n$ for $j = 1, \ldots, m$, and $\theta^G = (\lambda, \beta', \mathrm{vec}(\Gamma)', \delta)'$, then the set of moment functions for the GMM estimation is

$$g_n(\theta^G) = (\xi'_n(\theta^G)P_{1n}\xi_n(\theta^G), \ldots, \xi'_n(\theta^G)P_{mn}\xi_n(\theta^G), \xi'_n(\theta^G)Q_n, \mathrm{vec}(\varepsilon'_n(\theta^G)X_n)')',$$

where $\theta^G = (\lambda, \beta', \mathrm{vec}(\Gamma)', \delta')'$, $\xi_n(\theta^G) = S_n(\lambda)Y_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta$ and $\varepsilon_n(\theta^G) = Z_n - X_{2n}\Gamma$. Our GMM estimator of θ^G is derived from $\hat{\theta}_n^G = \arg\min_{\theta\in\Theta} g'_n(\theta^G)a'_na_ng_n(\theta^G)$, where a'_na_n is a positive definite matrix that may depend on the data.

4. Asymptotic properties of estimators

4.1. Key statistics

The 2SIV

For the 2SIV estimator $\widehat{\kappa}$ and G2SIV estimator $\widehat{\kappa}_G$,

$$\widehat{\kappa} - \kappa_0 = [(W_nY_n, X_{1n}, P_n^\perp Z_n)'Q_n(Q'_nQ_n)^{-1} \times Q'_n(W_nY_n, X_{1n}, P_n^\perp Z_n)]^{-1} \cdot (W_nY_n, X_{1n}, P_n^\perp Z_n)'Q_n(Q'_nQ_n)^{-1}Q'_n(\xi_n + P_n\varepsilon_n\delta_0)$$

and

$$\widehat{\kappa}_G - \kappa_0 = [(W_nY_n, X_{1n}, P_n^\perp Z_n)'\Pi_n^{-1}Q_n(Q'_n\Pi_n^{-1}Q_n)^{-1} \times Q'_n\Pi_n^{-1}(W_nY_n, X_{1n}, P_n^\perp Z_n)]^{-1} \cdot (W_nY_n, X_{1n}, P_n^\perp Z_n)'\Pi_n^{-1}Q_n(Q'_n\Pi_n^{-1}Q_n)^{-1} \times Q'_n\Pi_n^{-1}(\xi_n + P_n\varepsilon_n\delta_0),$$

where the subscript 0 on parameters denotes their true values. As $\Pi_n^{-1} = (\sigma_{\xi0}^2 I_n + \delta'_0\Sigma_{\varepsilon0}\delta_0 P_n)^{-1} = \frac{1}{\sigma_{\xi0}^2}(I_n - \frac{\delta'_0\Sigma_{\varepsilon0}\delta_0}{\sigma_{\xi0}^2+\delta'_0\Sigma_{\varepsilon0}\delta_0}P_n)$, for the consistency and asymptotic distribution of $\widehat{\kappa}$ and $\widehat{\kappa}_G$, terms we need to analyze are Q'_nQ_n, Q'_nX_{2n}, $Q'_n(W_nY_n, X_{1n}, P_n^\perp\varepsilon_n)$, $X'_{2n}(W_nY_n, X_{1n}, P_n^\perp\varepsilon_n)$, $Q'_n\xi_n$, and $Q'_nP_n\varepsilon_n\delta_0$. Here $W_nY_n = W_n(I_n - \lambda_0W_n)^{-1}(X_{1n}\beta_0 + \varepsilon_n\delta_0 + \xi_n) = G_n(X_{1n}\beta_0 + \varepsilon_n\delta_0 + \xi_n)$, where $G_n(\lambda) = W_n(I_n - \lambda W_n)^{-1}$ with $G_n = G_n(\lambda_0)$. Let X_n be an $n \times k$ matrix collecting all distinct column vectors in X_{1n} and X_{2n}. Then, for the choice of the IV matrix Q_n, its column vectors can be linear combinations of $X_n, W_nX_n, W_n^2X_n, \ldots$, and columns in $P_n^\perp Z_n$. For example, if we choose $Q_n = (G_nX_n, G_nZ_n, X_n, Z_n)$, which is an optimal choice of the IV matrix as derived in the following section, then terms which need to be analyzed for consistency via some law of large numbers (LLN) are

$$\frac{1}{n}X'_nG_nX_n,\ \frac{1}{n}X'_nG'_n\varepsilon_n,\ \frac{1}{n}X'_nG_n\varepsilon_n,\ \frac{1}{n}X'_nG'_nG_nX_n,\ \frac{1}{n}X'_nG'_nG_n\varepsilon_n,\ \frac{1}{n}\varepsilon'_nG_n\varepsilon_n,\ \frac{1}{n}\varepsilon'_nG'_nG_n\varepsilon_n,$$
$$\frac{1}{n}X'_nG_n\xi_n,\ \frac{1}{n}X'_nG'_n\xi_n,\ \frac{1}{n}\varepsilon'_nG_n\xi_n,\ \frac{1}{n}X'_nG'_nG_n\xi_n,\ \text{and}\ \frac{1}{n}\varepsilon'_nG'_nG_n\xi_n.$$

For asymptotic distribution of the estimator, we need to consider the stochastic convergence in distribution via central limit theorem (CLT) for some of those terms after proper rescaling.

The QMLE

To show the consistency of the QMLE $\widehat{\theta}$, first we need to show the uniform convergence of the log quasi-likelihood function to its expectation, i.e., $\sup_{\theta\in\Theta}\frac{1}{n}|\ln L_n(\theta) - E(\ln L_n(\theta))| = o_p(1)$. It is sufficient to show the uniform convergence of the sample averages of $\ln|S_n(\lambda)|$ and $\xi_n(\theta)'\xi_n(\theta)$, where $\xi_n(\theta) = [S_n(\lambda)Y_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta]$. Note that

$$\xi_n(\theta) = S_n(\lambda)S_n^{-1}(X_{1n}\beta_0 + V_n) - X_{1n}\beta - [X_{2n}(\Gamma_0 - \Gamma) + \varepsilon_n]\delta = (\lambda_0 - \lambda)G_n(X_{1n}\beta_0 + \varepsilon_n\delta_0) + X_{1n}(\beta_0 - \beta) - X_{2n}(\Gamma_0 - \Gamma)\delta + \varepsilon_n(\delta_0 - \delta) + [I_n - (\lambda - \lambda_0)G_n]\xi_n,$$

where $S_n = I_n - \lambda_0W_n$. From the Taylor expansion,

$$\frac{1}{n}\ln|S_n(\lambda)| = \frac{1}{n}\ln|I_n - \lambda W_n| = -\frac{1}{n}\sum_{i=1}^{n}\left[\sum_{l=1}^{\infty}\frac{\lambda^l}{l}(W_n^l)_{ii}\right].$$

Hence, in the log quasi-likelihood function, the terms which need to be analyzed are

$$\frac{1}{n}X'_nG'_nG_nX_n,\ \frac{1}{n}X'_nG_nX_n,\ \frac{1}{n}X'_nG'_n\varepsilon_n,\ \frac{1}{n}X'_nG_n\varepsilon_n,\ \frac{1}{n}X'_nG'_nG_n\varepsilon_n\frac{1}{n}\xi'_nG'_nG_nX_n,\ \frac{1}{n}\xi'_nG_nX_n,\ \frac{1}{n}\xi'_nG'_n\varepsilon_n;$$
$$\frac{1}{n}\xi'_nG_n\varepsilon_n,\ \frac{1}{n}\xi'_nG'_nG_n\varepsilon_n,\ \frac{1}{n}\varepsilon'_nG'_nG_n\varepsilon_n,\ \frac{1}{n}\varepsilon'_nG_n\varepsilon_n,\ \frac{1}{n}\xi'_n\varepsilon_n,\ \frac{1}{n}\xi'_nG'_nG_n\xi_n,\ \text{and}\ \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{l=1}^{\infty}\frac{\lambda^l}{l}(W_n^l)_{ii}\right]$$

for consistency via LLN, and some properly rescaled terms for their asymptotic distributions via CLT.

[6] As in Lin and Lee (2010), with an unknown heteroskedasticity in ξ_n, i.e., $E(\xi_{i,n}^2|\varepsilon_{i,n}) = \sigma^2(\varepsilon_{i,n})$, the quadratic moment may be modified to $E[\xi'_n(M_n - \mathrm{Diag}(M_n))\xi_n] = 0$, where $\mathrm{Diag}(A)$ for a square matrix A denotes the diagonal matrix formed by the diagonal elements of A, for consistent estimation.

The GMM

The GMM is based on the first two moments of ξ_n and ε_n. Some elements in $g_n(\theta^G)$ have similar expressions as those in the 2SIV estimator and QMLE. Some have new features to analyze, such as

$$\frac{1}{n}X_n'\overline{M}_n'X_n,\ \frac{1}{n}\varepsilon_n'\overline{M}_n'\varepsilon_n,\ \frac{1}{n}\xi_n'\overline{M}_n'\xi_n,\ \frac{1}{n}X_n'G_n'\overline{M}_n'X_n,$$
$$\frac{1}{n}\varepsilon_n'G_n'\overline{M}_n'\varepsilon_n,\ \frac{1}{n}\xi_n'G_n'\overline{M}_n'\xi_n,$$
$$\frac{1}{n}X_n'G_n'\overline{M}_n'G_nX_n,\ \frac{1}{n}\varepsilon_n'G_n'\overline{M}_n'G_n\varepsilon_n,\ \frac{1}{n}\xi_n'G_n'\overline{M}_n'G_n\xi_n,$$
$$\frac{1}{n}X_n'\overline{M}_n'\varepsilon_n,\ \frac{1}{n}X_n'\overline{M}_n'\xi_n,\ \frac{1}{n}\varepsilon_n'\overline{M}_n'\xi_n,$$
$$\frac{1}{n}X_n'G_n\overline{M}_n'\varepsilon_n,\ \frac{1}{n}X_n'G_n\overline{M}_n'\xi_n,\ \frac{1}{n}\varepsilon_n'G_n\overline{M}_n'\xi_n,$$
$$\frac{1}{n}X_n'G_n\overline{M}_n'G_n\varepsilon_n,\ \frac{1}{n}X_n'G_n\overline{M}_n'G_n\xi_n,\ \frac{1}{n}\varepsilon_n'G_n\overline{M}_n'G_n\xi_n,$$

where $\overline{M}_n = M_n - \mathrm{tr}(M_n)I_n/n$ and M_n is either G_n, G_n', or $G_n'G_n$ in our example if we choose $Q_n = (G_nX_n, G_nZ_n, X_n, Z_n)$. In general, M_n can be I_n, $W_n'^{m_1}W_n^{m_2}$, G_n, G_n', and $G_n'G_n$ for any nonnegative integers m_1 and m_2.

4.2. Assumptions and topological structures

To analyze terms in above key statistics, we need additional assumptions and topological structures.

Assumption 3. (3.1) For any i, j, and n, the spatial weight $w_{ij,n} \geq 0$, $w_{ii,n} = 0$, and $\sup_n \|W_n\|_\infty = c_w < \infty$.

(3.2) The parameter $\theta = (\lambda, \beta', \mathrm{vec}(\Gamma)', \sigma_\xi^2, \alpha', \delta')'$ is in a compact set Θ in the Euclidean space $\mathbf{R}^{k_\theta}$. Here $k_\theta = k_1 + 2 + k_2p_2 + p_2 + J$, where k_1 is the dimension of β, p_2 is the dimension of $\sigma_{v\varepsilon}$, k_2p_2 is the number of parameters in Γ, and J is the dimension of α with α being the vector of all distinct elements in Σ_ε. The true parameter θ_0 is contained in the interior of Θ. Furthermore, $\sup_{\lambda\in\Lambda}|\lambda|c_w < 1$, where Λ is the parameter space for λ.

(3.3) Let the $k \times n$ matrix X_n collect all distinct column vectors in X_{1n} and X_{2n}. All elements in X_n are deterministic and bounded in absolute value. $\lim_{n\to\infty}\frac{1}{n}X_n'X_n$ exists and is nonsingular.

Assumption 4. We consider two cases of W_n:

(4.1) Case 1: The spatial weight $w_{ij,n} = h_{ij}(z_{i,n}, z_{j,n}, \rho_{ij})$ for $i \neq j$, where $h_{ij}(\cdot)$'s are non-negative, uniformly bounded functions of some observable variable Z_n. $0 \leq w_{ij,n} \leq c_1\rho_{ij}^{-c_3d_0}$ for some $0 \leq c_1$ and $c_3 > 3$.[7] Furthermore, there exist at most K ($K \geq 1$) columns of W_n that the column sum exceeds c_w, where K is a fixed number that does not depend n.[8]

(4.2) Case 2: The spatial weight $w_{ij,n} = 0$ if $\rho_{ij} > \rho_c$, i.e., there exists a threshold $\rho_c > 1$ and if the geographic distance exceeds ρ_c, then the weight is zero. For $i \neq j$, $w_{ij,n} = h_{ij}(z_{i,n}, z_{j,n}) \cdot I(\rho_{ij} \leq \rho_c)$ or $w_{ij,n} = h_{ij}(z_{i,n}, z_{j,n}) \cdot I(\rho_{ij} \leq \rho_c)/\sum_{\rho_{ik}\leq\rho_c} h_{ik}(z_{i,n}, z_{k,n})$, where $h_{ij}(\cdot)$'s are non-negative, uniformly bounded functions.

Assumptions 3 and 4 provide the essential features of the weights matrix and parameters for the model. Assumptions (3.1) and (3.2) are standard assumptions in the spatial econometrics literature to limit the spatial correlation in a manageable degree. Assumption (3.3) requires that all distinct regressors in X_{1n} and X_{2n} are linearly independent. Note that Assumption (3.3) allows the special case that X_{1n} and X_{2n} are the same. Due to interactions of W_n and Y_n, and nonlinearity of Z_n in W_n, as contrary to a linear simultaneous equation system, exclusive restrictions on regressors for identification may not be needed. From Assumption 4, we can see that the geographic distance plays an important role in constraining magnitudes of our spatial weights. The spatial weight of two locations would be larger if they were closer to each other or when their economic indices were more similar, but their weights would become smaller when two units are further apart. Assumption (4.1) allows the situation that all agents are spatially correlated but the spatial weight decreases sufficiently fast at a certain rate as physical distances increase. Symmetry is not imposed on the spatial weight matrix. If W_n is indeed symmetric, then by Assumption (3.1), the column sum will also be uniformly bounded by c_w. In that case, the second part on the column sum norm condition in Assumption (4.1) will not be needed. However, in general, W_n can be asymmetric, i.e., $h_{ij}(z_{i,n}, z_{j,n}) \neq h_{ji}(z_{j,n}, z_{i,n})$. For an asymmetric W_n, the second part of Assumption (4.1) limits the number of columns which have large magnitudes relative to the row sum norm. For example, big countries may have great impact on small countries, but those small countries may have little or zero influence on big countries. In this example, we have some "stars" whose row sums are bounded by c_w, while their column sums can be much larger. Assumption (4.1) assumes that the number of such stars can only be finite and bounded. Assumption (4.2), also imposed in Qu and Lee (2012), allows for a row-normalized spatial weight matrix: $w_{ij,n} = h_{ij}(z_{i,n}, z_{j,n}) \cdot I(\rho_{ij} \leq \rho_c)/\sum_{\rho_{ik}\leq\rho_c} h_{ik}(z_{i,n}, z_{k,n})$. In this case, $w_{ij,n}$ might have agents linked in an area, which could be wide, but once the geographic distance between two agents exceeds a threshold, the two units are not spatially interacted.

Our asymptotic analysis of the proposed estimators will be based on inference under NED. The following notion of NED for random fields is similar to Jenish and Prucha (2012).

Definition 1. For any random vector Y, $\|Y\|_p = [E|Y|^p]^{1/p}$ denotes its L_p-norm where $|Y|$ is the Euclidean norm of Y. Denote $\mathcal{F}_{i,n}(s)$ as a σ-field generated by the random vectors $\varsigma_{j,n}$'s located within the ball $B_i(s)$, which is a ball centered at the location $l(i)$ with a radius s in a d_0-dimensional Euclidean space D.

Definition 2 (*NED*). Let $T = \{T_{i,n}, l(i) \in D_n, n \geq 1\}$ and $\varsigma = \{\varsigma_{i,n}, l(i) \in D_n, n \geq 1\}$ be random fields with $\|T_{i,n}\|_p < \infty$, $p \geq 1$, where $D_n \subset D$ and its cardinality $|D_n| = n$, and let $d = \{d_{i,n}, l(i) \in D_n, n \geq 1\}$ be an array of finite positive constants. Then the random field T is said to be L_p-near-epoch dependent on the random field ς if $\|T_{i,n} - E(T_{i,n}|\mathcal{F}_{i,n}(s))\|_p \leq d_{i,n}\varphi(s)$ for some sequence $\varphi(s) \geq 0$ such that $\lim_{s\to\infty}\varphi(s) = 0$. The $\varphi(s)$, which is, without loss of generality, assumed to be non-increasing, is called the NED coefficient, and the $d_{i,n}$'s are called NED scaling factors. T is said to be L_p-NED on ς of size $-\alpha$ if $\varphi(s) = O(s^{-\mu})$ for some $\mu > \alpha > 0$. Furthermore, if $\sup_n \sup_{l(i)\in D_n} d_{i,n} < \infty$, then T is said to be uniformly L_p-NED on ς. If $\varphi(s) = O(\rho^s)$, where $0 < \rho < 1$, then T is called geometrically L_p-NED on ς.

4.3. Asymptotic inference of key statistics

Let $\varsigma_{i,n}^*$ be a vector-valued function of the error term $\varsigma_{i,n} = (\varepsilon_{i,n}', \xi_{i,n})'$ and the observed X_n, i.e., $\varsigma_{i,n}^* = f_i(\varepsilon_{i,n}, \xi_{i,n}, X_n, \theta_0)$. As X_n is deterministic, $\varsigma_{i,n}^*$ is purely determined by the location $l(i)$ (individual i), independent of error terms associated with any other places (individuals). Let $M_n = A_n'B_n$, where A_n and B_n are either

[7] For example, $w_{ij,n} = \min(\frac{1}{\|z_{i,n}-z_{j,n}\|_p}, c_1\rho_{ij}^{-c_3d_0})$.

[8] As $c_0^{-\rho_{ij}}$ decreases faster than $\rho_{ij}^{-c_3d_0}$, all the results hold for the case of $0 \leq w_{ij,n}^d \leq c_1c_0^{-\rho_{ij}}$ with some $c_1 \geq 0$ and $c_0 > 1$.

$W_n^{m_1}$ or $G_n^{m_2}$ with m_1 and m_2 being finite non-negative integers. Denote $\varsigma_n^* = (\varsigma_{1,n}^*, \ldots, \varsigma_{n,n}^*)'$. The NED property of the statistic $a'\varsigma_n^{*\prime}M_n\varsigma_n^* b$ for some constant vectors a and b with $\varsigma_{i,n}$ as the basis for the NED is established in Appendix C.1 under Assumption (4.1) for Case 1 and in Appendix C.2 for Case 2 under Assumption (4.2). Then based on the asymptotic inference under NED, we have the following LLN.

Proposition 1. *Under Assumptions 1, (3.1), and 4, suppose* $\sup_{i,n}\|\varsigma_{i,n}^*\|_4 < \infty$, *then* $\frac{1}{n}E|a'\varsigma_n^{*\prime}M_n\varsigma_n^* b| = O(1)$ *and* $\frac{1}{n}[a'\varsigma_n^{*\prime}M_n\varsigma_n^* b - E(a'\varsigma_n^{*\prime}M_n\varsigma_n^* b)] = o_p(1)$, *where a and b are conformable vectors of constants.*

Furthermore, with the compactness of the parameter space of θ, we have the following ULLN.

Corollary 1. *Under Assumptions 1, (3.1), (3.2), and 4, suppose* $\sup_{i,n}\|\varsigma_{i,n}^*\|_4 < \infty$, *then* $\frac{1}{n}a'\varsigma_n^*(\theta)'G_n^{m_1}(\lambda)'G_n^{m_2}(\lambda)\varsigma_n^*(\theta)b$ *is stochastic equicontinuous and*

$$\sup_{\theta\in\Theta}\frac{1}{n}|a'\varsigma_n^*(\theta)'G_n^{m_1}(\lambda)'G_n^{m_2}(\lambda)\varsigma_n^*(\theta)b - E(a'\varsigma_n^*(\theta)'G_n^{m_1}(\lambda)'G_n^{m_2}(\lambda)\varsigma_n^*(\theta)b)| = o_p(1),$$

where $\varsigma_{i,n}^*(\theta) = f_i(\varepsilon_{i,n}, \xi_{i,n}, X_n, \theta)$ *with* θ *entering* f_i *polynomially,* m_1 *and* m_2 *are finite non-negative integers, and a and b are conformable vectors of constants.*

Denote

$$R_n = \sum_{j=1}^m [a_j'\varsigma_n^{*\prime}M_{jn}\varsigma_n^* b_j - E(a_j'\varsigma_n^{*\prime}M_{jn}\varsigma_n^* b_j)] = \sum_{i=1}^n r_{i,n},$$

where each M_{jn} matrix, $j = 1, \ldots, m$ can be expressed as $M_{jn} = A_{jn}'B_{jn}$ with A_{jn} and B_{jn} being either $W_n^{m_1}$ or $G_n^{m_2}$. Denote σ_{Rn}^2 as the variance of R_n and $r_{i,n} = \sum_{j=1}^m\sum_{k=1}^n [a_j'\varsigma_{i,n}^* M_{jn}(i,k)\varsigma_{k,n}^* b_j - E(a_j'\varsigma_{i,n}^* M_{jn}(i,k)\varsigma_{k,n}^* b_j)]$. Then $R_n = \sum_{i=1}^n r_{i,n}$ and $\sigma_{Rn}^2 = \mathrm{Var}(\sum_{i=1}^n r_{i,n})$. We have the following CLT for R_n.

Proposition 2. *Under Assumptions 1, 2, (3.1), and 4, suppose* $\sup_{i,n}\|\varsigma_{i,n}^*\|_{4+\delta_\varepsilon} < \infty$ *for some* $\delta_\varepsilon > 0$, *and* $\inf_n \frac{1}{n}\sigma_{Rn}^2 > 0$, *then* $R_n/\sigma_{Rn} \xrightarrow{d} N(0,1)$.

The LLN in Proposition 1 and the CLT in Proposition 2 provide the essential tools for asymptotic analysis of the consistency and asymptotic normality of the 2SIV, QMLE and GMM estimators in our model.

4.4. Consistency and asymptotic normality of estimators

The 2SIV

To show the consistency and asymptotic normality of the 2SIV and G2SIV estimators, in addition to the convergence of each separated term, we need some rank conditions on relevant limiting matrices.

Assumption 5. (5.1) Columns of Q_n are from M_nq_n and M_nZ_n, where q_n is a strictly exogenous vector and $M_n = A_n'B_n$, in which A_n and B_n are either $W_n^{m_1}$ or $G_n^{m_2}$ with m_1 and m_2 being finite non-negative integers.

(5.2) $\lim_{n\to\infty}\frac{1}{n}E(Q_n'Q_n)$ exists and is nonsingular;

(5.3) $\lim_{n\to\infty}\frac{1}{n}E[Q_n'(G_n(X_{1n}\beta_0 + \varepsilon_n\delta_0), X_{1n}, \varepsilon_n)]$ has full column rank.

It is of interest to note that endogeneity of W_n in our model may provide parameter identification via the IV estimation, even if there are no relevant regressors X_{1n} in the SAR equation. In the SAR with an exogenous W_n, if there are no regressors X_{1n} in the equation, i.e., $\beta_0 = 0$, its corresponding limiting matrix $\lim_{n\to\infty}\frac{1}{n}E[Q_n'(G_nX_{1n}\beta_0, X_{1n})] = [0, \lim_{n\to\infty}\frac{1}{n}Q_n'X_{1n}]$ would not have full column rank. However, with endogeneity, $\lim_{n\to\infty}\frac{1}{n}E[Q_n'(G_n\varepsilon_n\delta_0, X_{1n}, \varepsilon_n)]$ may have full column rank.

Theorem 1. *Under Assumptions 1–5, the 2SIV estimator* $\widehat{\kappa}$ *and the G2SIV estimator* $\widehat{\kappa}_G$ *are consistent estimators of* κ_0. *Furthermore,* $\sqrt{n}(\widehat{\kappa} - \kappa_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{IV}})$ *and* $\sqrt{n}(\widehat{\kappa}_G - \kappa_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{GIV}})$, *where*

$$\Sigma_{\mathrm{IV}} = \operatorname*{plim}_{n\to\infty}\frac{1}{n}(U_n'A_{qn}U_n)^{-1}U_n'A_{qn}\Pi_nA_{qn}U_n(U_n'A_{qn}U_n)^{-1} \quad \text{and}$$

$$\Sigma_{\mathrm{GIV}} = \operatorname*{plim}_{n\to\infty}\frac{1}{n}[U_n'\Pi_n^{-1}Q_n(Q_n'\Pi_n^{-1}Q_n)^{-1}Q_n'\Pi_n^{-1}U_n]^{-1}$$

with $U_n = [G_n(X_{1n}\beta_0 + \varepsilon_n\delta_0), X_{1n}, \varepsilon_n]$ *and* $A_{qn} = Q_n(Q_n'Q_n)^{-1}Q_n'$.

By the Cauchy–Schwarz inequality, $U_n'\Pi_n^{-1}Q_n(Q_n'\Pi_n^{-1}Q_n)^{-1}Q_n'\Pi_n^{-1}U_n \le U_n'\Pi_n^{-1}U_n$ and the "=" holds if the columns of U_n are in the linear space spanned by the columns of Q_n. Therefore, if column vectors in the IV matrix Q_n consist of G_nX_n, G_nZ_n, X_n, and Z_n, then the best G2SIV estimator based on this optimal Q_n has the smallest limiting variance $\Sigma_{\mathrm{BGIV}} = \operatorname{plim}_{n\to\infty}\frac{1}{n}(U_n'\Pi_n^{-1}U_n)^{-1}$.

However, the best G2SIV estimator is not feasible because $\sigma_{\xi0}^2$ and $\delta_0'\Sigma_{\varepsilon0}\delta_0$ in Π_n as well as λ_0 in G_n are unknown. In practice, we may use X_n W_nX_n, W_nZ_n, etc. as IV matrices to get an initial consistent estimate $\widehat{\kappa}$ by 2SIV, and then using $G_n(\widehat{\lambda})X_n$, $G_n(\widehat{\lambda})Z_n$, X_n, and Z_n as new IVs and substituting $\widehat{\Pi}_n = \widehat{\sigma}_\xi^2 I_n + \widehat{\delta}'\widehat{\Sigma}_\varepsilon\widehat{\delta}P_n$, where $\widehat{\Sigma}_\varepsilon = \frac{1}{n}Z_n'P_n^\perp Z_n$ and $\widehat{\sigma}_\xi^2 = \frac{1}{n}(Y_n - \widehat{\lambda}W_nY_n - X_{1n}\widehat{\beta} - P_n^\perp Z_n\widehat{\delta})'(Y_n - \widehat{\lambda}W_nY_n - X_{1n}\widehat{\beta} - P_n^\perp Z_n\widehat{\delta})$, for Π_n to obtain the feasible best G2SIV estimator $\widehat{\kappa}_{\mathrm{FBGIV}}$. The following theorem shows that $\widehat{\kappa}_{\mathrm{FBGIV}}$ has the same limiting distribution as the best G2SIV estimator.

Theorem 2. *Under Assumptions 1–5, the feasible best G2SIV estimator* $\widehat{\kappa}_{\mathrm{FBGIV}}$ *is a consistent estimator of* κ_0 *and* $\sqrt{n}(\widehat{\kappa}_{\mathrm{FBGIV}} - \kappa_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{BGIV}})$.

The QMLE

Assumption 6. Either (a) $\lim_{n\to\infty}\frac{1}{n}E[(G_n(X_{1n}\beta_0 + \varepsilon_n\delta_0), X_{1n}, \varepsilon_n)'(G_n(X_{1n}\beta_0 + \varepsilon_n\delta_0), X_{1n}, \varepsilon_n)]$ exists and is nonsingular, or (b) $S_n(\lambda)'S_n(\lambda)$ is not proportional to $S_n'S_n$ with probability one whenever $\lambda \ne \lambda_0$.

Assumption 6 is an identification condition for the model. Assumption 6(a) is a rank condition, which is similar to Assumption (5.3) for the 2SIV. Assumption 6(b) explores the i.i.d. disturbances of the model so that the reduced form of Y_n has a unique variance structure. A sufficient condition that guarantees the linear independence of $S_n(\lambda)'S_n(\lambda)$ with $S_n'S_n$ is that the matrices I_n, $(W_n + W_n')$ and $W_n'W_n$ are linearly independent.[9] Assumption 6 also implies that the information matrix of this model is nonsingular as shown in Claim C.3.2.

With identification, the uniform convergence of $\sup_{\theta\in\Theta}\frac{1}{n}\left|\ln L_n(\theta) - E\frac{1}{n}\ln L_n(\theta)\right| \xrightarrow{p} 0$ and the equicontinuity of $\lim_{n\to\infty}\frac{1}{n}E\ln L_n(\theta_0)$ together imply the consistency of the QMLE.

[9] Here is a simple proof: suppose that for some $c \ne 0$, $S_n(\lambda)'S_n(\lambda) = cS_n'S_n$ with probability one. It follows that $(1-c)I_n + (c\lambda_0 - \lambda)(W_n + W_n') + (\lambda^2 - c\lambda_0^2)W_n'W_n = 0$ with probability one. Under the linear independence of I_n, $(W_n + W_n')$, and $W_n'W_n$, it must be $c = 1$ and $\lambda_0 = \lambda$.

Theorem 3. *Under Assumptions 1–4, and 6, the QMLE $\widehat{\theta}$ is a consistent estimator of θ_0 and $\sqrt{n}(\widehat{\theta}-\theta_0) \xrightarrow{d} N(0, \Sigma_{\text{QMLE}})$, where*

$$\Sigma_{\text{QMLE}} = \left(\lim_{n\to\infty}\frac{1}{n}E\left(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right)\right)^{-1} \times \lim_{n\to\infty}\frac{1}{n}E\left(\frac{\partial \ln L_n(\theta_0)}{\partial\theta}\frac{\partial \ln L_n(\theta_0)}{\partial\theta'}\right) \times \left(\lim_{n\to\infty}\frac{1}{n}E\left(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right)\right)^{-1}.$$

Expressions for each term of Σ_{QMLE} are in Appendix A. In the special case that $(v_{i,n}, \varepsilon'_{i,n})'$ is jointly normal, QMLE becomes MLE and the asymptotic variance is simply $-\left(\lim_{n\to\infty}E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'})\right)^{-1}$.

The GMM

One advantage of the GMM approach compared to the QMLE method is that the GMM estimator can be computationally simpler as the determinant of the Jacobian transformation, $|I_n - \lambda W_n|$, needs not to be evaluated whereas with QMLE it does. To prove the consistency and asymptotic normality of the GMM estimator, we impose following assumptions.

Assumption 7. (7.1) The $n\times m^*$ IV matrix Q_n has its columns from M_nq_n and M_nZ_n, where q_n is a strictly exogenous vector and $M_n = A'_nB_n$, in which A_n and B_n are either $W_n^{m_1}$ or $G_n^{m_2}$ with m_1, and m_2 being non-negative integers. The $n \times n$ square matrices $P_{jn} = M_{jn} - \text{tr}(M_{jn})I_n/n$ ($j = 1, \dots, m$ for some finite m) have zero trace.

(7.2) $\text{plim}_{n\to\infty}\frac{1}{n}a_ng_n(\theta^G) = 0$ has a unique root at θ_0^G in Θ^G.

(7.3) $\text{plim}_{n\to\infty}\frac{1}{n}a_nD_n$ exists and has the full rank $(1+k_1+k_2p_2+p_2)$, where $D_n = -\text{plim}_{n\to\infty}\frac{1}{n}\frac{\partial g_n(\theta_0^G)}{\partial\theta^{G\prime}}$.

For simplicity, (7.2) in Assumption 7 is a high level sufficient condition for identification. Given specific moments as suggested in Section 3.3, it is possible to have Assumption (7.2) satisfied with some sufficient conditions on Q_n and P_{jn}'s as in Lee (2007). The simplest sufficient condition is the ability to construct consistent IV estimation of the model equations by some proper IV matrix Q_n, as in Assumption 5.

By applying Propositions 1 and 2, we have the following theorem.

Theorem 4. *Under Assumptions 1–4, and 7, the GMM estimator $\widehat{\theta}_n^G = \arg\min_{\theta\in\Theta} g'_n(\theta^G)a'_na_ng_n(\theta^G)$ is a consistent estimator of θ_0^G, and $\sqrt{n}(\widehat{\theta}_n^G-\theta_0^G) \xrightarrow{d} N(0, \Sigma_{\text{GMM}})$, where*

$$\Sigma_{\text{GMM}} = \lim_{n\to\infty}\frac{1}{n}(D'_na'_na_nD_n)^{-1}D'_na'_na_n\Omega_n(\theta_0^G)a'_n \times a_nD_n(D'_na'_na_nD_n)^{-1},$$

with $D_n = -\frac{1}{n}\frac{\partial(g_n(\theta_0^G))}{\partial\theta^{G\prime}}$ and $\Omega_n(\theta_0^G) = \text{Var}(g_n(\theta_0^G))$.

Detailed expressions of D_n and $\Omega_n(\theta_0^G)$ are in (C.5) and (C.6) of Appendix C. By the generalized Cauchy–Schwarz inequality, the optimal weighting matrix for the GMM estimation with the moment functions $g_n(\theta^G)$ is $[\Omega_n(\theta_0^G)]^{-1}$. Then, with a consistent estimator $\widehat{\Omega}_n$ of $\Omega_n(\theta_0^G)$, the feasible "optimal" GMM is obtained from $\min_{\theta\in\Theta} g'_n(\theta^G)\widehat{\Omega}_n^{-1}g_n(\theta^G)$ and it will have the smallest asymptotic variance $(\lim_{n\to\infty}\frac{1}{n}D'_n[\Omega_n(\theta_0^G)]^{-1}D_n)^{-1}$.

4.5. Estimated variance–covariance matrix of estimators

For QMLE, all parameters in θ are jointly estimated, so directly we have a consistent estimator of $\sigma^2_{\xi_0}$. For 2SIV and GMM methods, we do not estimate $\sigma^2_{\xi_0}$ directly and therefore need to construct a consistent estimator for it. Expressions for the estimated variance–covariance matrix of Σ_{IV} and Σ_{BGIV} are based on the following result.

Claim 1. *Suppose $(\widehat{\lambda}, \widehat{\beta}', \widehat{\gamma}', \widehat{\delta})'$ is a consistent estimator of $(\lambda_0, \beta'_0, \gamma'_0, \delta_0)'$, then $\widehat{\sigma}^2_\xi = \frac{1}{n}\widehat{\xi}'_n\widehat{\xi}_n$ is a consistent estimator of $\sigma^2_{\xi_0}$, where $\widehat{\xi}_n = S_n(\widehat{\lambda})Y_n - X_{1n}\widehat{\beta} - (Z_n - X_{2n}\widehat{\Gamma})\widehat{\delta}$. Furthermore, if $(\lambda_0, \beta'_0, \gamma'_0, \delta_0)'$ is replaced with $(\widehat{\lambda}, \widehat{\beta}', \widehat{\gamma}', \widehat{\delta})'$ and ε_n with $\widehat{\varepsilon}_n = Z_n - X_{2n}\widehat{\Gamma}$ in Σ_{IV} and Σ_{BGIV} to obtain, respectively, empirical estimates $\widehat{\Sigma}_{\text{IV}}$ and $\widehat{\Sigma}_{\text{BGIV}}$, then $\widehat{\Sigma}_{\text{IV}} \xrightarrow{p} \Sigma_{\text{IV}}$ and $\widehat{\Sigma}_{\text{BGIV}} \xrightarrow{p} \Sigma_{\text{BGIV}}$.*

Based on this claim, the estimated asymptotic variance–covariance matrices for the 2SIV estimator $\widehat{\kappa}$ and the feasible best G2SIV estimator $\widehat{\kappa}_{\text{FBGIV}}$ are, respectively,

$$\frac{1}{n}\widehat{\Sigma}_{\text{IV}} = (\widehat{U}'_nA_{qn}\widehat{U}_n)^{-1}\widehat{U}'_nA_{qn}\widehat{\Pi}_nA_{qn}\widehat{U}_n(\widehat{U}'_nA_{qn}\widehat{U}_n)^{-1} \quad \text{and}$$

$$\frac{1}{n}\widehat{\Sigma}_{\text{BGIV}} = (\widehat{U}'_n\widehat{\Pi}_n^{-1}\widehat{U}_n)^{-1},$$

where

$$\widehat{U}_n = [G_n(\widehat{\lambda})(X_{1n}\widehat{\beta} + P_n^{\perp}Z_n\widehat{\delta}), X_{1n}, P_n^{\perp}Z_n] \quad \text{and}$$

$$\widehat{\Pi}_n = \widehat{\sigma}^2_\xi I_n + \widehat{\delta}'\widehat{\Sigma}_\varepsilon\widehat{\delta}P_n \quad \text{with}$$

$$\widehat{\Sigma}_\varepsilon = \frac{1}{n}Z'_nP_n^{\perp}Z_n \quad \text{and}$$

$$\widehat{\sigma}^2_\xi = \frac{1}{n}(Y_n - \widehat{\lambda}W_nY_n - X_{1n}\widehat{\beta} - P_n^{\perp}Z_n\widehat{\delta})' \times (Y_n - \widehat{\lambda}W_nY_n - X_{1n}\widehat{\beta} - P_n^{\perp}Z_n\widehat{\delta}).$$

For Σ_{QMLE} and Σ_{GMM}, we have not only similar terms as those in Σ_{IV}, but also special ones involving the third and fourth orders of ξ_{in}, such as $\frac{1}{n}\sum_{i=1}^n E[\xi^3_{i,n}G_{in}(X_{1n}\beta_0 + \varepsilon_n\delta_0)G_{ii,n}]$ and $\frac{1}{n}\sum_{i=1}^n E(\xi^4_{i,n}G_{ii,n})$. But they can be estimated by empirical moments with estimated coefficients.

Claim 2. *If θ_0 is replaced with a consistent estimator $\widehat{\theta}$, ε_n with $\widehat{\varepsilon}_n = Z_n - X_{2n}\widehat{\Gamma}$, and ξ_{in} with $\widehat{\xi}_{in}$, where $\widehat{\xi}_{in}$ is the ith element of $\widehat{\xi}_n = S_n(\widehat{\lambda})Y_n - X_{1n}\widehat{\beta} - (Z_n - X_{2n}\widehat{\Gamma})\widehat{\delta}$, in Σ_{QMLE} and Σ_{GMM} to obtain, respectively, empirical estimates $\widehat{\Sigma}_{\text{QMLE}}$ and $\widehat{\Sigma}_{\text{GMM}}$, then $\widehat{\Sigma}_{\text{QMLE}} \xrightarrow{p} \Sigma_{\text{QMLE}}$ and $\widehat{\Sigma}_{\text{GMM}} \xrightarrow{p} \Sigma_{\text{GMM}}$.*

4.6. Comparison of the three estimation methods

If error terms are jointly normally distributed, the QMLE becomes MLE and achieves the asymptotic efficiency. The 2SIV estimation is based on linear moments only and therefore is asymptotically less efficient. For the GMM estimator based on some proper linear and quadratic moment conditions, it can be asymptotically as efficient as the ML estimator under normality.

In the absence of normality, the QMLE is no longer asymptotic efficient. The optimum GMM estimator based on linear and quadratic moment conditions obtained from relevant components in the first order condition of the QMLE function might be asymptotically more efficient than the QMLE, because it adopts the best weighting matrix for those moment conditions, while the QMLE gives each moment an equal weight. As that GMM estimator is based on linear and quadratic moment conditions, it can also be asymptotically more efficient than the 2SIV. However,

Table 1
Estimates from spatial weight matrices with weak endogeneity (small sample).

$\rho = 0.2$	*WS* ($n = 49$)				*WO* ($n = 98$)			
$\lambda = 0.2$	IV	2SIV	SAR	MLE	IV	2SIV	SAR	MLE
$\widehat{\lambda}$	0.1229 (0.2519) [0.2560]	0.1922 (0.2422) [0.2659]	0.1391 (0.1362) [0.1293]	0.1808 (0.1306) [0.1286]	0.0784 (0.2085) [0.2043]	0.1749 (0.1986) [0.2095]	0.1241 (0.1142) [0.1027]	0.1777 (0.1087) [0.1010]
$\widehat{\beta}_1$	1.0993 (0.3743) [0.3717]	1.0036 (0.3632) [0.3773]	1.0808 (0.2324) [0.2283]	1.0234 (0.2266) [0.2324]	1.1680 (0.2993) [0.3029]	1.0382 (0.2873) [0.3044]	1.1048 (0.1834) [0.1792]	1.0327 (0.1774) [0.1832]
$\widehat{\beta}_2$	0.9759 (0.1604) [0.1684]	0.9815 (0.1606) [0.1686]	0.9884 (0.1505) [0.1588]	0.9915 (0.1505) [0.1593]	0.9675 (0.1173) [0.1228]	0.9852 (0.1182) [0.1230]	0.9810 (0.1103) [0.1158]	0.9906 (0.1105) [0.1172]
$\widehat{\gamma}_1$		1.0044 (0.1419) [0.1498]		1.0044 (0.1390) [0.1498]		1.0020 (0.1013) [0.1056]		1.0020 (0.1002) [0.1056]
$\widehat{\gamma}_2$		0.7987 (0.1533) [0.1604]		0.7987 (0.1502) [0.1604]		0.8038 (0.1102) [0.1108]		0.8038 (0.1090) [0.1108]
$\widehat{\delta}$		0.2012 (0.1545) [0.1616]		0.1996 (0.1416) [0.1523]		0.1994 (0.1072) [0.1093]		0.1998 (0.1003) [0.1023]
$\lambda = 0.4$	IV	2SIV	SAR	MLE	IV	2SIV	SAR	MLE
$\widehat{\lambda}$	0.3198 (0.2402) [0.2537]	0.3857 (0.2305) [0.2593]	0.3287 (0.1253) [0.1229]	0.3736 (0.1184) [0.1262]	0.2792 (0.2015) [0.2028]	0.3717 (0.1883) [0.2029]	0.3159 (0.1046) [0.0994]	0.3742 (0.0977) [0.1005]
$\widehat{\beta}_1$	1.1386 (0.4614) [0.4784]	1.0173 (0.4456) [0.4799]	1.1273 (0.2676) [0.2705]	1.0453 (0.2571) [0.2684]	1.2201 (0.3742) [0.3876]	1.0557 (0.3519) [0.3822]	1.1525 (0.2117) [0.2141]	1.0487 (0.2008) [0.2105]
$\widehat{\beta}_2$	0.9803 (0.1600) [0.1670]	0.9815 (0.1598) [0.1670]	0.9924 (0.1509) [0.1587]	0.9929 (0.1503) [0.1587]	0.9729 (0.1158) [0.1207]	0.9855 (0.1159) [0.1204]	0.9836 (0.1100) [0.1150]	0.9912 (0.1096) [0.1158]
$\widehat{\gamma}_1$		1.0044 (0.1419) [0.1498]		1.0044 (0.1390) [0.1498]		1.0020 (0.1013) [0.1056]		1.0020 (0.1002) [0.1056]
$\widehat{\gamma}_2$		0.7987 (0.1533) [0.1604]		0.7987 (0.1502) [0.1604]		0.8038 (0.1102) [0.1108]		0.8038 (0.1090) [0.1108]
$\widehat{\delta}$		0.2002 (0.1535) [0.1599]		0.1991 (0.1412) [0.1415]		0.1995 (0.1060) [0.1076]		0.1999 (0.0998) [0.1016]

Note: Observations $n = 49$ or 98, $\beta_1 = \beta_2 = \gamma_1 = 1$, and $\gamma_2 = 0.8$. Estimated standard error based on an asymptotic variance–covariance matrix is in parentheses; and empirical standard deviation is in brackets.

the asymptotic efficiency of the 2SIV and the QMLE cannot be directly compared in general.[10]

The conditional homoskedasticity assumption in Assumption 2 is required for the consistency of QMLE as in the traditional SAR framework. However, we can relax this assumption for the 2SIV and some GMM estimation. Therefore, the 2SIV and GMM methods have the merit of computational simplicity and robustness.

5. Some extensions

5.1. Extension to nonlinear conditional mean

Our previous analysis is based on the linear conditional mean $E(v_{i,n}|\varepsilon_{i,n}) = \varepsilon_{i,n}\delta$ in Assumption 2. As a possible generalization, the linear conditional mean can be relaxed to a polynomial function with little additional complication for our proposed estimators. For simplicity, assume $p_2 = 1$ and $E(v_{i,n}|\varepsilon_{i,n}) = \sum_{m=1}^{\overline{m}} \varepsilon_{i,n}^m \delta_m$, where $\overline{m}$ is a finite positive integer. For an $n \times 1$ vector $b = (b_i)$, b^m denotes an $n \times 1$ vector with the ith element as b_i^m. Then Eq. (2.3) can be generalized to

$$Y_n = \lambda W_n Y_n + X_{1n}\beta + \sum_{m=1}^{\overline{m}} (Z_n - X_{2n}\gamma)^m \delta_m + \xi_n.$$

The log quasi-likelihood function is

$$\begin{aligned} \ln L_n(\theta) &= \ln[f(Z_n) f(Y_n|Z_n)] = -n\ln(2\pi) - \frac{n}{2}\ln \sigma_\xi^2 \sigma_\varepsilon^2 \\ &\quad + \ln|S_n(\lambda)| - \frac{1}{2\sigma_\varepsilon^2}(Z_n - X_{2n}\gamma)'(Z_n - X_{2n}\gamma) \\ &\quad - \frac{1}{2\sigma_\xi^2}\left(S_n(\lambda)Y_n - X_{1n}\beta - \sum_{m=1}^{\overline{m}}(Z_n - X_{2n}\gamma)^m\delta_m\right)' \\ &\quad \times \left(S_n(\lambda)Y_n - X_{1n}\beta - \sum_{m=1}^{\overline{m}}(Z_n - X_{2n}\gamma)^m\delta_m\right). \end{aligned}$$

10 Liu et al. (2010) have derived the best selection of moments for GMM estimation. However, due to complexity of the model with endogenous spatial weights matrix, the construction of the best GMM moments remains an open question.

Table 2
Estimates from spatial weight matrices with weak endogeneity (large sample).

$\rho = 0.2$	WA ($n = 361$)				WC ($n = 361$)			
$\lambda = 0.2$	IV	2SIV	SAR	MLE	IV	2SIV	SAR	MLE
$\widehat{\lambda}$	0.1187	0.1986	0.1512	0.1963	0.1536	0.1987	0.1743	0.1954
	(0.0908)	(0.0850)	(0.0555)	(0.0531)	(0.0744)	(0.0691)	(0.0418)	(0.0403)
	[0.0868]	[0.0859]	[0.0541]	[0.0544]	[0.0731]	[0.0697]	[0.0417]	[0.0403]
$\widehat{\beta}_1$	1.1082	1.0036	1.0654	1.0066	1.0630	1.0037	1.0356	1.0079
	(0.1300)	(0.1232)	(0.0896)	(0.0871)	(0.1116)	(0.1053)	(0.0762)	(0.0747)
	[0.1275]	[0.1238]	[0.0882]	[0.0868]	[0.1114]	[0.1060]	[0.0753]	[0.0731]
$\widehat{\beta}_2$	0.9919	0.9994	0.9961	1.0003	0.9945	0.9990	0.9981	1.0001
	(0.0562)	(0.0561)	(0.0554)	(0.0554)	(0.0562)	(0.0560)	(0.0554)	(0.0553)
	[0.0565]	[0.0563]	[0.0552]	[0.0553]	[0.0564]	[0.0561]	[0.0553]	[0.0553]
$\widehat{\gamma}_1$		0.9966		0.9966		0.9966		0.9966
		(0.0528)		(0.0526)		(0.0528)		(0.0526)
		[0.0543]		[0.0543]		[0.0543]		[0.0543]
$\widehat{\gamma}_2$		0.8005		0.8005		0.8005		0.8005
		(0.0555)		(0.0553)		(0.0555)		(0.0553)
		[0.0555]		[0.0555]		[0.0555]		[0.0555]
$\widehat{\delta}$		0.2010		0.2005		0.2008		0.2004
		(0.0536)		(0.0521)		(0.0524)		(0.0516)
		[0.0553]		[0.0543]		[0.0541]		[0.0535]
$\lambda = 0.4$	IV	2SIV	SAR	MLE	IV	2SIV	SAR	MLE
$\widehat{\lambda}$	0.3233	0.3980	0.3501	0.3952	0.3591	0.3982	0.3764	0.3955
	(0.0855)	(0.0783)	(0.0504)	(0.0476)	(0.0649)	(0.0590)	(0.0353)	(0.0337)
	[0.0819]	[0.0792]	[0.0499]	[0.0487]	[0.0637]	[0.0596]	[0.0361]	[0.0337]
$\widehat{\beta}_1$	1.1355	1.0053	1.0886	1.0100	1.0734	1.0051	1.0430	1.0096
	(0.1582)	(0.1462)	(0.1024)	(0.0981)	(0.1256)	(0.1161)	(0.0812)	(0.0790)
	[0.1549]	[0.1477]	[0.1022]	[0.0983]	[0.1251]	[0.1172]	[0.0813]	[0.0774]
$\widehat{\beta}_2$	0.9974	0.9994	0.9994	1.0006	1.0000	0.9991	1.0012	1.0007
	(0.0560)	(0.0556)	(0.0554)	(0.0552)	(0.0561)	(0.0556)	(0.0554)	(0.0552)
	[0.0556]	[0.0556]	[0.0551]	[0.0552]	[0.0554]	[0.0554]	[0.0551]	[0.0551]
$\widehat{\gamma}_1$		0.9966		0.9966		0.9966		0.9966
		(0.0528)		(0.0526)		(0.0528)		(0.0526)
		[0.0543]		[0.0543]		[0.0543]		[0.0543]
$\widehat{\gamma}_2$		0.8005		0.8005		0.8005		0.8005
		(0.0555)		(0.0553)		(0.0555)		(0.0553)
		[0.0555]		[0.0555]		[0.0555]		[0.0555]
$\widehat{\delta}$		0.2010		0.2005		0.2008		0.2006
		(0.0529)		(0.0519)		(0.0521)		(0.0516)
		[0.0547]		[0.0540]		[0.0538]		[0.0534]

Note: Observations $n = 361$, $\beta_1 = \beta_2 = \gamma_1 = 1$, and $\gamma_2 = 0.8$. Estimated standard error based on an asymptotic variance–covariance matrix is in parentheses; and empirical standard deviation is in brackets.

And the possible set of linear moments for GMM estimation is $E(X_n'\varepsilon_n) = 0$, $E(X_n'\xi_n) = 0$, $E(Z_n'\xi_n) = 0$, $E((G_nX_n)'\xi_n) = 0$, and $E((G_n(Z_n - X_{2n}\gamma)^m)'\xi_n) = 0$ for $m = 1, \ldots, \overline{m}$. Note that

$$\begin{aligned}\xi_n(\theta) &= S_n(\lambda)Y_n - X_{1n}\beta - \sum_{m=1}^{\overline{m}}(Z_n - X_{2n}\gamma)^m\delta_m \\ &= S_n(\lambda)S_n^{-1}\left(X_{1n}\beta_0 + \sum_{m=1}^{\overline{m}}\varepsilon_n^m\delta_{m0} + \xi_n\right) - X_{1n}\beta \\ &\quad - \sum_{m=1}^{\overline{m}}[X_{2n}(\gamma_0 - \gamma) + \varepsilon_n]^m\delta_m \\ &= (\lambda_0 - \lambda)G_n\left(X_{1n}\beta_0 + \sum_{m=1}^{\overline{m}}\varepsilon_n^m\delta_{m0}\right) + X_{1n}(\beta_0 - \beta) \\ &\quad + \sum_{m=1}^{\overline{m}}\varepsilon_n^m(\delta_{m0} - \delta_m) - \sum_{m=1}^{\overline{m}}\{[X_{2n}(\gamma_0 - \gamma) + \varepsilon_n]^m \\ &\quad - \varepsilon_n^m\}\delta_m + [I_n - (\lambda - \lambda_0)G_n]\xi_n.\end{aligned}$$

Then in this general setting, the new additional statistics involve higher orders of ε_n, e.g., $\frac{1}{n}\varepsilon_n^{l_1\prime}M_n\varepsilon_n^{l_2}$ for $l_1, l_2, = 1, \ldots, \overline{m}$. But Claims C.1.6, C.1.7, C.2.5 and C.2.6 are general enough to ensure the NED property of these statistics, so the 2SIV, QMLE and GMM approaches can still be applied here.

5.2. An LM test of the endogeneity of the spatial weight matrix

To test the endogeneity of W_n, our null hypothesis test is $H_0 : \varepsilon_n$ and V_n are independent and the alternative hypothesis is $H_1 : \varepsilon_n$ and V_n are not independent.

Under the jointly normal assumption of ε_n and V_n, i.e., $(v_{i,n}, \varepsilon_{i,n}')' \sim$ i.i.d. $N\left(0, \begin{pmatrix}\sigma_v^2 & \sigma_{v\varepsilon}' \\ \sigma_{v\varepsilon} & \Sigma_\varepsilon\end{pmatrix}\right)$, to test the independence of ε_n and V_n, we can test whether the covariance $\sigma_{v\varepsilon}$ is zero, or equivalently, whether the corresponding vector of correlation coefficients is zero. Under the joint normality, our hypothesis test is equivalent to $H_0 : \eta = 0$ vs. $H_1 : \eta \neq 0$, where $\eta = \Sigma_\varepsilon^{-1/2}\sigma_{v\varepsilon}/\sigma_v$ is the vector of generalized correlation coefficients. From the quasi log likelihood function (3.4), the LM test can be constructed from

$$g_n(\widehat{\tau}) = \frac{1}{\widetilde{\sigma}_v}\widetilde{\Sigma}_\varepsilon^{-1/2}(Z_n - X_{2n}\widetilde{\Gamma})'(S_n(\widetilde{\lambda})Y_n - X_{1n}\widetilde{\beta}),$$

Table 3
Estimates from spatial weight matrices with medium endogeneity (small sample).

$\rho = 0.5$	*WS* ($n = 49$)				*WO* ($n = 98$)			
$\lambda = 0.2$	IV	2SIV	SAR	MLE	IV	2SIV	SAR	MLE
$\widehat{\lambda}$	0.0318	0.1967	0.0839	0.1835	−0.0455	0.1805	0.0531	0.1819
	(0.2554)	(0.2004)	(0.1384)	(0.1164)	(0.2163)	(0.1602)	(0.1169)	(0.0954)
	[0.2378]	[0.2171]	[0.1292]	[0.1227]	[0.1924]	[0.1672]	[0.1048]	[0.0974]
$\widehat{\beta}_1$	1.2234	0.9985	1.1560	1.0197	1.3352	1.0298	1.2005	1.0268
	(0.3781)	(0.3124)	(0.2344)	(0.2123)	(0.3097)	(0.2401)	(0.1863)	(0.1633)
	[0.3550]	[0.3207]	[0.2339]	[0.2145]	[0.2964]	[0.2508]	[0.1875]	[0.1689]
$\widehat{\beta}_2$	0.9626	0.9860	0.9794	0.9920	0.9342	0.9866	0.9615	0.9905
	(0.1601)	(0.1603)	(0.1501)	(0.1510)	(0.1185)	(0.1177)	(0.1100)	(0.1109)
	[0.1684]	[0.1653]	[0.1580]	[0.1588]	[0.1268]	[0.1214]	[0.1158]	[0.1173]
$\widehat{\gamma}_1$		1.0033		1.0033		1.0023		1.0023
		(0.1422)		(0.1392)		(0.1014)		(0.1004)
		[0.1475]		[0.1475]		[0.1051]		[0.1051]
$\widehat{\gamma}_2$		0.7984		0.7984		0.8022		0.8022
		(0.1536)		(0.1505)		(0.1103)		(0.1092)
		[0.1596]		[0.1596]		[0.1130]		[0.1130]
$\widehat{\delta}$		0.5050		0.5014		0.4995		0.4989
		(0.1376)		(0.1257)		(0.0960)		(0.0893)
		[0.1476]		[0.1384]		[0.1008]		[0.0954]
$\lambda = 0.4$								
$\widehat{\lambda}$	0.2318	0.3921	0.2843	0.3773	0.1593	0.3788	0.2606	0.3792
	(0.2510)	(0.1909)	(0.1287)	(0.1062)	(0.2144)	(0.1515)	(0.1082)	(0.0864)
	[0.2417]	[0.2116]	[0.1250]	[0.1127]	[0.1951]	[0.1600]	[0.1025]	[0.0894]
$\widehat{\beta}_1$	1.2981	1.0072	1.2078	1.0384	1.4337	1.0418	1.2510	1.0396
	(0.4796)	(0.3792)	(0.2727)	(0.2394)	(0.3972)	(0.2901)	(0.2175)	(0.1841)
	[0.4638]	[0.4033]	[0.2791]	[0.2462]	[0.3824]	[0.3075]	[0.2242]	[0.1926]
$\widehat{\beta}_2$	0.9733	0.9858	0.9874	0.9930	0.9471	0.9870	0.9705	0.9910
	(0.1609)	(0.1589)	(0.1509)	(0.1505)	(0.1177)	(0.1153)	(0.1099)	(0.1099)
	[0.1670]	[0.1637]	[0.1579]	[0.1580]	[0.1246]	[0.1190]	[0.1152]	[0.1159]
$\widehat{\gamma}_1$		1.0033		1.0033		1.0023		1.0023
		(0.1353)		(0.1392)		(0.1048)		(0.1004)
		[0.1475]		[0.1475]		[0.1051]		[0.1051]
$\widehat{\gamma}_2$		0.7984		0.7984		0.8022		0.8022
		(0.1422)		(0.1505)		(0.1103)		(0.1092)
		[0.1596]		[0.1596]		[0.1130]		[0.1130]
$\widehat{\delta}$		0.5043		0.5013		0.5000		0.4992
		(0.1536)		(0.1248)		(0.0938)		(0.0884)
		[0.1441]		[0.1371]		[0.0983]		[0.0943]

Note: Observations $n = 49$ or 98, $\beta_1 = \beta_2 = \gamma_1 = 1$, and $\gamma_2 = 0.8$. Estimated standard error based on an asymptotic variance–covariance matrix is in parentheses; and empirical standard deviation is in brackets.

where $\widehat{\tau} = (\widetilde{\lambda}, \widetilde{\beta}', \mathrm{vec}(\widetilde{\Gamma})', \widetilde{\sigma}_v^2, \alpha(\widetilde{\Sigma}_\varepsilon)')'$ is the QMLE from the constrained model, i.e., $\widetilde{\Gamma}$ and $\widetilde{\Sigma}_\varepsilon$ are the QMLE of the model (2.1) and $(\widetilde{\lambda}, \widetilde{\beta}', \widetilde{\sigma}_v^2)'$ is the QMLE of (2.2). From this expression, we can see the LM statistic is based on the estimated sample correlation coefficients of ε_n and V_n. If W_n is exogenous, we would expect this coefficient vector very close to zero. Denote τ_0 as the true value, then $g_n(\tau_0) = \frac{1}{\sigma_{v0}} \Sigma_{\varepsilon 0}^{-1/2} \varepsilon_n' V_n$. By the mean value theorem, we have

$$g_n(\widehat{\tau}) = g_n(\tau_0) - \frac{1}{n}\frac{\partial g_n(\overline{\tau})}{\partial \tau'}\left(\frac{1}{n}\frac{\partial^2 \ln L_c(\widetilde{\tau})}{\partial \tau \partial \tau'}\right)^{-1}\frac{\partial \ln L_c(\tau_0)}{\partial \tau},$$

where $\overline{\tau}$ and $\widetilde{\tau}$ are values between $\widehat{\tau}$ and τ_0, and $\ln L_c(\tau)$ is the log likelihood function of the constrained model. Under the null of independence between ε_n and V_n, $E\left(\frac{1}{n}\frac{\partial g_n(\tau_0)}{\partial \tau'}\right) = 0$ and hence,

$$\frac{1}{\sqrt{n}} g_n(\widehat{\tau}) = \frac{1}{\sqrt{n}} g_n(\tau_0) + o_p(1) \xrightarrow{d} N(0, I_p).$$

Therefore, our LM statistic is

$$\mathrm{LM} = \frac{1}{\widetilde{\sigma}_v^2}(S_n(\widetilde{\lambda})Y_n - X_{1n}\widetilde{\beta})'(Z_n - X_{2n}\widetilde{\Gamma})\widetilde{\Sigma}_\varepsilon^{-1} \times (Z_n - X_{1n}\widetilde{\Gamma})'(S_n(\widetilde{\lambda})Y_n - X_{2n}\widetilde{\beta}) \xrightarrow{d} \chi_p^2.$$

Although it is derived from the log likelihood function under the joint normality of ε_n and V_n, asymptotic distribution of this LM statistic is valid in the absence of the jointly normal assumption. Proof is straightforward under the null of independence between ε_n and V_n, so it is omitted in the paper and available upon request. Alternatively, a test can be set up via a Wald test. Under normality, likelihood ratio test can be constructed too. In a more general setting, how to test the endogeneity of spatial weights is an open question worth further research.

6. Monte Carlo simulations

6.1. Data generating process

In this section, we evaluate four estimation methods of a SAR with an endogenous W_n. The data generating process (DGP) is

$Y_n = (I_n - \lambda W_n)^{-1}(X_n\beta + V_n)$,

where $x_{i,n} = (x_{i1,n}, x_{i2,n})'$ with $x_{i1,n} = 1$ and $x_{i2,n} \sim N(0, 1)$; $\beta_1 = \beta_2 = 1$. Here we let $X_{1n} = X_{2n} = X_n$. The endogenous, row-normalized $W_n = (w_{ij,n})$ is constructed as follows:

Table 4
Estimates from spatial weight matrices with medium endogeneity (large sample).

$\rho = 0.5$	WA ($n = 361$)				WC ($n = 361$)			
$\lambda = 0.2$	IV	2SIV	SAR	MLE	IV	2SIV	SAR	MLE
$\widehat{\lambda}$	0.0168 (0.0933) [0.0806]	0.2004 (0.0698) [0.0694]	0.0853 (0.0567) [0.0536]	0.1976 (0.0465) [0.0468]	0.0935 (0.0772) [0.0710]	0.2002 (0.0580) [0.0562]	0.1441 (0.0426) [0.0431]	0.1970 (0.0359) [0.0352]
$\widehat{\beta}_1$	1.2419 (0.1330) [0.1235]	1.0012 (0.1059) [0.1034]	1.1520 (0.0907) [0.0901]	1.0048 (0.0805) [0.0781]	1.1426 (0.1150) [0.1109]	1.0016 (0.0932) [0.0902]	1.0756 (0.0770) [0.0780]	1.0057 (0.0708) [0.0683]
$\widehat{\beta}_2$	0.9748 (0.0566) [0.0575]	1.0001 (0.0563) [0.0563]	0.9851 (0.0553) [0.0551]	1.0005 (0.0555) [0.0554]	0.9855 (0.0568) [0.0570]	0.9998 (0.0560) [0.0560]	0.9934 (0.0554) [0.0552]	1.0003 (0.0554) [0.0553]
$\widehat{\gamma}_1$		0.9976 (0.0528) [0.0542]		0.9976 (0.0527) [0.0542]		0.9976 (0.0528) [0.0542]		0.9976 (0.0527) [0.0542]
$\widehat{\gamma}_2$		0.8009 (0.0555) [0.0560]		0.8009 (0.0553) [0.0560]		0.8009 (0.0555) [0.0560]		0.8009 (0.0553) [0.0560]
$\widehat{\delta}$		0.5000 (0.0479) [0.0498]		0.4992 (0.0464) [0.0484]		0.4998 (0.0464) [0.0478]		0.4991 (0.0457) [0.0472]
$\lambda = 0.4$								
$\widehat{\lambda}$	0.2264 (0.0899) [0.0781]	0.3998 (0.0646) [0.0642]	0.2993 (0.0521) [0.0505]	0.3966 (0.0420) [0.0423]	0.3065 (0.0687) [0.0631]	0.3997 (0.0496) [0.0481]	0.3574 (0.0360) [0.0366]	0.3969 (0.0302) [0.0296]
$\widehat{\beta}_1$	1.3050 (0.1657) [0.1523]	1.0019 (0.1248) [0.1223]	1.1774 (0.1050) [0.1052]	1.0075 (0.0902) [0.0882]	1.1658 (0.1320) [0.1261]	1.0024 (0.1018) [0.0987]	1.0764 (0.0823) [0.0831]	1.0071 (0.0746) [0.0719]
$\widehat{\beta}_2$	0.9882 (0.0565) [0.0563]	1.0001 (0.0557) [0.0556]	0.9941 (0.0554) [0.0550]	1.0006 (0.0553) [0.0552]	0.9989 (0.0567) [0.0556]	0.9998 (0.0555) [0.0553]	1.0005 (0.0556) [0.0551]	1.0008 (0.0552) [0.0551]
$\widehat{\gamma}_1$		0.9976 (0.0528) [0.0542]		0.9976 (0.0527) [0.0542]		0.9976 (0.0528) [0.0542]		0.9976 (0.0527) [0.0542]
$\widehat{\gamma}_2$		0.8009 (0.0555) [0.0560]		0.8009 (0.0553) [0.0560]		0.8009 (0.0555) [0.0560]		0.8009 (0.0553) [0.0560]
$\widehat{\delta}$		0.4999 (0.0469) [0.0487]		0.4992 (0.0459) [0.0478]		0.4997 (0.0459) [0.0474]		0.4994 (0.0455) [0.0470]

Note: Observations $n = 361$, $\beta_1 = \beta_2 = \gamma_1 = 1$, and $\gamma_2 = 0.8$. Estimated standard error based on an asymptotic variance–covariance matrix is in parentheses; and empirical standard deviation is in brackets.

1. Generate bivariate normal random variables $(v_{i,n}, \varepsilon_{i,n})$ from i.i.d. $N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ as disturbances in the outcome equation and the spatial weight equation.
2. Construct the spatial weight matrix as the Hadamard product $W_n = W_n^d \circ W_n^e$, i.e., $w_{ij,n} = w_{ij,n}^d w_{ij,n}^e$, where W_n^d is a predetermined matrix based on geographic distance: $w_{ij,n}^d = 1$ if the two locations are neighbors and otherwise 0; W_n^e is a matrix based on economic similarity: $w_{ij,n}^e = 1/|z_{i,n} - z_{j,n}|$ if $i \neq j$ and $w_{ii,n}^e = 0$, where elements of Z_n are generated by $z_{i,n} = 1 + 0.8x_{i2,n} + \varepsilon_{i,n}$.
3. Row-normalize W_n.

For the predetermined W_n^d, we use four examples. First, the US states spatial weight matrix $WS(49 \times 49)$, based on the contiguity of the 48 contiguous states and D.C.; second, the Toledo spatial weight matrix $WO(98 \times 98)$, based on the 5 nearest neighbors of 98 census tracts in Toledo, Ohio; third, the Iowa "Adjacency" spatial weight matrix $WA(361 \times 361)$, based on the adjacency of 361 school districts in Iowa in 2009; and lastly, the Iowa "County" spatial weight matrix $WC(361 \times 361)$, based on whether the school districts are in the same county in Iowa in 2009.

In the simulation, we compare four different estimation methods: conventional IV, 2SIV, conventional MLE of SAR, and the MLE in Section 3.2. The conventional method refers to the case of treating W_n as exogenous. We refer to these four methods as IV, 2SIV, SAR, and MLE in tables. Here 2SIV and MLE correctly treat W_n as endogenous, but the conventional IV and SAR methods only estimate the outcome equation (Z_n equation is not estimated) since they treat W_n as exogenous. Of particular interest, we want to see how large the bias is for the two conventional estimation methods when W_n is endogenous. To generate different degrees of endogeneity, we choose correlation coefficients $\rho = 0.2$, 0.5, and 0.8. We also let the spatial correlation to be $\lambda = 0.2$ and 0.4 to investigate how the spatial correlation parameter affects estimates. 1000 replications are carried out for each setting.[11]

6.2. Monte Carlo results

Tables 1–6 report the empirical mean of each estimator, the empirical mean of its estimated standard error based on the

[11] We try the DGP of some other values of β and γ. The results are similar.

Table 5
Estimates from spatial weight matrices with strong endogeneity (small sample).

$\rho = 0.8$	WS $(n = 49)$				WO $(n = 98)$			
$\lambda = 0.2$	IV	2SIV	SAR	MLE	IV	2SIV	SAR	MLE
$\widehat{\lambda}$	−0.0469 (0.2525) [0.2206]	0.2002 (0.1309) [0.1377]	0.0152 (0.1400) [0.1289]	0.1921 (0.0830) [0.0881]	−0.1427 (0.2193) [0.1797]	0.1913 (0.1021) [0.1066]	−0.0345 (0.1191) [0.1066]	0.1917 (0.0667) [0.0690]
$\widehat{\beta}_1$	1.3325 (0.3728) [0.3428]	0.9968 (0.2332) [0.2340]	1.2506 (0.2352) [0.2416]	1.0092 (0.1814) [0.1807]	1.4673 (0.3133) [0.2915]	1.0140 (0.1729) [0.1779]	1.3194 (0.1880) [0.1991]	1.0133 (0.1353) [0.1394]
$\widehat{\beta}_2$	0.9436 (0.1594) [0.1668]	0.9934 (0.1581) [0.1607]	0.9642 (0.1484) [0.1564]	0.9955 (0.1515) [0.1583]	0.8974 (0.1176) [0.1302]	0.9913 (0.1145) [0.1179]	0.9319 (0.1083) [0.1169]	0.9931 (0.1104) [0.1166]
$\widehat{\gamma}_1$		1.0015 (0.1427) [0.1418]		1.0028 (0.1398) [0.1450]		1.0024 (0.1015) [0.1028]		1.0024 (0.1005) [0.1029]
$\widehat{\gamma}_2$		0.7982 (0.1542) [0.1582]		0.7983 (0.1511) [0.1588]		0.7999 (0.1104) [0.1149]		0.7999 (0.1093) [0.1149]
$\widehat{\delta}$		0.8047 (0.0960) [0.1025]		0.8011 (0.0876) [0.0954]		0.7985 (0.0676) [0.0705]		0.7978 (0.0628) [0.0664]
$\lambda = 0.4$	IV	2SIV	SAR	MLE	IV	2SIV	SAR	MLE
$\widehat{\lambda}$	0.1543 (0.2530) [0.2274]	0.3983 (0.1247) [0.1323]	0.2281 (0.1320) [0.1278]	0.3885 (0.0765) [0.0820]	0.0636 (0.2219) [0.1862]	0.3905 (0.0965) [0.1017]	0.1898 (0.1121) [0.1066]	0.3900 (0.0611) [0.0640]
$\widehat{\beta}_1$	1.4404 (0.4826) [0.4471]	0.9998 (0.2726) [0.2793]	1.3104 (0.2777) [0.2902]	1.0181 (0.1990) [0.2013]	1.6052 (0.4105) [0.3791]	1.0194 (0.2017) [0.2104]	1.3778 (0.2233) [0.2393]	1.0200 (0.1487) [0.1547]
$\widehat{\beta}_2$	0.9603 (0.1606) [0.1654]	0.9931 (0.1570) [0.1596]	0.9784 (0.1499) [0.1562]	0.9959 (0.1509) [0.1579]	0.9173 (0.1178) [0.1285]	0.9915 (0.1130) [0.1167]	0.9501 (0.1089) [0.1162]	0.9932 (0.1097) [0.1157]
$\widehat{\gamma}_1$		1.0015 (0.0933) [0.1418]		1.0011 (0.1398) [0.1427]		1.0024 (0.1015) [0.1028]		1.0024 (0.1005) [0.1029]
$\widehat{\gamma}_2$		0.7982 (0.1427) [0.1582]		0.7987 (0.1510) [0.1587]		0.7999 (0.1104) [0.1149]		0.7998 (0.1093) [0.1150]
$\widehat{\delta}$		0.8044 (0.1542) [0.0989]		0.8008 (0.0863) [0.0934]		0.7988 (0.0652) [0.0680]		0.7980 (0.0616) [0.0651]

Note: Observations $n = 49$ or 98, $\beta_1 = \beta_2 = \gamma_1 = 1$, and $\gamma_2 = 0.8$. Estimated standard error based on an asymptotic variance–covariance matrix is in parentheses; and empirical standard deviation is in brackets.

corresponding asymptotic variance–covariance matrix (in parentheses), and the empirical standard deviation of the estimator (in brackets) based on 1000 replications using WS, WO, WA, or WC as the predetermined spatial weight matrices. In each table, the upper panel shows the results for $\lambda = 0.2$ and the lower panel for $\lambda = 0.4$. To see how the different estimation methods behave under different degrees of endogeneity, we conduct three sets of simulations: results for weak endogeneity ($\rho = 0.2$) are in Tables 1 and 2, medium endogeneity ($\rho = 0.5$) in Tables 3 and 4, and strong endogeneity ($\rho = 0.8$) in Tables 5 and 6.

The simulation results are summarized as follows.

1. For the biases of parameter estimators, our 2SIV and MLE estimators have very small biases in all cases. For conventional IV and SAR estimators, the higher the degree of endogeneity is, i.e., the larger the correlation coefficient ρ is, the larger the bias of estimator is. The biases for estimators of the spatial correlation $\widehat{\lambda}$ are, in general, much higher than those for β. $\widehat{\lambda}$ from IV and SAR suffers severe downward bias when $\rho = 0.5$ or 0.8, in some cases with bias exceeding 100%. The conventional IV performs much worse than the conventional SAR.
2. For the variances of parameter estimators, we provide both the empirical standard deviation based on 1000 replications and the mean of estimated standard error based on the asymptotic variance–covariance matrix. From Tables 1–6, we can see that these two values are very close in all cases. Comparing variances of estimators from different estimation methods, we can see that IV is close to 2SIV and SAR is close to MLE. It seems that estimators based on the likelihood estimation method have smaller variances than those based on the IV methods.
3. The biases of IV and SAR estimators vary with the spatial correlation λ. When true $\lambda = 0.2$, $\widehat{\lambda}$ from the IV and SAR has large biases relative to its true value than when $\lambda = 0.4$. It seems that the conventional methods produce even more severe bias in the situation of weak spatial correlation.
4. Comparing Table 1 to Table 2, Table 3 to Table 4, and Table 5 to Table 6, as the sample size increases while the number of neighbors for each agent grows at a slower rate, the bias and standard error of estimators decrease.

Table 6
Estimates from spatial weight matrices with strong endogeneity (large sample).

$\rho = 0.8$	WA ($n = 361$)				WC ($n = 361$)			
$\lambda = 0.2$	IV	2SIV	SAR	MLE	IV	2SIV	SAR	MLE
$\widehat{\lambda}$	−0.0712 (0.0944) [0.0765]	0.2009 (0.0453) [0.0435]	0.0047 (0.0576) [0.0528]	0.1988 (0.0325) [0.0316]	0.0386 (0.0792) [0.0697]	0.2005 (0.0383) [0.0362]	0.1039 (0.0435) [0.0443]	0.1985 (0.0257) [0.0254]
$\widehat{\beta}_1$	1.3579 (0.1340) [0.1231]	1.0005 (0.0797) [0.0752]	1.2580 (0.0914) [0.0931]	1.0033 (0.0677) [0.0641]	1.2153 (0.1174) [0.1111]	1.0011 (0.0733) [0.0692]	1.1288 (0.0778) [0.0810]	1.0037 (0.0626) [0.0603]
$\widehat{\beta}_2$	0.9509 (0.0562) [0.0589]	1.0011 (0.0561) [0.0557]	0.9657 (0.0546) [0.0555]	1.0010 (0.0555) [0.0553]	0.9732 (0.0571) [0.0579]	1.0009 (0.0558) [0.0556]	0.9851 (0.0553) [0.0554]	1.0009 (0.0554) [0.0553]
$\widehat{\gamma}_1$		0.9991 (0.0528) [0.0532]		0.9991 (0.0526) [0.0532]		0.9991 (0.0528) [0.0532]		0.9991 (0.0526) [0.0532]
$\widehat{\gamma}_2$		0.8014 (0.0555) [0.0562]		0.8014 (0.0553) [0.0562]		0.8014 (0.0555) [0.0562]		0.8014 (0.0553) [0.0562]
$\widehat{\delta}$		0.7993 (0.0336) [0.0339]		0.7986 (0.0325) [0.0327]		0.7992 (0.0322) [0.0322]		0.7986 (0.0317) [0.0319]
$\lambda = 0.4$	IV	2SIV	SAR	MLE	IV	2SIV	SAR	MLE
$\widehat{\lambda}$	0.1408 (0.0929) [0.0762]	0.4006 (0.0421) [0.0404]	0.2344 (0.0539) [0.0517]	0.3983 (0.0296) [0.0290]	0.2576 (0.0720) [0.0631]	0.4002 (0.0328) [0.0310]	0.3323 (0.0370) [0.0376]	0.3985 (0.0217) [0.0217]
$\widehat{\beta}_1$	1.4552 (0.1706) [0.1539]	1.0006 (0.0907) [0.0856]	1.2912 (0.1075) [0.1104]	1.0046 (0.0738) [0.0700]	1.2516 (0.1375) [0.1278]	1.0014 (0.0782) [0.0738]	1.1205 (0.0837) [0.0858]	1.0035 (0.0649) [0.0650]
$\widehat{\beta}_2$	0.9720 (0.0565) [0.0576]	1.0010 (0.0556) [0.0553]	0.9832 (0.0551) [0.0553]	1.0010 (0.0553) [0.0551]	0.9948 (0.0572) [0.0561]	1.0008 (0.0555) [0.0552]	0.9986 (0.0556) [0.0551]	1.0007 (0.0552) [0.0573]
$\widehat{\gamma}_1$		0.9991 (0.0528) [0.0532]		0.9991 (0.0526) [0.0532]		0.9991 (0.0528) [0.0532]		0.9977 (0.0527) [0.0610]
$\widehat{\gamma}_2$		0.8014 (0.0555) [0.0562]		0.8014 (0.0553) [0.0562]		0.8014 (0.0555) [0.0562]		0.8011 (0.0553) [0.0591]
$\widehat{\delta}$		0.7992 (0.0326) [0.0329]		0.7986 (0.0320) [0.0321]		0.7991 (0.0318) [0.0318]		0.7985 (0.0315) [0.0329]

Note: Observations $n = 361$, $\beta_1 = \beta_2 = \gamma_1 = 1$, and $\gamma_2 = 0.8$. Estimated standard error based on an asymptotic variance–covariance matrix is in parentheses; and empirical standard deviation is in brackets.

7. Conclusion

In this paper, we consider the specification and estimation of a cross-sectional SAR model with an endogenous spatial weight matrix. First, we specify two sets of equations: one is for the SAR outcome, and the other is for entries of the spatial weight matrix. The source of endogeneity is the correlation between the disturbances in the SAR outcome equation and the errors in the spatial weight entry equation. Second, under the conditional moment assumptions on disturbances, we propose three estimation methods: 2SIV, QMLE, and GMM. We consider two types of spatial weight matrices: one is sparse and another one has its entries decreasing sufficiently fast as the physical distance increases. By employing the theory of asymptotic inference under near-epoch dependence, we prove the consistency and asymptotic normality of these three estimators. In generalized 2SIV, we also provide the optimal choice for IV matrices.

To examine the behavior of our proposed estimators in finite samples, we conduct a Monte Carlo simulation study. The simulation results indicate that the commonly used estimates under exogenous weight matrix suffer serious downward bias when the true weight matrix is endogenous. On the other hand, our estimates have good finite sample properties. As the sample size increases and the number of neighbors grows more slowly, our estimates quickly converge to true parameters.

This paper focuses on estimating a cross-sectional SAR model with a specified source of endogeneity for the spatial weight matrix. In future research, we may extend our cross-sectional model to a spatial panel data setting where the spatial weight matrix varies over time due to changing economic conditions. Another issue that needs future research is to consider an endogenous spatial weight matrix purely constructed with economic distances. This could be a technical challenging issue as the near-epoch assumption may not be met. Thus alternative large sample theorems may need to be developed.

Appendix A. Expressions related to the statistics

A.1. First order derivatives and the expectation of the log quasi-likelihood function

The expectation of the log quasi-likelihood function in (3.4) is

$$\frac{1}{n}E(\ln L_n(\theta)) = -\ln(2\pi) - \frac{1}{2}\ln(\sigma_\xi^2) - \frac{1}{2}\ln|\Sigma_\varepsilon| + \frac{1}{n}E(\ln|S_n(\lambda)|) - \frac{1}{2}\mathrm{tr}(\Sigma_\varepsilon^{-1}\Sigma_{\varepsilon 0}) - \frac{1}{2n}\sum_{i=1}^n x_{2,in}'(\Gamma_0-\Gamma)\Sigma_\varepsilon^{-1}(\Gamma_0-\Gamma)'x_{2,in} - \frac{1}{2n}\frac{\sigma_{\xi 0}^2}{\sigma_\xi^2}E[\mathrm{tr}(S_n^{-1\prime}S_n(\lambda)'S_n(\lambda)S_n^{-1})] - \frac{1}{2\sigma_\xi^2}((\lambda_0-\lambda),(\beta_0-\beta)',((\Gamma-\Gamma_0)\delta)',(\delta_0-\delta)')H_{1n}((\lambda_0-\lambda),(\beta_0-\beta)',((\Gamma-\Gamma_0)\delta)',(\delta_0-\delta)')',$$

where $H_{1n} = \frac{1}{n}E[(G_n(X_{1n}\beta_0+\varepsilon_n\delta_0),X_{1n},X_{2n},\varepsilon_n)'(G_n(X_{1n}\beta_0+\varepsilon_n\delta_0),X_{1n},X_{2n},\varepsilon_n)]$.

The first order derivatives are

$$\frac{\partial \ln L_n(\theta)}{\partial\lambda} = \frac{1}{\sigma_\xi^2}(W_nY_n)'\xi_n(\theta) - \mathrm{tr}[W_nS_n^{-1}(\lambda)];$$

$$\frac{\partial \ln L_n(\theta)}{\partial\beta} = \frac{1}{\sigma_\xi^2}X_{1n}'\xi_n(\theta);$$

$$\frac{\partial \ln L_n(\theta)}{\partial\,\mathrm{vec}(\Gamma)} = (\Sigma_\varepsilon^{-1}\otimes X_{2n}')\mathrm{vec}(Z_n - X_{2n}\Gamma) - \frac{1}{\sigma_\xi^2}\delta\otimes(X_{2n}'\xi_n(\theta));$$

$$\frac{\partial \ln L_n(\theta)}{\partial\sigma_\xi^2} = -\frac{n}{2\sigma_\xi^2} + \frac{1}{2\sigma_\xi^4}\xi_n(\theta)'\xi_n(\theta);$$

$$\frac{\partial \ln L_n(\theta)}{\partial\delta} = \frac{1}{\sigma_\xi^2}\varepsilon_n(\theta)'\xi_n(\theta);$$

$$\frac{\partial \ln L_n(\theta)}{\partial\alpha} = -\frac{n}{2}\frac{\partial\ln|\Sigma_\varepsilon|}{\partial\alpha} - \frac{1}{2}\frac{\partial}{\partial\alpha}\mathrm{tr}[\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)],$$

where $\xi_n(\theta) = S_n(\lambda)Y_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta$ and $\varepsilon_n(\Gamma) = Z_n - X_{2n}\Gamma$. As α is a J-dimensional column vector of distinct elements in Σ_ε, the J-dimensional vector $\frac{\partial\ln|\Sigma_\varepsilon|}{\partial\alpha}$ has the j th element $\mathrm{tr}(\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_j})$ and $\frac{\partial}{\partial\alpha}\mathrm{tr}[\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)]$ has its jth element $-\mathrm{tr}\left(\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_j}\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)\right)$ for $j = 1,\ldots,J$.

A.2. Second order derivatives and the variance–covariance matrix

The second order derivatives are

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\lambda} = -\mathrm{tr}[W_nS_n^{-1}(\lambda)]^2 - \frac{1}{\sigma_\xi^2}(W_nY_n)'W_nY_n;$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\sigma_\xi^2} = -\frac{1}{\sigma_\xi^4}(W_nY_n)'\xi_n(\theta);$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\beta} = -\frac{1}{\sigma_\xi^2}X_{1n}'W_nY_n;$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\mathrm{vec}(\Gamma)} = \frac{1}{\sigma_\xi^2}\delta\otimes(X_{2n}'W_nY_n); \qquad \frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\alpha} = 0;$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\delta} = -\frac{1}{\sigma_\xi^2}\varepsilon_n(\Gamma)'(W_nY_n);$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\beta'} = -\frac{1}{\sigma_\xi^2}X_{1n}'X_{1n};$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\,\mathrm{vec}(\Gamma)'} = \frac{1}{\sigma_\xi^2}\delta\otimes(X_{2n}'X_{1n});$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\alpha'} = 0; \qquad \frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\delta'} = -\frac{1}{\sigma_\xi^2}X_{1n}'\varepsilon_n(\Gamma);$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\sigma_\xi^2} = -\frac{1}{\sigma_\xi^4}X_{1n}'\xi_n(\theta);$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\,\mathrm{vec}(\Gamma)\partial\mathrm{vec}(\Gamma)'} = -\Sigma_\varepsilon^{-1}\otimes(X_{2n}'X_{2n}) - \frac{1}{\sigma_\xi^2}\delta\delta'\otimes(X_{2n}'X_{2n});$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\,\mathrm{vec}(\Gamma)\partial\sigma_\xi^2} = \frac{1}{\sigma_\xi^4}\delta\otimes(X_{2n}'\xi_n(\theta));$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\mathrm{vec}(\Gamma)\partial\alpha'} = [I_{p_2}\otimes(X_{2n}'\varepsilon_n(\Gamma))]\frac{\partial\,\mathrm{vec}(\Sigma_\varepsilon^{-1})}{\partial\alpha'};$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\delta\partial\,\mathrm{vec}(\Gamma)'} = -\frac{1}{\sigma_\xi^2}I_{p2}\otimes(X_{2n}'\xi_n(\theta)) + \frac{1}{\sigma_\xi^2}\delta\otimes(X_{2n}'\varepsilon_n(\Gamma));$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\sigma_\xi^2\partial\alpha} = 0;$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\sigma_\xi^2\partial\sigma_\xi^2} = \frac{n}{2\sigma_\xi^4} - \frac{1}{\sigma_\xi^6}\xi_n(\theta)'\xi_n(\theta);$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\sigma_\xi^2\partial\delta} = -\frac{1}{\sigma_\xi^4}\varepsilon_n(\theta)'\xi_n(\theta); \qquad \frac{\partial^2 \ln L_n(\theta)}{\partial\alpha\partial\delta'} = 0;$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\delta\partial\delta'} = -\frac{1}{\sigma_\xi^2}\varepsilon_n(\theta)'\varepsilon_n(\theta);$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\alpha\partial\alpha'} = -\frac{n}{2}\frac{\partial^2\ln|\Sigma_\varepsilon|}{\partial\alpha\partial\alpha'} - \frac{1}{2}\frac{\partial^2}{\partial\alpha\partial\alpha'}\mathrm{tr}[\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)]$$

where $\frac{\partial^2\ln|\Sigma_\varepsilon|}{\partial\alpha\partial\alpha'}$ is a $J\times J$ matrix with the (j,k)th element $\frac{\partial^2\ln|\Sigma_\varepsilon|}{\partial\alpha_j\partial\alpha_k} = -\mathrm{tr}(\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_k}\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_j})$ and the (j,k)th element of $\frac{\partial^2}{\partial\alpha\partial\alpha'}\mathrm{tr}[\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)]$ is

$$\frac{\partial^2}{\partial\alpha_j\partial\alpha_k}\mathrm{tr}[\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)] = \mathrm{tr}\left(\Sigma_\varepsilon^{-1}\left(\frac{\partial\Sigma_\varepsilon}{\partial\alpha_k}\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_j} + \frac{\partial\Sigma_\varepsilon}{\partial\alpha_j}\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_k}\right)\times\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)\right)$$

for $j,k = 1,\ldots,J$. Therefore,

$$E\left(\frac{\partial^2\ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right) = \frac{1}{\sigma_{\xi 0}^2}\begin{pmatrix} I_{\lambda\lambda} & I_{\lambda\beta}' & I_{\lambda\Gamma}' & -E[\mathrm{tr}(G_n)] & 0 & I_{\lambda\delta}' \\ * & -X_{1n}'X_{1n} & \delta_0'\otimes(X_{1n}'X_{2n}) & 0 & 0 & 0 \\ * & * & I_{\Gamma\Gamma} & 0 & 0 & 0 \\ * & 0 & 0 & -\frac{n}{2\sigma_{\xi 0}^2} & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{\alpha\alpha} & 0 \\ * & 0 & 0 & 0 & 0 & -n\Sigma_{\varepsilon 0}\end{pmatrix},$$

with

$$I_{\lambda\lambda} = -\sigma_{\xi 0}^2\mathrm{tr}[E(G_n^2 + G_nG_n')] - E[(X_{1n}\beta_0+\varepsilon_n\delta_0)'G_n'(X_{1n}\beta_0+\varepsilon_n\delta_0)];$$

$$I_{\lambda\beta} = -X_{1n}'E(G_nX_{1n}\beta_0 + G_n\varepsilon_n\delta_0);$$

$$I_{\lambda\Gamma} = \delta_0\otimes[X_{2n}'E(G_nX_{1n}\beta_0 + G_n\varepsilon_n\delta_0)];$$

$$I_{\lambda\delta} = -E[\varepsilon_n'G_n(X_{1n}\beta_0+\varepsilon_n\delta_0)];$$

$$I_{\Gamma\Gamma} = -(\sigma_{\xi 0}^2\Sigma_{\varepsilon 0}^{-1} + \delta_0\delta_0')\otimes(X_{2n}'X_{2n});$$

$$(I_{\alpha\alpha})_{kj} = -\frac{n\sigma_{\xi 0}^2}{2}\mathrm{tr}\left(\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_k}\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_j}\right)$$

for $j, k = 1, \ldots, J$.

And

$$E\left(\frac{\partial \ln L_n(\theta_0)}{\partial \theta}\frac{\partial \ln L_n(\theta_0)}{\partial \theta'}\right) = -E\left(\frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'}\right) + \Omega_{\theta_0}^{ML},$$

where $\Omega_{\theta_0}^{ML}$ is given in Box I with

$$R_{\lambda\lambda} = \sum_{i=1}^n E[2\xi_{i,n}^3 G_{in}(X_{1n}\beta_0 + \varepsilon_n\delta_0)G_{ii,n} + G_{ii,n}^2(\xi_{i,n}^4 - 3\sigma_{\xi 0}^4)];$$

$$R_{\lambda\beta} = \sum_{i=1}^n E(\xi_{i,n}^3 G_{ii,n})x'_{1,in};$$

$$R_{\lambda\Gamma} = -\delta_0\sum_{i=1}^n E(\xi_{i,n}^3 G_{ii,n})x'_{2,in};$$

$$R_{\Gamma\alpha} = \frac{\sigma_{\xi 0}^4}{2}\left[l'_n \otimes E\left(\varepsilon'_{i,n}\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_j}\Sigma_{\varepsilon 0}^{-1}\varepsilon_{i,n} \times \Sigma_{\varepsilon 0}^{-1}\varepsilon_{i,n}\right)\otimes I_{k_2}\right]\mathrm{vec}(X'_{2n});$$

$$\begin{aligned}(R_{\alpha\alpha})_{kj} = \frac{n\sigma_{\xi 0}^4}{4}\Bigg[&E\left(\varepsilon'_{i,n}\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_j}\Sigma_{\varepsilon 0}^{-1}\varepsilon_{i,n}\varepsilon'_{i,n}\,\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_k}\Sigma_{\varepsilon 0}^{-1}\varepsilon_{i,n}\right) \\ &- \mathrm{tr}\left(\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_j}\right)\mathrm{tr}\left(\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_k}\right) \\ &- 2\mathrm{tr}\left(\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_k}\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_j}\right)\Bigg];\end{aligned}$$

$$\begin{aligned}R_{\lambda\xi} = \frac{1}{2\sigma_{\xi 0}^2}\sum_{i=1}^n &\{E[\xi_{i,n}^3 l'_n G_n(X_{1n}\beta_0 + \varepsilon_n\delta_0)] \\ &+ E[(\xi_{i,n}^4 - 3\sigma_{\xi 0}^4)G_{ii,n}]\}.\end{aligned}$$

Appendix B. Some basic properties of NED of random fields

In the following proofs, we will adopt asymptotic inference under near-epoch dependence and let $\varsigma_n = (\varepsilon_n, \xi_n)$ be the basis for NED processes. The following claims are some basic results. The first Claim B.1 is due to the topological structure in Assumption 1. The other claims are some basic properties for NED processes.

Claim B.1. *For any distance ρ, there are at most $c_5\rho^{d_0}$ points in $B_i(\rho)$ and at most $c_4\rho^{d_0-1}$ points in the space $B_i(\rho+1)\setminus B_i(\rho)$, where c_4 and c_5 are positive constants.*

Claim B.1 is directly from Jenish and Prucha (2012).[12]

Claim B.2. *For any random field $T = \{T_{i,n},\ i \in D_n,\ n \geq 1\}$ with $\|T_{i,n}\|_p < \infty$, $\|T_{i,n} - E(T_{i,n}|\mathcal{F}_{i,n}(s))\|_p \leq 2\|T_{i,n}\|_p$ with $p \geq 1$.*

This result follows from the Minkowski and conditional Jensen's inequalities: $\|T_{i,n} - E(T_{i,n}|\mathcal{F}_{i,n}(s))\|_p \leq \|T_{i,n}\|_p + \|E(T_{i,n}|\mathcal{F}_{i,n}(s))\|_p \leq 2\|T_{i,n}\|_p$.

Claim B.3. *If $\|t_{1i,n} - E(t_{1i,n}|\mathcal{F}_{i,n}(s))\|_4 \leq C_1\varphi_1(s)$ and $\|t_{2i,n} - E(t_{2i,n}|\mathcal{F}_{i,n}(s))\|_4 \leq C_2\varphi_2(s)$, with $\max(\|t_{1i,n}\|_4, \|t_{2i,n}\|_4) \leq C$, then $\|t_{1i,n}t_{2i,n} - E(t_{1i,n}t_{2i,n}|\mathcal{F}_{i,n}(s))\|_2 \leq C(C_1+C_2)\varphi(s)$, where $\varphi(s) = \max(\varphi_1(s), \varphi_2(s))$.*

Proof of Claim B.3. For the product of $t_{1i,n}t_{2i,n}$,

$$\begin{aligned}&\|t_{1i,n}t_{2i,n} - E(t_{1i,n}t_{2i,n}|\mathcal{F}_{i,n}(s))\|_2 \\ &\quad\leq \|t_{1i,n}t_{2i,n} - E(t_{1i,n}|\mathcal{F}_{i,n}(s))E(t_{2i,n}|\mathcal{F}_{i,n}(s))\|_2 \\ &\quad\leq \|t_{2i,n}[t_{1i,n} - E(t_{1i,n}|\mathcal{F}_{i,n}(s))]\|_2 \\ &\qquad+ \|E(t_{1i,n}|\mathcal{F}_{i,n}(s))[t_{2i,n} - E(t_{2i,n}|\mathcal{F}_{i,n}(s))]\|_2 \\ &\quad\leq \|t_{2i,n}\|_4\cdot\|t_{1i,n} - E(t_{1i,n}|\mathcal{F}_{i,n}(s))\|_4 \\ &\qquad+ \|t_{1i,n}\|_4\cdot\|t_{2i,n} - E(t_{2i,n}|\mathcal{F}_{i,n}(s))\|_4 \\ &\quad\leq C(C_1\varphi_1(s) + C_2\varphi_2(s)) \leq C(C_1+C_2)\max(\varphi_1(s), \varphi_2(s)).\end{aligned}$$

The third inequality follows from Hölder's inequality. ■

From Jenish and Prucha (2012), we have the following two Claims for LLN and CLT under NED.

Claim B.4. *Under Assumption 1, if the random field $\{T_{i,n},\ i \in D_n,\ n \geq 1\}$ is L_1-NED, the base $\{\varsigma_{i,n}\}$'s are i.i.d., and $\{T_{i,n}\}$'s are uniformly L_p bounded for some $p > 1$, then $\frac{1}{n}\sum_{i=1}^n(T_{i,n} - ET_{i,n}) \xrightarrow{L_1} 0$.*

Claim B.5. *Let $\{T_{i,n},\ l(i) \in D_n,\ n \geq 1\}$ be a random field that is L_2-NED on an i.i.d. random field ς. If Assumption 1 and the following conditions are met:*

(1) *$\{T_{i,n},\ i \in D_n,\ n \geq 1\}$ is uniformly $L_{2+\delta}$-bounded for some $\delta > 0$,*
(2) *$\inf_n \frac{1}{n}\sigma_n^2 > 0$ where $\sigma_n^2 = \mathrm{Var}(\sum_{i=1}^n T_{i,n})$,*
(3) *NED coefficients satisfy $\sum_{r=1}^\infty r^{d_0-1}\varphi(r) < \infty$,*
(4) *NED scaling factors satisfy $\sup_{n,i\in D} d_{i,n} < \infty$,*

then $\sigma_n^{-1}\sum_{i=1}^n(T_{i,n} - ET_{i,n}) \xrightarrow{d} N(0,1)$.

Appendix C. Proofs of NED properties for relevant statistics

C.1. NED properties in Case 1 under Assumption (4.1)

Claim C.1.1. *Under Assumptions 1, (3.1), and (4.1), $\sup_n\|W_n\|_1 < \infty$.*[13]

Proof of Claim C.1.1. For any i, divide the whole space D into subsets $B_i(\rho+1)\setminus B_i(\rho)$, $\rho = 1, 2, \ldots,$ and $B_i(1)$. Under Assumption (4.1), $0 \leq w_{ij,n} \leq c_1\rho_{ij}^{-c_3 d_0}$. Then $w_{ji,n} \leq c_1\rho^{-c_3 d_0}$ for any $j \in B_i(\rho+1)\setminus B_i(\rho)$ with $\rho \geq 1$. There are at most $c_4\rho^{d_0-1}$ points in $B_i(\rho+1)\setminus B_i(\rho)$. Therefore, $\sum_{j\in B_i(\rho+1)\setminus B_i(\rho)} w_{ji,n} \leq c_4 c_1\rho^{(1-c_3)d_0-1}$. For the special case of $B_i(1)$, as $w_{ii,n} = 0$, it must be $\rho_{ij} = 1$ from Assumption 1 and hence, $w_{ji,n} \leq c_1$. Since $D_n \subset D = B_i(1)\bigcup\left(\cup_{\rho=1}^\infty B_i(\rho+1)\setminus B_i(\rho)\right)$, we have $\sum_{j=1}^n w_{ji,n} = \sum_{\rho=0}^\infty\sum_{j\in B_i(\rho+1)\setminus B_i(\rho)} w_{ji,n} \leq c_4 c_1\left(1 + \sum_{\rho=1}^\infty \rho^{(1-c_3)d_0-1}\right) < \infty$ when $c_3 > 1$. ■

Claim C.1.2. *Under Assumptions 1, (3.1), and (4.1), for any n and positive integer q, $\|W_n^q\|_1 \leq (q-1)c_u K c_w^{q-1} + c_u c_w^{q-1} \leq q c_u K c_w^{q-1}$, where $c_u = \sup_n\|W_n\|_1$ and $c_w = \sup_n\|W_n\|_\infty$.*

Proof of Claim C.1.2. Denote an index set V_n with $c_w \leq \sum_{j=1}^n w_{ji,n} < c_u$ if $i \in V_n$ and $\sum_{j=1}^n w_{ji,n} < c_w$ if $i \notin V_n$. Then Assumption (4.1) constrains that $|V_n| \leq K$ for any n. Consider the kth column sum of W_n^q, i.e., $e'_n W_n^q e_{k,n}$, where $e_{k,n} = (0, \ldots, 0, 1, 0, \ldots, 0)'$ is the

---

12 These two results are special cases of those in Jenish and Prucha (2012) where the base random field can be spatial mixing processes. Here we have the base being i.i.d. variables for simplicity, which is sufficient for our model.

13 For this claim, it is sufficient to have $c_3 > 1$ in Assumption (4.1) instead of the larger c_3.

$$\Omega_{\theta_0}^{ML} = \frac{1}{\sigma_{\xi 0}^4}\begin{pmatrix} R_{\lambda\lambda} & R_{\lambda\beta} & R_{\lambda\Gamma} & R_{\lambda\xi} & 0 & \sum_{i=1}^n E(\xi_{i,n}^3\varepsilon_{i,n}G_{ii,n}) \\ * & 0 & 0 & \frac{1}{2\sigma_{\xi 0}^2}\sum_{i=1}^n E(\xi_{i,n}^3)x_{1,in}' & 0 & 0 \\ * & * & 0 & -\frac{\delta_0}{2\sigma_{\xi 0}^2}\sum_{i=1}^n E(\xi_{i,n}^3)x_{2,in}' & R_{\Gamma\alpha} & 0 \\ * & * & * & \frac{n}{4\sigma_{\xi 0}^4}(\mu_{\xi 4} - 3\sigma_{\xi 0}^4) & 0 & \frac{1}{2\sigma_{\xi 0}^2}\sum_{i=1}^n E(\xi_{i,n}^3\varepsilon_{i,n}) \\ * & * & * & * & (R_{\alpha\alpha})_{kj} & 0 \\ * & * & * & * & * & 0 \end{pmatrix}$$

Box I.

unit column vector with one in its kth entry and zeros in its other entries and $e_n = (1, \ldots, 1)' = \sum_{k=1}^n e_{k,n}$. As $I_n = \sum_{i=1}^n e_{i,n}e_{i,n}'$,

$$\begin{aligned} e_n'W_n^q e_{k,n} &= \sum_{i=1}^n e_n'W_n e_{i,n}e_{i,n}'W_n^{q-1}e_{k,n} \\ &= \sum_{i\in V_n} e_n'W_n e_{i,n}e_{i,n}'W_n^{q-1}e_{k,n} + \sum_{i\notin V_n} e_n'W_n e_{i,n}e_{i,n}'W_n^{q-1}e_{k,n} \\ &\le K\left(\max_{i\in V_n} e_n'W_n e_{i,n}\right)\left(\max_{i\in V_n} e_{i,n}'W_n^{q-1}e_{k,n}\right) \\ &\quad + \left(\max_{i\notin V_n} e_n'W_n e_{i,n}\right)\sum_{i\notin V_n} e_{i,n}'W_n^{q-1}e_{k,n} \\ &\le Kc_u\|W_n^{q-1}\|_\infty + c_w\|W_n^{q-1}\|_1 \\ &\le Kc_u c_w^{q-1} + c_w\|W_n^{q-1}\|_1. \end{aligned}$$

As this inequality holds for any $k = 1, \ldots, n$, we have $\|W_n^q\|_1 \le c_uKc_w^{q-1} + c_w\|W_n^{q-1}\|_1$. By deduction, we have $\|W_n^q\|_1 \le (q-1)c_uKc_w^{q-1} + c_uc_w^{q-1} \le qc_uKc_w^{q-1}$. ■

Claim C.1.3. *Under Assumptions 1, (3.1), (3.2), and (4.1), $\sup_{\lambda\in\Lambda}\|G_n(\lambda)\|_\infty < \infty$ and $\sup_{\lambda\in\Lambda}\|G_n(\lambda)\|_1 < \infty$.*

Proof of Claim C.1.3. As $G_n(\lambda) = \sum_{l=0}^\infty \lambda^l W_n^{l+1}$ and $\|W_n^{l+1}\|_\infty \le \|W_n\|_\infty^{l+1}$, we have

$$\sup_{\lambda\in\Lambda}\|G_n(\lambda)\|_\infty \le \sum_{l=0}^\infty \sup_{\lambda\in\Lambda}|\lambda|^l\|W_n^{l+1}\|_\infty \le c_w\sum_{l=0}^\infty \sup_{\lambda\in\Lambda}|\lambda c_w|^l < \infty.$$

From Claim C.1.2, $\|W_n^{l+1}\|_1 \le c_uK(l+1)c_w^l$, and hence,

$$\begin{aligned} \sup_{\lambda\in\Lambda}\|G_n(\lambda)\|_1 &\le \sum_{l=0}^\infty \sup_{\lambda\in\Lambda}|\lambda|^l\|W_n^{l+1}\|_1 \\ &\le c_uK\sum_{l=0}^\infty (l+1)\sup_{\lambda\in\Lambda}|\lambda c_w|^l < \infty. \quad ■ \end{aligned}$$

Claim C.1.4. *Suppose W is an $n\times n$ square matrix which can be decomposed into the sum of two $n\times n$ matrices such that $W = A+B$. Denote $|A|_{\max} = \max\{|a_{ij}| : i, j = 1, \ldots, n\}$. Then for any positive integer k and any $i, j = 1, \ldots, n$,*

$$(W^k - B^k)_{ij} \le |A|_{\max}\sum_{m=0}^{k-1}\|B\|_\infty^m \cdot \|W^{k-1-m}\|_1.$$

Proof of Claim C.1.4. By expansion, $W^k - B^k = \sum_{m=0}^{k-1} B^mA W^{k-1-m}$. Then $(W^k - B^k)_{ij} = \sum_{m=0}^{k-1} e_{i,n}'B^mAW^{k-1-m}e_{j,n}$. For any matrix M and vector e of dimension n, it is easy to see that $\|Me\|_\infty \le |M|_{\max}\|e\|_1$. Thus, for any integer $m = 0, \ldots, k-1$,

$$\begin{aligned} e_{i,n}'B^mAW^{k-1-m}e_{j,n} &\le \|B'^m e_{i,n}\|_1 \cdot \|AW^{k-1-m}e_{j,n}\|_\infty \\ &\le \|B^m\|_\infty \cdot |AW^{k-1-m}|_{\max} \\ &\le |A|_{\max}\cdot\|B^m\|_\infty\cdot\|W^{k-1-m}\|_1. \end{aligned}$$

Together, we have the result. ■

Claim C.1.5. *For any $\alpha > 0$ and $s \ge 2$, $\sum_{\rho=[s]}\rho^{-\alpha-1} < \frac{3^\alpha}{\alpha}s^{-\alpha}$, where $[s]$ denotes the largest integer less than or equal to s.*

Proof of Claim C.1.5. For any $\rho_0 \ge 2$,

$$\begin{aligned} \frac{\rho_0^{-\alpha}}{\alpha} &= \int_{\rho_0}^\infty x^{-\alpha-1}dx < \sum_{\rho=\rho_0}\rho^{-\alpha-1} \\ &< \int_{\rho_0-1}^\infty x^{-\alpha-1}dx = \frac{(\rho_0-1)^{-\alpha}}{\alpha} \le \frac{2^\alpha\rho_0^{-\alpha}}{\alpha}. \end{aligned}$$

The last inequality holds because $\rho_0 - 1 \ge \rho_0/2$ and hence $(\rho_0-1)^{-\alpha} \le 2^\alpha\rho_0^{-\alpha}$. Therefore, we can find a positive constant $1/\alpha < c_{\alpha 1} < 2^\alpha/\alpha$ such that $\sum_{\rho=\rho_0}\rho^{-\alpha-1} = c_{\alpha 1}\rho_0^{-\alpha}$. Now with $1 \le s/[s] < 1 + (s-[s])/[s] < 3/2$, there exists a constant $1 < c_{\alpha 2} < (3/2)^\alpha$ such that $c_{\alpha 2}s^{-\alpha} = [s]^{-\alpha}$. Together, we have $\sum_{\rho=[s]}\rho^{-\alpha-1} = c_{\alpha 1}c_{\alpha 2}s^{-\alpha} < (3^\alpha/\alpha)s^{-\alpha}$. ■

Claim C.1.6. *Let $t_{i,n}(m)$ be the ith element of the vector $W_n^m\varsigma_n^*a$, where $\varsigma_{i,n}^* = f_i(\varsigma_{i,n}, X_n)$ with $\varsigma_n = (\varepsilon_n, \xi_n)$ is a vector-valued function and a is any conformable vector of constants. Under Assumptions 1, (3.1), and (4.1), suppose $\sup_{i,n}\|\varsigma_{i,n}^*\|_p < \infty$, then $\sup_{i,n}\|t_{i,n}(m)\|_p \le m^{c_3d_0+2}c_w^mC_{ap}$ and $\sup_{i,n}\|t_{i,n}(m) - E(t_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \le C_{ap}c_w^m m^{3+c_3d_0}s^{(2-c_3)d_0}$ with C_{ap} being a finite constant.*

Proof of Claim C.1.6. First we show that $\|t_{i,n}(m) - E(t_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \le C_{ap}c_w^m m^{3+c_3d_0}s^{(2-c_3)d_0}$ for any i and n. Note that for any integer $1 \le k \le m$ and $i_k \notin B_i(s)$, we can show that

$$(W_n^k)_{ii_k} \le C_0m^{c_3d_0+2}c_w^k\rho_{ii_k}^{-c_3d_0} \quad \text{and} \quad \sum_{i_k\notin B_i(s)}(W_n^k)_{ii_k} \le C_1c_w^k m^{c_3d_0+2}s^{(1-c_3)d_0} \tag{C.1}$$

with C_0 and C_1 being positive constants, which do not depend on k. To show this, we construct two matrices A_n and B_n as follows: $a_{ij,n} = w_{ij,n}I(w_{ij,n} \le c_1(\rho_{ii_k}/m)^{-c_3d_0})$ and $b_{ij,n} = w_{ij,n}I(w_{ij,n} > c_1(\rho_{ii_k}/m)^{-c_3d_0})$, then $W_n = A_n + B_n$ and $a_{ij,n}b_{ij,n} = 0$. As $i_k \notin B_i(s)$, at least one of the items $w_{ii_1,n}, w_{i_1i_2,n}, \ldots, w_{i_{k-1}i_k,n}$, say $w_{i_{q-1}i_q,n}$, satisfies that $w_{i_{q-1}i_q,n} \le c_1(\rho_{ii_k}/k)^{-c_3d_0} \le c_1(\rho_{ii_k}/m)^{-c_3d_0}$, because there exist at least two neighboring nodes in the chain $i \to i_1 \to \cdots \to i_k$ such that their distance is at least ρ_{ii_k}/k. Hence, $(B_n^k)_{ii_k} = \sum_{i_1} \cdots \sum_{i_{(k-1)}} w_{ii_1,n} w_{i_1 i_2,n} \cdots w_{i_{(k-1)} i_k,n} I(\text{all } w\text{'s} > c_1(\rho_{ii_k}/m)^{-c_3 d_0}) = 0$, and we have

$$\begin{aligned}(W_n^k)_{ii_k} &= (W_n^k - B_n^k)_{ii_k} \le |A_n|_{\max} \sum_{q=0}^{k-1} \|B_n\|_\infty^q \|W_n^{k-1-q}\|_1 \\ &\le c_1 \left(\frac{\rho_{ii_k}}{m}\right)^{-c_3 d_0} \sum_{q=0}^{k-1} c_w^q (k-1-q) K c_u c_w^{k-q-2} \\ &\le K c_u c_1 (\rho_{ii_k}/m)^{-c_3 d_0} k^2 c_w^{k-2} \\ &\le C_0 m^{c_3 d_0 + 2} c_w^k \rho_{ii_k}^{-c_3 d_0},\end{aligned}$$

where the first inequality is from Claim C.1.4; the second one is from Claim C.1.2 and all elements in A_n are $\le c_1(\rho_{ii_k}/m)^{-c_3 d_0}$. And hence, for any i,

$$\begin{aligned}\sum_{i_k \notin B_i(s)} (W_n^k)_{ii_k} &\le C_0 m^{c_3 d_0 + 2} c_w^k \sum_{i_k \notin B_i(s)} \rho_{ii_k}^{-c_3 d_0} \\ &\le C_0 m^{c_3 d_0 + 2} c_w^k c_4 \sum_{\rho=[s]}^{\infty} \rho^{(1-c_3) d_0 - 1} \\ &\le C_1 c_w^k m^{c_3 d_0 + 2} s^{(1-c_3) d_0}.\end{aligned}$$

The last inequality is from Claim C.1.5.

Any chain of W_n^m starting from i in $t_{i,n}(m)$ involves m steps. We can divide these chains into two sets: one set has all its paths staying within $B_i(s)$, and the other set has some paths falling outside $B_i(s)$. For the first set with all nodes in $B_i(s)$, obviously $t_{i,n}(m) - E(t_{i,n}(m)|\mathcal{F}_{i,n}(s)) = 0$. For the second set, divide it into m mutually exclusive subsets:

(i) $\sum_{i_m \notin B_i(s)} (W_n^m)_{ii_m} \varsigma_{i_m,n}^* a$;

(ii) $\sum_{i_{m-1} \notin B_i(s)} \sum_{i_m \in B_i(s)} (W_n^{m-1})_{ii_{m-1}} w_{i_{m-1} i_m,n} \varsigma_{i_m,n}^* a$; etc.

Such a subset can be written as $\sum_{i_{m-k} \notin B_i(s)} \sum_{i_{m-k+1} \in B_i(s)} \cdots \sum_{i_m \in B_i(s)} (W_n^{m-k})_{ii_{m-k}} w_{i_{m-k} i_{m-k+1},n} \cdots w_{i_{m-1} i_m,n} \varsigma_{i_m,n}^* a$ for $k = 0, \ldots, m-1$.

Consider (i): as

$$\begin{aligned}\left|\sum_{i_m \notin B_i(s)} (W_n^m)_{ii_m} \varsigma_{i_m,n}^* a\right| &\le \sum_{i_m \notin B_i(s)} (W_n^m)_{ii_m} |\varsigma_{i_m,n}^* a| \\ &\le \sum_{i_m \notin B_i(s)} |\varsigma_{i_m,n}^* a| C_0 m^{c_3 d_0 + 2} c_w^m \rho_{ii_m}^{-c_3 d_0},\end{aligned}$$

we have

$$\begin{aligned}&\left\|\sum_{i_m \notin B_i(s)} (W_n^m)_{ii_m} \varsigma_{i_m,n}^* a\right\|_p \\ &\le c_{ap} C_0 m^{c_3 d_0 + 2} c_w^m \sum_{\rho=[s]}^{\infty} c_4 \rho^{(1-c_3) d_0 - 1} \\ &\le C_2 c_{ap} m^{c_3 d_0 + 2} c_w^m s^{(1-c_3) d_0}, \qquad \text{(C.2)}\end{aligned}$$

where $c_{ap} = (E|\varsigma_{i_m,n} a|^p)^{1/p}$ for any i_m and n. For (ii),

$$\begin{aligned}&\left|\sum_{i_{m-1} \notin B_i(s)} \sum_{i_m \in B_i(s)} (W_n^{m-1})_{ii_{m-1}} w_{i_{m-1} i_m,n}\right| \\ &\le \|W_n\|_\infty \sum_{i_{m-1} \notin B_i(s)} (W_n^{m-1})_{ii_{m-1}} \le c_w^m m^{c_3 d_0 + 2} s^{(1-c_3) d_0} C_1,\end{aligned}$$

and hence,

$$\begin{aligned}&\left\|\sum_{i_{m-1} \notin B_i(s)} \sum_{i_m \in B_i(s)} (W_n^{m-1})_{ii_{m-1}} w_{i_{m-1} i_m,n} \varsigma_{i_m,n}^* a\right\|_p \\ &\le C_1 c_w^m m^{c_3 d_0 + 2} s^{(1-c_3) d_0} \sum_{i_m \in B_i(s)} c_{ap} \\ &\le C_1 c_w^m m^{c_3 d_0 + 2} s^{(1-c_3) d_0} c_5 s^{d_0} c_{ap} \\ &\le C_3 c_w^m m^{c_3 d_0 + 2} s^{(2-c_3) d_0} c_{ap} \quad \text{by (C.1)}.\end{aligned}$$

And similarly, for subset (k) with $i_{m-k} \notin B_i(s)$ where $1 \le k \le m-1$, we have

$$\begin{aligned}&\sum_{i_{m-k} \notin B_i(s)} \sum_{i_{m-k+1} \in B_i(s)} \cdots \sum_{i_m \in B_i(s)} (W_n^{m-k})_{ii_{m-k}} w_{i_{m-k} i_{m-k+1}} \cdots w_{i_{m-1} i_m,n} \\ &\le \|W_n^k\|_\infty \sum_{i_{m-k} \notin B_i(s)} (W_n^{m-k})_{ii_{m-k}} \\ &\le c_w^m m^{c_3 d_0 + 2} s^{(1-c_3) d_0} C_1, \quad \text{because } \|W_n^k\|_\infty \le c_w^k.\end{aligned}$$

Hence,

$$\begin{aligned}&\left\|\sum_{i_{m-k} \notin B_i(s)} \sum_{i_{m-k+1} \in B_i(s)} \cdots \sum_{i_m \in B_i(s)} (W_n^{m-k})_{ii_{m-k}} \right. \\ &\qquad \left. \times w_{i_{m-k} i_{m-k+1}} \cdots w_{i_{m-1} i_m,n} \varsigma_{i_m,n}^* a\right\|_p \\ &\le C_3 c_w^m m^{c_3 d_0 + 2} s^{(2-c_3) d_0} c_{ap}.\end{aligned}$$

These together imply

$$\begin{aligned}&\|t_{i,n}(m) - E(t_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \\ &\le 2\|\text{summation of paths in } t_{i,n}(m) \text{ with at least} \\ &\qquad \text{one node } i_{m-k} \notin B_i(s)\|_p \\ &\le 2\left(C_2 m^{c_3 d_0 + 2} c_w^m s^{(1-c_3) d_0} c_{ap} + C_3 \sum_{k=1}^{m-1} c_w^m m^{c_3 d_0 + 2} s^{(2-c_3) d_0} c_{ap}\right) \\ &\le c_w^m m^{c_3 d_0 + 3} s^{(2-c_3) d_0} C_{ap}.\end{aligned}$$

To conclude, for any integer m, $\|t_{i,n}(m) - E(t_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \le C_{ap} c_w^m m^{3 + c_3 d_0} \varphi(s)$ with $\varphi(s) = s^{(2-c_3) d_0}$.

Now we show $\|t_{i,n}(m)\|_p \le c_w^m C_{ap1}$. Divide the whole space D into exclusive subsets $B_i(1)$ and $B_i(\rho+1) \setminus B_i(\rho)$, $\rho = 1, 2, \ldots$. Consider the case with $i_m \in B_i(\rho+1) \setminus B_i(\rho)$. For each $\rho \ge 1$, from Eq. (C.2), we have

$$\left\|\sum_{i_m \in B_i(\rho+1) \setminus B_i(\rho)} (W_n^m)_{ii_m} \varsigma_{i_m,n}^* a\right\|_p \le C_2 c_{ap} m^{c_3 d_0 + 2} c_w^m \rho^{(1-c_3) d_0}.$$

For $B_i(1)$, there are two cases: $i_m = i$ and $i_m \ne i$. For the case $i_m = i$, we have $\|(W_n^m)_{ii} \varsigma_{i,n}^* a\|_p \le c_{ap} c_w^m$. For the case $i_m \ne i$, it must be $\rho_{ii_m} = 1$ from Assumption 1. Hence, $\|\sum_{i_m, \rho_{ii_m} = 1} (W_n^m)_{ii_m} \varsigma_{i_m,n}^* a\|_p \le c_{ap} c_5 c_w^m$. Since

$$\begin{aligned}t_{i,n}(m) = e_{i,n} W_n^m \varsigma_n^* a &= \sum_{\rho=1}^{\infty} \sum_{i_m \in B_i(\rho+1) \setminus B_i(\rho)} (W_n^m)_{ii_m} \varsigma_{i_m,n}^* a \\ &\quad + \sum_{i_m \in B_i(1)} (W_n^m)_{ii_m} \varsigma_{i_m,n}^* a,\end{aligned}$$

we have

$$\begin{aligned}\|t_{i,n}(m)\|_p &\le \sum_{\rho=1}^{\infty} \left\|\sum_{i_m \in B_i(\rho+1) \setminus B_i(\rho)} (W_n^m)_{ii_m} \varsigma_{i_m,n}^* a\right\|_p \\ &\quad + \left\|\sum_{i_m \in B_i(1)} (W_n^m)_{ii_m} \varsigma_{i_m,n}^* a\right\|_p\end{aligned}$$

$$\leq C_2 c_{ap} m^{c_3 d_0+2} c_w^m \sum_{\rho=1}^{\infty} \rho^{(1-c_3)d_0} + c_{ap} c_5 c_w^m + c_w^m c_{ap} \leq m^{c_3 d_0+2} c_w^m C_{ap}. \quad \blacksquare$$

Claim C.1.7. *Let $g_{i,n}(m) = e_{i,n}' G_n^m(\lambda) \varsigma_n^* a$, where ς_n^* and a are the same as in Claim C.1.6. Under Assumptions 1, (3.1), and (4.1), suppose $\sup_{i,n} \|\varsigma_{i,n}^*\|_p < \infty$, then $\sup_{i,n} \|g_{i,n}(m)\|_p < \infty$ and $\sup_{i,n} \|g_{i,n}(m) - E(g_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \leq C_{apm} s^{(2-c_3)d_0}$ with C_{apm} being a finite constant.*

Proof of Claim C.1.7. Suppose $|x| < 1$, taking the $(m-1)$th order derivative on both sides of $(1-x)^{-1} = \sum_{k=0}^{\infty} x^k$, we have $(1-x)^{-m}(m-1)! = \sum_{k=m-1}^{\infty} k(k-1)\cdots(k-m+2)x^{k-(m-1)}$. Hence, $G_n^m(\lambda) = (I_n - \lambda W_n)^{-m} W_n^m = \sum_{l=0}^{\infty} C_l^{l+m-1} \lambda^l W_n^{l+m}$, where C_l^{l+m-1} is a binomial coefficient, and by using the results for $t_{i,n}(l+m)$ in Claim C.1.6, we have

$$\|g_{i,n}(m)\|_p \leq \sum_{l=0}^{\infty} (l+m-1)^{m-1} |\lambda|^l \|t_{i,n}(l+m)\|_p \leq c_w^m C_{ap} \sum_{l=0}^{\infty} (l+m)^{m+c_3 d_0+1} |\lambda c_w|^l < \infty$$

and

$$\begin{aligned} &\|g_{i,n}(m) - E(g_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \\ &\leq \sum_{l=0}^{\infty} (l+m-1)^{m-1} |\lambda|^l \|t_{i,n}(l+m) - E(t_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \\ &\leq C_{ap} \sum_{l=0}^{\infty} |\lambda|^l c_w^{l+m} (l+m)^{m+2+c_3 d_0} s^{(2-c_3)d_0} \\ &\leq C_{apm} s^{(2-c_3)d_0}. \quad \blacksquare \end{aligned}$$

C.2. NED properties in Case 2 under Assumption (4.2)

Claim C.2.1. *Under Assumptions 1, (3.1), and (4.2), for any positive integer q, $\sup_n \|W_n^q\|_1 \leq c_w^q c_5 \rho_c^{d_0}$.*

Proof of Claim C.2.1. Consider the kth column sum of W_n^q, as all elements in W_n are non-negative,

$$e_n' W_n^q e_{k,n} = \sum_{i=1}^{n} e_n' W_n^{q-1} e_{i,n} e_{i,n}' W_n e_{k,n} \leq \|W_n^{q-1}\|_\infty \cdot \sum_{i=1}^{n} e_{i,n}' W_n e_{k,n}.$$

Under Assumption (4.2), $w_{ij,n} = 0$ if $j \notin B_i(\rho_c)$, so $\sum_{i=1}^{n} e_{i,n}' W_n e_{k,n} = \sum_{i \in B_k(\rho_c)} w_{ik,n} \leq c_w c_5 \rho_c^{d_0}$. Hence, $e_n' W_n^q e_{k,n} \leq c_w^q c_5 \rho_c^{d_0}$. As "$\leq$" holds for any k and n, we have $\sup_n \|W_n^q\|_1 \leq c_w^q c_5 \rho_c^{d_0}$. $\blacksquare$

Claim C.2.2. *Under Assumptions 1, (3.1), (3.2), and (4.2), $\sup_{\lambda \in \Lambda} \|G_n(\lambda)\|_\infty < \infty$ and $\sup_{\lambda \in \Lambda} \|G_n(\lambda)\|_1 < \infty$.*

Proof of Claim C.2.2. As $G_n(\lambda) = \sum_{l=0}^{\infty} \lambda^l W_n^{l+1}$ and $\sup_{\lambda \in \Lambda} |\lambda c_w| < 1$, we have $\sup_{\lambda \in \Lambda} \|G_n(\lambda)\|_\infty \leq \sum_{l=0}^{\infty} \sup_{\lambda \in \Lambda} |\lambda|^l \|W_n^{l+1}\|_\infty \leq c_w \sum_{l=0}^{\infty} \sup_{\lambda \in \Lambda} |\lambda c_w|^l < \infty$. By Claim C.2.1 on $\|W_n^l\|_1$,

$$\sup_{\lambda \in \Lambda} \|G_n(\lambda)\|_1 \leq \sum_{l=0}^{\infty} \sup_{\lambda \in \Lambda} |\lambda|^l \|W_n^{l+1}\|_1 \leq c_w c_5 \rho_c^{d_0} \sum_{l=0}^{\infty} \sup_{\lambda \in \Lambda} |\lambda c_w|^l < \infty. \quad \blacksquare$$

Claim C.2.3. *If the i,jth element of W_n^m is not zero, then $\rho_{ij} \leq m\rho_c$.*

Proof of Claim C.2.3. The i,jth element of W_n^m is $\sum_{i_1} \sum_{i_2} \cdots \sum_{i_{m-1}} w_{ii_1,n} w_{i_1 i_2,n} \cdots w_{i_{m-1}j,n}$. If it is not zero, then there exists at least one path $i \to i_1 \to \cdots i_{m-1} \to j$ such that all $w_{ii_1,n}, w_{i_1 i_2,n}, \ldots, w_{i_{m-1}j,n}$ are positive. As $w_{ij,n} = 0$ if $j \notin B_i(\rho_c)$, it must be $i_1 \in B_i(\rho_c), i_2 \in B_{i_1}(\rho_c), \ldots, j \in B_{i_{m-1}}(\rho_c)$. Therefore, $\rho_{ij} \leq \rho_{ii_1} + \rho_{i_1 i_2} + \cdots + \rho_{i_{m-1}j} \leq m\rho_c$. $\blacksquare$

Claim C.2.4. *For any positive integer p and $0 < q < 1$, if $s \geq p/(-\ln q) + 1$, then there exists a finite constant c such that $\sum_{l=[s]} l^p q^l < c s^p q^s$, where $[s]$ denotes the largest integer less than or equal to s.*

Proof of Claim C.2.4. Let $f(x) = x^p q^x$, then $f'^{p-1} q^x (p + x \ln q) < 0$ if $x > p/(-\ln q)$. As $s \geq p/(-\ln q) + 1$,

$$\sum_{l=s} l^p q^l < \int_s^\infty x^p q^x dx = -\frac{s^p q^s}{\ln q} - \frac{p}{\ln q} \int_s^\infty x^{p-1} q^x dx < c_0 s^p q^s,$$

where c_0 is a constant. The first inequality holds because the sequence $l^p q^l$ is monotonically decreasing when $l > p/(-\ln q)$. The equality is from integration by parts, and the last inequality is from induction for $\int_s^\infty x^r q^x dx$ for $r = 0, 1, \ldots, p$. Therefore, $\sum_{l=[s]} l^p q^l < \sum_{l=s-1} l^p q^l < c_0 (s-1)^p q^{s-1} < c s^p q^s$. $\blacksquare$

Claim C.2.5. *Let $t_{i,n}(m) = e_{i,n}' W_n^m \varsigma_n^* a$, where ς_n^* and a are the same as in Claim C.1.6. Under Assumptions 1, (3.1) and (4.2), suppose $\sup_{i,n} \|\varsigma_{i,n}^*\|_p < \infty$, then $\sup_{i,n} \|t_{i,n}(m)\|_p \leq C_{ap} m^{d_0} c_w^m$ and $\sup_{i,n} \|t_{i,n}(m) - E(t_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \leq C_{ap1} \varphi(s)$ with C_{ap} and C_{ap1} being positive constants; $\varphi(s) = 1$ if $s \leq m\rho_c$ and $\varphi(s) = 0$ if $s > m\rho_c$.*

Proof of Claim C.2.5. From Claim C.2.3, $e_{i,n}' W_n^m e_{k,n} = 0$ if $k \notin B_i(m\rho_c)$. Therefore,

$$|t_{i,n}(m)| = \left| \sum_k e_{i,n}' W_n^m e_{k,n} e_{k,n}' \varsigma_n^* a \right| = \left| \sum_{k \in B_i(m\rho_c)} e_{i,n}' W_n^m e_{k,n} e_{k,n}' \varsigma_n^* a \right| \leq \max_{k,n} |e_{i,n}' W_n^m e_{k,n}| \sum_{k \in B_i(m\rho_c)} |\varsigma_{k,n}^* a|$$

and hence,

$$\|t_{i,n}(m)\|_p \leq c_w^m \sum_{k \in B_i(m\rho_c)} \|\varsigma_{k,n}^* a\|_p \leq c_w^m c_5 (m\rho_c)^{d_0} c_{ap} = C_{ap} c_w^m m^{d_0},$$

where $c_{ap} = \sup_{i,n} \|\varsigma_{i,n}^* a\|_p$ and $C_{ap} = c_{ap} c_5 \rho_c^{d_0}$.

Next, we show the NED property. For the spatial weight matrix without row-normalization, $w_{ij,n}$ is a function of $\varsigma_{i,n}$ and $\varsigma_{j,n}$, and $w_{ij,n} = 0$ if $j \notin B_i(\rho_c)$. For the row-normalized case, $w_{ij,n}$ may be related to many points in $B_i(\rho_c)$ and in general is a function of ς's at those locations. In both cases, all the locations of nodes in the chains of $e_{i,n}' W_n^m$ related to $t_{i,n}(m)$ are within the ball $B_i(m\rho_c)$. Hence, when $s > m\rho_c$, $t_{i,n}(m) - E(t_{i,n}(m)|\mathcal{F}_{i,n}(s)) = 0$. With $s \leq m\rho_c$,

$$\|t_{i,n}(m) - E(t_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \leq 2\|t_{i,n}(m)\|_p \leq 2C_{ap} c_w^m m^{d_0}.$$

Therefore, the NED property follows if we choose $\varphi(s) = 1$ for $s \leq m\rho_c$ and $\varphi(s) = 0$ for $s > m\rho_c$. $\blacksquare$

Claim C.2.6. *Denote $g_{i,n}(m) = e_{i,n}G_n^m(\lambda)\varsigma_n^* a$, where ς_n^* and a are the same as in Claim C.1.6. Under Assumptions 1, (3.1), and (4.2), suppose $\sup_{i,n}\|\varsigma_{i,n}^*\|_p < \infty$, then $\sup_{i,n}\|g_{i,n}(m)\|_p < \infty$ and $\sup_{i,n}\|g_{i,n}(m) - E(g_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \le C_{apm}\varphi(s)$ with C_{apm} being a finite constant; $\varphi(s) = 1$ if $s \le m\rho_c$ and $\varphi(s) = s^{d_0+m-1}|\lambda c_w|^{s/\rho_c}$ if $s > m\rho_c$.*

Proof of Claim C.2.6. From the proof of Claim C.1.7, $g_{i,n}(m) = \sum_{l=0}^{\infty} C_l^{l+m-1}\lambda^l t_{i,n}(l+m)$. If $\lambda = 0$, then $g_{i,n}(m) = t_{i,n}(m)$ and the claim follows from Claim C.2.5. For $\lambda \neq 0$, by Claim C.2.5, for any i and n,

$$\|g_{i,n}(m)\|_p \le c_w^m C_{ap}\sum_{l=0}^{\infty}|\lambda c_w|^l (l+m)^{d_0+m-1},$$

which is finite and denoted as C_m. Thus, for $s > 0$, $\|g_{i,n}(m) - E(g_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \le 2\|g_{i,n}(m)\|_p \le 2C_m$. Now consider the case when $s > m\rho_c$. Given such an s, from Claim C.2.5, $t_{i,n}(m+l) - E(t_{i,n}(m+l)|\mathcal{F}_{i,n}(s)) = 0$ for any nonnegative integer l such that $s > (m+l)\rho_c$. Such a set of l will be determined by $l < (\frac{s}{\rho_c} - m)$. Therefore, when $s > m\rho_c$,

$$\begin{aligned}
&\|g_{i,n}(m) - E(g_{i,n}(m)|\mathcal{F}_{i,n}(s))\|_p \\
&= \left\| \sum_{l=\left[\frac{s}{\rho_c}-m\right]}^{\infty} C_l^{l+m-1}\lambda^l [t_{i,n}(l+m) - E(t_{i,n}(l+m)|\mathcal{F}_{i,n}(s))] \right\|_p \\
&\le 2\sum_{l=\left[\frac{s}{\rho_c}-m\right]}^{\infty} (l+m)^{m-1}|\lambda|^l \|t_{i,n}(l+m)\|_p \\
&\le 2C_{ap}c_w^m \sum_{l=\left[\frac{s}{\rho_c}-m\right]}^{\infty} |\lambda c_w|^l (l+m)^{m-1+d_0},
\end{aligned}$$

where the last inequality follows from Claim C.2.5. By the inequality in Claim C.2.4, as $s/\rho_c > m$, we have

$$\begin{aligned}
\sum_{l=\left[\frac{s}{\rho_c}-m\right]}^{\infty} |\lambda c_w|^{l+m}(l+m)^{m-1+d_0}/|\lambda|^m &= \sum_{l=\left[\frac{s}{\rho_c}\right]}^{\infty} |\lambda c_w|^l l^{m-1+d_0}/|\lambda|^m \\
&= O(s^{m+d_0-1}|\lambda c_w|^{s/\rho_c}).
\end{aligned}$$

The claim would follow if we set $\varphi(s) = 1$ if $s \le m\rho_c$ and $\varphi(s) = s^{d_0+m-1}|\lambda c_w|^{s/\rho_c}$ if $s > m\rho_c$. ■

C.3. Proofs of main results

Proof of Proposition 1. As $M_n = A_n' B_n$, if we denote $a'\varsigma_n^{*\prime}M_n\varsigma_n^* b = \sum_{i=1}^n q_{i,n}$, then $q_{i,n} = a_{i,n}^* b_{i,n}^*$, where $a_{i,n}^* = e_{i,n}A_n\varsigma_n^* a$ and $b_{i,n}^* = e_{i,n}B_n\varsigma_n^* b$ can be either $t_{i,n}(m_1)$ or $g_{i,n}(m_2)$ for any finite integers m_1 and m_2. Under Assumption (4.1), Claims C.1.6, C.1.7, and B.3 give us $\|q_{i,n}\|_{p/2} \le \|a_{i,n}^*\|_p \cdot \|b_{i,n}^*\|_p < \infty$ and $\|q_{i,n} - E[q_{i,n}|\mathcal{F}_{i,n}(s)]\|_2 \le C_m s^{(2-c_3)d_0}$, with C_m being a finite constant. Under Assumption (4.2) Claims C.2.5, C.2.6 and B.3 give us $\|q_{i,n}\|_{p/2} \le \|a_{i,n}^*\|_p \cdot \|b_{i,n}^*\|_p < \infty$ and $\|q_{i,n} - E[q_{i,n}|\mathcal{F}_{i,n}(s)]\|_2 \le C_q\varphi(s)$ with $\varphi(s) = 1$ if $s \le s_m$ and $\varphi(s) = s^{d_0+m-1}|\lambda c_w|^{s/\rho_c}$ if $s > s_m$, where C_m and s_m are some finite constants. For both cases of W_n, conditions in Claim B.4 are satisfied. Therefore, $\frac{1}{n}E|a'\varsigma_n^{*\prime}M_n\varsigma_n^* b| = O(1)$ and $\frac{1}{n}[a'\varsigma_n^{*\prime}M_n\varsigma_n^* b - E(a'\varsigma_n^{*\prime}M_n\varsigma_n^* b)] = o_p(1)$. ■

Proof of Corollary 1. We have $\frac{1}{n}[a'\varsigma_n^*(\theta)'G_n^{m_1}(\lambda)'G_n^{m_2}(\lambda)\varsigma_n^*(\theta)b - E(a'\varsigma_n^*(\theta)G_n^{m_1}(\lambda)'G_n^{m_2}(\lambda)\varsigma_n^*(\theta)b)] = o_p(1)$ pointwisely for any θ from Proposition 1. As θ enters $\varsigma_n^*(\theta)$ polynomially and the parameter space of θ is compact, to show the ULLN, we only need to show the stochastic equicontinuity of $\frac{1}{n}a'\varsigma_n^{*\prime}G_n^{m_1}(\lambda)'G_n^{m_2}(\lambda)\varsigma_n^* b$. By the mean value theorem,

$$\begin{aligned}
&\left|a'\varsigma_n^{*\prime}G_n^{m_1}(\lambda_1)'G_n^{m_2}(\lambda_1)\varsigma_n^* b - a'\varsigma_n^{*\prime}G_n^{m_1}(\lambda_2)'G_n^{m_2}(\lambda_2)\varsigma_n^* b\right| \\
&= \left|(\lambda_1-\lambda_2)a'\varsigma_n^{*\prime}A_n(\bar\lambda)\varsigma_n^* b\right| \\
&\le |\lambda_1-\lambda_2|\left(a'\varsigma_n^{*\prime}\varsigma_n^* a\right)^{\frac{1}{2}}\left(b'\varsigma_n^{*\prime}A_n(\bar\lambda)'A_n(\bar\lambda)\varsigma_n^* b\right)^{\frac{1}{2}} \\
&\le |\lambda_1-\lambda_2|\left(a'\varsigma_n^{*\prime}\varsigma_n^* a\right)^{\frac{1}{2}}\left(b'\varsigma_n^{*\prime}\varsigma_n^* b\right)^{\frac{1}{2}}[\mu_{\max}(A_n(\bar\lambda)'A_n(\bar\lambda))]^{\frac{1}{2}} \\
&\le |\lambda_1-\lambda_2|\left(a'\varsigma_n^{*\prime}\varsigma_n^* a\right)^{1/2}\left(b'\varsigma_n^{*\prime}\varsigma_n^* b\right)^{1/2} \\
&\quad \times \left(\sup_{\lambda\in\Lambda}\|A_n'(\lambda)A_n(\lambda)\|_\infty\right)^{1/2},
\end{aligned}$$

where $\bar\lambda$ is between λ_1 and λ_2, $A_n(\lambda) = G_n^{m_1}(\lambda)'[m_2G_n(\lambda) + m_1G_n^{m_1}(\lambda)']G_n^{m_2}(\lambda)$, and $\mu_{\max}(\cdot)$ is the largest eigenvalue of the matrix inside. The first inequality is from the Cauchy–Schwarz inequality, the second inequality holds as $A_n(\bar\lambda)'A_n(\bar\lambda)$ is non-negative definite, and the last inequality is from the spectral radius theorem. From Claims C.1.3 and C.2.2, $\sup_{\lambda\in\Lambda}\|G_n(\lambda)\|_\infty < \infty$ and $\sup_{\lambda\in\Lambda}\|G_n(\lambda)\|_1 < \infty$, so $\sup_{\lambda\in\Lambda}\|A_n'(\lambda)A_n(\lambda)\|_\infty < \infty$. As $\frac{1}{n}a'\varsigma_n^{*\prime}\varsigma_n^* a = O_p(1)$ and $\frac{1}{n}b'\varsigma_n^{*\prime}\varsigma_n^* b = O_p(1)$, we have

$$\begin{aligned}
&\sup_{|\lambda_1-\lambda_2|<\delta^*}\frac{1}{n}|a'\varsigma_n^{*\prime}G_n^{m_1}(\lambda_1)'G_n^{m_2}(\lambda_1)\varsigma_n^* b \\
&\quad - a'\varsigma_n^{*\prime}G_n^{m_1}(\lambda_2)'G_n^{m_2}(\lambda_2)\varsigma_n^* b| = O_p(\delta^*).
\end{aligned}$$

Then the ULLN follows. ■

Proof of Proposition 2. Similarly to the proof of Proposition 1, denote $a_j'\varsigma_n^{*\prime}M_{jn}\varsigma_n^* b_j = \sum_{i=1}^n q_{i,n}(j)$, then $r_{i,n} = \sum_{j=1}^m q_{i,n}(j)$. Each $q_{i,n}(j)$ is L_2-NED on the i.i.d. random field $\varsigma = (\varepsilon, \xi)$ with a finite NED scaling factor. It is straightforward to show $\|r_{i,n}\|_{2+\delta_\varepsilon} < \infty$. For the case in Assumption (4.1), Claims C.1.6 and C.1.7 give the same NED coefficient $\varphi(s) = s^{(2-c_3)d_0}$ for each $q_{i,n}(j)$. Therefore, by Claim B.3, the NED coefficient for $r_{i,n}$ is also $\varphi(s) = s^{(2-c_3)d_0}$. As $c_3 > 3$, $\sum_{r=1}^{\infty} r^{d_0-1}\varphi(r) = \sum_{r=1}^{\infty} r^{(3-c_3)d_0-1} < \infty$. For the case in Assumption (4.2), Claims C.2.5, C.2.6, and B.3 give the NED coefficient $\varphi(s) = s^{d_0+m-1}|\lambda c_w|^{s/\rho_c}$ if $s > \overline{m}\rho_c$, otherwise, $\varphi(s) = 1$, where $\overline{m}$ is the highest power of G_n^m in M_{jn}'s. Therefore, $\sum_{r=1}^{\infty} r^{d_0-1}\varphi(r) = \sum_{r=1}^{\overline{m}\rho_c} r^{d_0-1} + \sum_{r=\overline{m}\rho_c+1}^{\infty} r^{d_0+m-1}|\lambda c_w|^{r/\rho_c} < \infty$. All the four conditions in Claim B.5 are satisfied and hence, $R_n/\sigma_{Rn} \xrightarrow{d} N(0,1)$. ■

Proof of Theorem 1. Under Assumptions 1–5, by applying Proposition 1, $\widehat\kappa - \kappa_0 \xrightarrow{p} a\lim_{n\to\infty}\frac{1}{n}E(Q_n'\xi_n) + b\lim_{n\to\infty}\frac{1}{n}E(X_{2n}'\varepsilon_n\delta_0)$, where

$$a = \left(H_q'\left[\lim_{n\to\infty}E\left(\frac{Q_n'Q_n}{n}\right)\right]^{-1}H_q\right)^{-1}H_q'\left[\lim_{n\to\infty}E\left(\frac{Q_n'Q_n}{n}\right)\right]^{-1}$$

and $\quad b = a\lim_{n\to\infty}E\left(\frac{Q_n'X_{2n}}{n}\right)\left(\lim_{n\to\infty}\frac{X_{2n}'X_{2n}}{n}\right)^{-1}$

with $H_q = \lim_{n\to\infty}\frac{1}{n}[E(Q_n'G_n)X_{1n}\beta_0 + E(Q_n'G_n\varepsilon_n)\delta_0, E(Q_n')X_{1n}, E(Q_n'\varepsilon_n)]$. As $E(Q_n'\xi_n) = 0$ and $E(X_{2n}'\varepsilon_n) = 0$, we have $\widehat\kappa - \kappa_0 \xrightarrow{p} 0$. Under given assumptions, since $\widehat\kappa - \kappa_0$ can be written as a form of R_n in Proposition 2, $\sqrt{n}(\widehat\kappa - \kappa_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{IV}})$. Similarly, we can show $\sqrt{n}(\widehat\kappa_G - \kappa_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{GIV}})$. ■

Proof of Theorem 2. Let $\widehat\kappa_{\mathrm{BGIV}}$ be the best G2SIV estimator with the corresponding optimal IV matrix $Q_n^* = [G_nX_{1n}, G_nZ_n, X_n, Z_n]$. As $\sqrt{n}(\widehat\kappa_{\mathrm{BGIV}} - \kappa_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{BGIV}})$ from Theorem 1, to show $\sqrt{n}(\widehat\kappa_{\mathrm{FBGIV}} - \kappa_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{BGIV}})$, it is sufficient to show $\sqrt{n}(\widehat\kappa_{\mathrm{FBGIV}} -$

$\widehat{\kappa}_{\mathrm{BGIV}}) = o_p(1)$. Denote $\widehat{Q}_n^* = [G_n(\widehat{\lambda})X_{1n}, G_n(\widehat{\lambda})Z_n, X_n, Z_n]$, $\upsilon_0 = \frac{\delta_0'\Sigma_{\varepsilon 0}\delta_0/\sigma_{\xi 0}^2}{\sigma_{\xi 0}^2+\delta_0'\Sigma_{\varepsilon 0}\delta_0}$, and $\widehat{\upsilon} = \frac{\widehat{\delta}'\widehat{\Sigma}_\varepsilon\widehat{\delta}/\widehat{\sigma}_\xi^2}{\widehat{\sigma}_\xi^2+\widehat{\delta}'\widehat{\Sigma}_\varepsilon\widehat{\delta}}$. Then

$$\begin{aligned}\widehat{\kappa}_{\mathrm{FBGIV}} - \kappa_0 = {} & [(W_nY_n, X_{1n}, P_n^{\perp}Z_n)'\widehat{\Pi}_n^{-1}\widehat{Q}_n^*(\widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}\widehat{Q}_n^*)^{-1} \\ & \times \widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}(W_nY_n, X_{1n}, P_n^{\perp}Z_n)]^{-1} \\ & \cdot (W_nY_n, X_{1n}, P_n^{\perp}Z_n)'\widehat{\Pi}_n^{-1}\widehat{Q}_n^*(\widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}\widehat{Q}_n^*)^{-1} \\ & \times \widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}(\xi_n + P_n\varepsilon_n\delta_0).\end{aligned}$$

We will show $\frac{1}{n}(\widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}\widehat{Q}_n^* - Q_n^{*\prime}\Pi_n^{-1}Q_n^*) = o_p(1)$, $\frac{1}{n}[\widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}(W_nY_n, X_{1n}, P_n^{\perp}Z_n) - Q_n^{*\prime}\Pi_n^{-1}(W_nY_n, X_{1n}, P_n^{\perp}Z_n)] = o_p(1)$, and $\frac{1}{\sqrt{n}}\widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}(\xi_n + P_n\varepsilon_n\delta_0) - \frac{1}{\sqrt{n}}Q_n^{*\prime}\Pi_n^{-1}(\xi_n + P_n\varepsilon_n\delta_0) = o_p(1)$. As

$$\frac{1}{n}\left(\widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}\widehat{Q}_n^* - Q_n^{*\prime}\Pi_n^{-1}Q_n^*\right) = \frac{1}{n}\left(\frac{1}{\widehat{\sigma}_\xi^2}\widehat{Q}_n^{*\prime}\widehat{Q}_n^* - \frac{1}{\sigma_{\xi 0}^2}Q_n^{*\prime}Q_n^*\right) - \frac{1}{n}\left(\widehat{\upsilon}\widehat{Q}_n^{*\prime}P_n\widehat{Q}_n^* - \upsilon_0 Q_n^{*\prime}P_nQ_n^*\right),$$

we can show each part is $o_p(1)$. From the proof of Corollary 1, $\sup_\lambda \|\frac{1}{n}\widehat{Q}_n^{*\prime}\widehat{Q}_n^*\| = O_p(1)$ and $\sup_\lambda \frac{1}{n}\|\widehat{Q}_n^{*\prime}\widehat{Q}_n^* - Q_n^{*\prime}Q_n^*\| = o_p(1)$, so

$$\begin{aligned}&\frac{1}{n}\left(\frac{1}{\widehat{\sigma}_\xi^2}\widehat{Q}_n^{*\prime}\widehat{Q}_n^* - \frac{1}{\sigma_{\xi 0}^2}Q_n^{*\prime}Q_n^*\right) \\ &\quad = \left(\frac{1}{\widehat{\sigma}_\xi^2} - \frac{1}{\sigma_{\xi 0}^2}\right)\frac{1}{n}\widehat{Q}_n^{*\prime}\widehat{Q}_n^* + \frac{1}{\sigma_{\xi 0}^2}\left(\frac{1}{n}\widehat{Q}_n^{*\prime}\widehat{Q}_n^* - \frac{1}{n}Q_n^{*\prime}Q_n^*\right) \\ &\quad = o_p(1).\end{aligned}$$

With same arguments, $\frac{1}{n}\left(\widehat{\upsilon}\widehat{Q}_n^{*\prime}P_n\widehat{Q}_n^* - \upsilon_0Q_n^{*\prime}P_nQ_n^*\right) = o_p(1)$. Together, we have $\frac{1}{n}(\widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}\widehat{Q}_n^* - Q_n^{*\prime}\Pi_n^{-1}Q_n^*) = o_p(1)$. Similarly, we can show $\frac{1}{n}[\widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}(W_nY_n, X_{1n}, P_n^{\perp}Z_n) - Q_n^{*\prime}\Pi_n^{-1}(W_nY_n, X_{1n}, P_n^{\perp}Z_n)] = o_p(1)$. It remains to show $\frac{1}{\sqrt{n}}\widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}(\xi_n + P_n\varepsilon_n\delta_0) - \frac{1}{\sqrt{n}}Q_n^{*\prime}\Pi_n^{-1}(\xi_n + P_n\varepsilon_n\delta_0) = o_p(1)$. From Propositions 1 and 2, and Corollary 1, $\sqrt{n}(\frac{1}{\widehat{\sigma}_\xi^2} - \frac{1}{\sigma_{\xi 0}^2}) = O_p(1)$, $\frac{1}{n}\widehat{Q}_n^{*\prime}(\xi_n + P_n\varepsilon_n\delta_0) = o_p(1)$, $\frac{1}{\sqrt{n}}Q_n^{*\prime}(\xi_n + P_n\varepsilon_n\delta_0) = O_p(1)$, and $\frac{1}{\sqrt{n}}(\widehat{Q}_n^* - Q_n^*)'(\xi_n + P_n\varepsilon_n\delta_0) = o_p(1)$ as initial estimates are $\sqrt{n}$-consistent, so

$$\frac{1}{\sqrt{n}\widehat{\sigma}_\xi^2}\widehat{Q}_n^{*\prime}(\xi_n + P_n\varepsilon_n\delta_0) - \frac{1}{\sqrt{n}\sigma_{\xi 0}^2}Q_n^{*\prime}(\xi_n + P_n\varepsilon_n\delta_0) = o_p(1).$$

Similarly, $\frac{1}{\sqrt{n}}\upsilon_0Q_n^{*\prime}P_n(\xi_n + P_n\varepsilon_n\delta_0) - \frac{1}{\sqrt{n}}\widehat{\upsilon}\widehat{Q}_n^{*\prime}P_n(\xi_n + P_n\varepsilon_n\delta_0) = o_p(1)$. As $\Pi_n^{-1} = \frac{1}{\sigma_{\xi 0}}I_n - \upsilon_0P_n$,

$$\frac{1}{\sqrt{n}}\widehat{Q}_n^{*\prime}\widehat{\Pi}_n^{-1}(\xi_n + P_n\varepsilon_n\delta_0) - \frac{1}{\sqrt{n}}Q_n^{*\prime}\Pi_n^{-1}(\xi_n + P_n\varepsilon_n\delta_0) = o_p(1).$$

These together complete the proof $\sqrt{n}(\widehat{\kappa}_{\mathrm{FBGIV}} - \widehat{\kappa}_{\mathrm{BGIV}}) = o_p(1)$. ■

Claim C.3.1. *Under Assumptions 1–4, and 6, θ_0 is the unique maximizer of $\lim_{n\to\infty}\frac{1}{n}E\ln L_n(\theta)$.*

Proof of Claim C.3.1. We want to show $\lim_{n\to\infty}\frac{1}{n}[E(\ln L_n(\theta)) - E(\ln L_n(\theta_0))] \le 0$ and the equality holds iff $\theta = \theta_0$. From Appendix A.1, we have

$$\begin{aligned}&\frac{1}{n}[E(\ln L_n(\theta)) - E(\ln L_n(\theta_0))] \\ &\quad = -\frac{1}{2}\ln\frac{\sigma_\xi^2}{\sigma_{\xi 0}^2} - \frac{1}{2}\ln\frac{|\Sigma_\varepsilon|}{|\Sigma_{\varepsilon 0}|} + \frac{1}{n}E\left(\ln\frac{|S_n(\lambda)|}{|S_n|}\right) \\ &\qquad - \frac{1}{2}\mathrm{tr}(\Sigma_{\varepsilon 0}^{1/2}\Sigma_\varepsilon^{-1}\Sigma_{\varepsilon 0}^{1/2} - I_{p_2}) \\ &\qquad - \frac{1}{2n}\sum_{i=1}^n x_{2,in}'(\Gamma_0 - \Gamma)\Sigma_\varepsilon^{-1}(\Gamma_0 - \Gamma)'x_{2,in} \\ &\qquad + \frac{1}{2} - \frac{\sigma_{\xi 0}^2}{2n\sigma_\xi^2}E[\mathrm{tr}(S_n^{-1\prime}S_n(\lambda)'S_n(\lambda)S_n^{-1\prime})] \\ &\qquad - \frac{1}{2\sigma_\xi^2}((\lambda_0 - \lambda), (\beta_0 - \beta)', ((\Gamma - \Gamma_0)\delta)', (\delta_0 - \delta)') \\ &\qquad \times H_{1n}((\lambda_0 - \lambda), (\beta_0 - \beta)', ((\Gamma - \Gamma_0)\delta)', (\delta_0 - \delta)')' \\ &\quad = -\frac{1}{2}[\mathrm{tr}(\Sigma_{\varepsilon 0}^{1/2}\Sigma_\varepsilon^{-1}\Sigma_{\varepsilon 0}^{1/2}) - \ln|\Sigma_{\varepsilon 0}^{1/2}\Sigma_\varepsilon^{-1}\Sigma_{\varepsilon 0}^{1/2}| - p_2] \\ &\qquad - \frac{1}{2n}\sum_{i=1}^n x_{2,in}'(\Gamma_0 - \Gamma)\Sigma_\varepsilon^{-1}(\Gamma_0 - \Gamma)'x_{2,in} \\ &\qquad - \frac{1}{2\sigma_\xi^2}((\lambda_0 - \lambda), (\beta_0 - \beta)', ((\Gamma - \Gamma_0)\delta)', (\delta_0 - \delta)') \\ &\qquad \times H_{1n}((\lambda_0 - \lambda), (\beta_0 - \beta)', ((\Gamma - \Gamma_0)\delta)', (\delta_0 - \delta)')' \\ &\qquad - \frac{1}{2n}E\left[\mathrm{tr}\left(\frac{\sigma_{\xi 0}^2}{\sigma_\xi^2}S_n^{-1\prime}S_n(\lambda)'S_n(\lambda)S_n^{-1}\right)\right. \\ &\qquad \left. - \ln\left|\frac{\sigma_{\xi 0}^2}{\sigma_\xi^2}S_n^{-1\prime}S_n(\lambda)'S_n(\lambda)S_n^{-1}\right| - n\right]. \end{aligned} \tag{C.3}$$

First we show $\frac{1}{n}[E(\ln L_n(\theta)) - E(\ln L_n(\theta_0))] \le 0$. By the concavity of $\ln x$, for any $x > 0$, the function $f(x) = x - \ln x - 1 \ge 0$ and it is minimized only at $x = 1$. Also for any positive definite real value matrix M, $f(M) = \mathrm{tr}(M) - \ln|M| - m = \sum_{i=1}^m(\varphi_i - \ln\varphi_i - 1) \ge 0$ and is minimized only at $M = I_m$, where m is the dimension of M and $\varphi_i's$ $(i = 1, \ldots, m)$ are eigenvalues of M. Therefore, $\frac{1}{n}[E(\ln L_n(\theta)) - E(\ln L_n(\theta_0))] \le 0$.

Now we show that $\lim_{n\to\infty}\frac{1}{n}[E(\ln L_n(\theta)) - E(\ln L_n(\theta_0))] = 0$ implies $\theta = \theta_0$. All the four terms in (C.3) are zero. Since $f(\Sigma_{\varepsilon 0}^{1/2}\Sigma_\varepsilon^{-1}\Sigma_{\varepsilon 0}^{1/2}) = 0$, it must be $\Sigma_\varepsilon = \Sigma_{\varepsilon 0}$. As $\lim_{n\to\infty}\frac{1}{n}X_{2n}'X_{2n}$ is p.d., it must be $\Gamma_0 = \Gamma$. The third and fourth terms imply $\lim_{n\to\infty}((\lambda_0 - \lambda), (\beta_0 - \beta)', ((\Gamma - \Gamma_0)\delta)', (\delta_0 - \delta)')H_{1n} = 0$ and $\frac{\sigma_{\xi 0}^2}{\sigma_\xi^2}S_n^{-1\prime}S_n(\lambda)'S_n(\lambda)S_n^{-1} = I_n$ with probability one. With $\Gamma_0 = \Gamma$, $\lim_{n\to\infty}((\lambda_0 - \lambda), (\beta_0 - \beta)', ((\Gamma - \Gamma_0)\delta)', (\delta_0 - \delta)')H_{1n} = 0$ is equivalent to $((\lambda_0 - \lambda), (\beta_0 - \beta)', (\delta_0 - \delta)')H_n = 0$, where $H_n = \frac{1}{n}E[(G_n(X_{1n}\beta_0 + \varepsilon_n\delta_0), X_{1n}, \varepsilon_n)'(G_n(X_{1n}\beta_0 + \varepsilon_n\delta_0), X_{1n}, \varepsilon_n)]$.

Under Assumption 6(a) that H_n is p.d., we have $\lambda_0 = \lambda$, $\beta_0 = \beta$, and $\delta_0 = \delta$. Under Assumption 6(b), as $S_n(\lambda)'S_n(\lambda)$ is linearly independent of $S_n'S_n$ with probability one, i.e., for any $\lambda \ne \lambda_0$, no value of σ_ξ^2 can make the equality $\frac{\sigma_{\xi 0}^2}{\sigma_\xi^2}S_n^{-1\prime}S_n(\lambda)'S_n(\lambda)S_n^{-1} = I_n$ hold with probability one, then, it must be $\lambda = \lambda_0$ and $\sigma_\xi^2 = \sigma_{\xi 0}^2$. Since $\lim_{n\to\infty}\frac{1}{n}X_{1n}'X_{1n}$ is p.d., the third term being zero implies $\beta = \beta_0$ and $\delta = \delta_0$. ■

Claim C.3.2. *Under Assumptions 1–3, and 6, the information matrix I_{θ_0} is positive definite.*

Proof of Claim C.3.2. The $I_{\theta_0} = -\lim_{n\to\infty}E\left(\frac{1}{n}\frac{\partial^2\ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right)$. Since X_n is made of all distinct column vectors of X_{1n} and X_{2n}, we can write $X_{1n}\beta_0 = X_n\beta_0^+$ and $X_{2n}\Gamma_0 = X_n\Gamma_0^+$, where some elements in β^+ and γ^+ are zero. To show I_{θ_0} is p.d., it is sufficient to show that $I_{\theta_0}^+$ is p.d., where $I_{\theta_0}^+$ is the information matrix for $L_n(\theta^+)$ and $\theta^+ = (\lambda, \beta^{+\prime}, \mathrm{vec}(\Gamma^+)', \sigma_\xi^2, \alpha', \delta')'$ without constraints on some elements of β_0^+ and Γ_0^+ being zero. Let $C_I =$

$(c_{I1}, c'_{I2}, \mathrm{vec}(c_{I3})', c_{I4}, c'_{I5}, c'_{I6})'$ be a $(k + kp_2 + J + p_2 + 2)$ dimensional column vector of constants, where c_{I1} and c_{I4} are constants; c_{I2}, c_{I5}, and c_{I6} are column vectors of dimension k, J, and p_2; c_3 is a $k \times p_2$ matrix. To prove $I^+_{\theta_0}$ is p.d., it is sufficient to show that the $C_I = 0$ is the only solution to $I^+_{\theta_0} C_I = 0$. From the second row block of the linear equation system $I^+_{\theta_0} C_I = 0$, we have

$$\lim_{n\to\infty} \frac{1}{n}[-c_{I1}X'_n E(G_n X_n \beta^+_0 + G_n \varepsilon_n \delta_0) - X'_n X_n c_{I2} + (\delta'_0 \otimes (X'_n X_n))\mathrm{vec}(c_{I3})] = 0.$$

From the third row block, we have

$$\lim_{n\to\infty} \frac{1}{n}[c_{I1}(\delta_0 \otimes X'_n)E(G_n X_n \beta_0 + G_n \varepsilon_n \delta_0) + (\delta_0 \otimes (X'_n X_n))c_{I2} - ((\sigma^2_{\xi 0}\Sigma^{-1}_{\varepsilon 0} + \delta_0\delta'_0) \otimes (X'_n X_n))\mathrm{vec}(c_{I3})] = 0.$$

By canceling out $\lim_{n\to\infty} \frac{1}{n} X'_n X_n c_{I2}$ in above two equations, we have $\lim_{n\to\infty} \frac{1}{n}(\Sigma^{-1}_{\varepsilon 0} \otimes (X'_n X_n))\mathrm{vec}(c_{I3})\sigma^2_{\xi 0} = 0$. As $\lim_{n\to\infty} \frac{X'_n X_n}{n}$ is p.d., it follows that $c_{I3} = 0$. Now with $c_{I3} = 0$, $c_{I2} = -c_{I1}(\lim_{n\to\infty} \frac{X'_n X_n}{n})^{-1} \lim_{n\to\infty} \frac{1}{n}[X'_n E(G_n X_n \beta^+_0 + G_n \varepsilon_n \delta_0)]$. From the fourth row block, we have $c_{I4} = 2\sigma^2_{\xi 0} \lim_{n\to\infty} \frac{1}{n} E[-c_{I1}\mathrm{tr}(G_n)]$. From the fifth row block, we have $c_{I5} = 0$. From the sixth row block, we have $c_{I6} = -c_{I1}\Sigma^{-1}_{\varepsilon 0} \lim_{n\to\infty} \frac{1}{n} E[\varepsilon'_n G_n (X_{1n}\beta^+_0 + \varepsilon_n \delta_0)]$. From the first row block, we have

$$\begin{aligned} 0 = \lim_{n\to\infty} \frac{1}{n}\{&-c_{I1}[\sigma^2_{\xi 0}\mathrm{tr}[E(G^2_n + G_n G'_n)] \\ &+ E[(X_n\beta^+_0 + G_n \varepsilon_n \delta_0)' G'_n G_n (X_n \beta^+_0 + G_n \varepsilon_n \delta_0)]] \\ &- c'_{I2} X'_n E(G_n X_n \beta^+_0 + G_n \varepsilon_n \delta_0) + \mathrm{vec}(c_{I3})' \\ &\times [\delta_0 \otimes X'_n E(G_n X_n \beta^+_0 + G_n \varepsilon_n \delta_0)] - c_{I4} E[\mathrm{tr}(G_n)] \\ &- c'_{I6} E(\varepsilon'_n G_n (X_n \beta^+_0 + \varepsilon_n \delta_0))\}. \end{aligned}$$

Plugging in $c_{I2}, \ldots, c_{I6}$ from the above, we have

$$\begin{aligned} 0 = &-c_{I1} \lim_{n\to\infty} \frac{1}{n}[\sigma^2_{\xi 0}\mathrm{tr}[E(G^2_n + G_n G'_n)]] \\ &- \lim_{n\to\infty} c_{I1} \frac{1}{n} E(H'_n H_n) + 2c_{I1}\sigma^2_{\xi 0}\left(\lim_{n\to\infty} E\frac{\mathrm{tr}(G_n)}{n}\right)^2 \\ &+ c_{I1} \lim_{n\to\infty} \frac{1}{n} E(H'_n) X_n (X'_n X_n)^{-1} X'_n E(H_n) \\ &+ c_{I1}\left(\lim_{n\to\infty} E\frac{\varepsilon'_n H_n}{n}\right)' \Sigma^{-1}_{\varepsilon 0}\left(\lim_{n\to\infty} E\frac{\varepsilon'_n H_n}{n}\right), \end{aligned}$$

where $H_n = G_n(X_n \beta^+_0 + \varepsilon_n \delta_0) = G_n(X_{1n}\beta_0 + \varepsilon_n \delta_0)$. By Cauchy–Schwarz inequality, $E(H'_n H_n) - E(H'_n)E(H_n) \geq \frac{1}{n} E(H'_n \varepsilon_n)\Sigma^{-1}_{\varepsilon 0} E(\varepsilon'_n H_n)$. Hence,

$$\begin{aligned} &E(H'_n H_n) - \frac{1}{n} E(H'_n \varepsilon_n)\Sigma^{-1}_{\varepsilon 0} E(\varepsilon'_n H_n) - E(H'_n) X_n (X'_n X_n)^{-1} X'_n E(H_n) \\ &\geq E(H'_n)E(H_n) - E(H'_n)[X_n (X'_n X_n)^{-1} X'_n] E(H_n) \\ &= E(H'_n)[I_n - X_n (X'_n X_n)^{-1} X'_n] E(H_n) \geq 0. \end{aligned}$$

As $E[\mathrm{tr}(G^2_n + G_n G'_n)] - \frac{2}{n} E^2[\mathrm{tr}(G_n)] \geq E[\mathrm{tr}(G^2_n + G_n G'_n - \frac{2}{n}\mathrm{tr}^2(G_n))] = \frac{1}{2} E[\mathrm{tr}(G_n + G'_n - 2\mathrm{tr}(G_n) I_n / n)^2] \geq 0$ by Assumption 6(b) and $\lim_{n\to\infty} \frac{1}{n} E(H'_n)[I_n - X_n (X'_n X_n)^{-1} X'_n] E(H_n)$ is p.d. by Assumption 6(a), it follows that $c_{I1} = 0$, and therefore c_{I2}, c_{I4}, and c_{I6} are all zeros. ■

Proof of Theorem 3. First we check two conditions for consistency of the QMLE in two steps.

Step 1: Uniform convergence of the log quasi-likelihood function. All terms in the log quasi-likelihood function in Appendix A.1 can be expressed in the general terms M_n in Proposition 1. The pointwise convergence is straightforward. Since all parameters are bounded and they enter the log quasi-likelihood function polynomially except for the term $\ln|S_n(\lambda)|$, we only need to show the stochastic equicontinuity of $\frac{1}{n}\ln|S_n(\lambda)|$ to have the uniform convergence. Applying the mean value theorem,

$$\begin{aligned} &\left|\frac{1}{n}(\ln|S_n(\lambda_1)| - \ln|S_n(\lambda_2)|)\right| \\ &= \left|(\lambda_2 - \lambda_1)\frac{1}{n}\mathrm{tr}(G_n(\bar{\lambda}))\right| \leq |\lambda_2 - \lambda_1| C, \end{aligned} \tag{C.4}$$

where $\bar{\lambda}$ is between λ_1 and λ_2 and C is a constant not depending on n. The inequality is implied by $\sup_\lambda \|G_n(\lambda)\|_\infty < \infty$. From this, we have $\sup_{\theta\in\Theta} |\frac{1}{n}\ln L_n(\theta) - E[\frac{1}{n}\ln L_n(\theta)]| \xrightarrow{p} 0$.

Step 2: Uniform equicontinuity of $\lim_{n\to\infty} E\left(\frac{1}{n}\ln L_n(\theta)\right)$. By inequality in (C.4), variance parameters being bounded away from zero in compact parameter spaces, and earlier result $\frac{1}{n}E|\varsigma^{*\prime}_n M_n \varsigma^*_n| = O(1)$, we have that $E\left(\frac{1}{n}\ln L_n(\theta)\right)$ is uniformly equicontinuous in $\theta \in \Theta$.

As θ_0 is the unique maximizer of $\lim_{n\to\infty} E[\frac{1}{n}\ln L_n(\theta)]$ from Claim C.3.1, these together imply $\widehat{\theta} \xrightarrow{p} \theta_0$.

Next we show the asymptotic normality of $\widehat{\theta}$. The second derivatives in Appendix A.2 can be written in the general form in Corollary 1, so we have the uniform convergence that $\sup_{\theta\in\Theta} \frac{1}{n}\left\|\frac{\partial^2 \ln L_n(\theta)}{\partial\theta\partial\theta'} - E\left(\frac{\partial^2 \ln(L_n(\theta))}{\partial\theta\partial\theta'}\right)\right\| \xrightarrow{p} 0$. Applying the CLT in Proposition 2 to $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta}$ in Appendix A.2, we have

$$\begin{aligned} \sqrt{n}(\widehat{\theta} - \theta_0) &= -\left(\frac{1}{n}\frac{\partial^2 \ln L_n(\widetilde{\theta})}{\partial\theta\partial\theta'}\right)^{-1} \frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta} \\ &= -\left[E\left(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right)\right]^{-1} \\ &\quad \times \frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta} + o_p(1) \\ &\xrightarrow{d} N\left(0, \left(\lim_{n\to\infty}\frac{1}{n}E\left(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right)\right)^{-1}\right. \\ &\quad \times \lim_{n\to\infty}\frac{1}{n}E\left(\frac{\partial \ln L_n(\theta_0)}{\partial\theta}\frac{\partial \ln L_n(\theta_0)}{\partial\theta'}\right) \\ &\quad \left.\times \left(\lim_{n\to\infty}\frac{1}{n}E\left(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right)\right)^{-1}\right). \quad ■ \end{aligned}$$

Proof of Theorem 4. As $\xi_n(\theta^G) = (\lambda_0 - \lambda)G_n(X_{1n}\beta_0 + \varepsilon_n\delta_0) + X_{1n}(\beta_0 - \beta) - X_{2n}(\Gamma_0 - \Gamma)\delta + \varepsilon_n(\delta_0 - \delta) + [I_n - (\lambda - \lambda_0)G_n]\xi_n$, we have $\xi_n(\theta^G) = M_n\varsigma^*_n b_1(\theta^G_0 - \theta^G) + X_{2n}(\Gamma_0 - \Gamma)(\delta_0 - \delta) + \xi_n$, where ς^*_n and M_n are expressed as in Propositions 1 and 2. Therefore

$$\begin{aligned} \frac{1}{n}\xi'_n(\theta^G)Q_n &\xrightarrow{p} (\theta^G_0 - \theta^G)' \lim_{n\to\infty}\frac{1}{n}E(b'_1\varsigma^{*\prime}_n M'_n Q_n) \\ &+ [(\Gamma_0 - \Gamma)(\delta_0 - \delta)]' \lim_{n\to\infty}\frac{1}{n}X'_{2n}Q_n. \end{aligned}$$

For $P_{jn} = M_{jn} - \frac{1}{n}\mathrm{tr}(M_{jn})I_n$,

$$\begin{aligned} \xi'_n(\theta^G)P_{jn}\xi_n(\theta^G) &= (\theta^G_0 - \theta^G)' b'_1 \varsigma^{*\prime}_n M'_n P_{jn} M_n \varsigma^*_n b_1(\theta^G_0 - \theta^G) \\ &+ 2(\theta^G_0 - \theta^G)' b'_1 \varsigma^{*\prime}_n M'_n P_{jn}\xi_n + \xi'_n P_{jn}\xi_n \\ &+ [(\Gamma_0 - \Gamma)(\delta_0 - \delta)]' X'_{2n} P_{jn}[X_{2n}(\Gamma_0 - \Gamma)(\delta_0 - \delta) \\ &+ M_n\varsigma^*_n b_1(\theta^G_0 - \theta^G) + \xi_n]. \end{aligned}$$

Proposition 1 implies that $\frac{1}{n}\xi_n' P_{jn}\xi_n \xrightarrow{p} \lim_{n\to\infty}\frac{1}{n}E(\xi_n' M_{jn}\xi_n) - \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}[E(M_{jn})]\lim_{n\to\infty}\frac{1}{n}E(\xi_n'\xi_n) = 0$ and $\frac{1}{n}\varsigma_n^{*\prime}M_n' P_{jn}M_n\varsigma_n^* \xrightarrow{p} \lim_{n\to\infty}\frac{1}{n}E(\varsigma_n^{*\prime}M_n' M_{jn}M_n\varsigma_n^*) - \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}[E(M_{jn})]\lim_{n\to\infty}\frac{1}{n}E(\varsigma_n^{*\prime}M_n' M_n\varsigma_n^*)$. Therefore,

$$\begin{aligned}
\frac{1}{n}\xi_n'(\theta^G)P_{jn}\xi_n(\theta^G) &\xrightarrow{p} 2(\theta_0^G-\theta^G)' b_1' \\
&\times\left(\lim_{n\to\infty}\frac{1}{n}E(\varsigma_n^{*\prime}M_n^{*\prime}M_{jn}\xi_n) - \lim_{n\to\infty}\frac{1}{n}E(\varsigma_n^{*\prime}M_n^{*\prime}\xi_n)\lim_{n\to\infty}\frac{1}{n}E(M_{jn})\right) \\
&+(\theta_0^G-\theta^G)' b_1'\left(\lim_{n\to\infty}\frac{1}{n}E(\varsigma_n^{*\prime}M_n^{*\prime}M_{jn}M_n^*\varsigma_n^*) - \lim_{n\to\infty}\frac{1}{n}E(\varsigma_n^{*\prime}M_n^{*\prime}M_n^*\varsigma_n^*)\lim_{n\to\infty}\frac{1}{n}\mathrm{tr}E(M_{jn})\right) b_1(\theta_0^G-\theta^G) \\
&+[(\Gamma_0-\Gamma)(\delta_0-\delta)]'\left(\lim_{n\to\infty}\frac{1}{n}X_{2n}'E(P_{jn})X_{2n}\right) \\
&\times(\Gamma_0-\Gamma)(\delta_0-\delta)+[(\Gamma_0-\Gamma)(\delta_0-\delta)]' \\
&\times\left[\lim_{n\to\infty}\frac{1}{n}E(X_{2n}'M_{jn}M_n\varsigma_n^*) - \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}E(M_{jn})\lim_{n\to\infty}\frac{1}{n}E(X_{2n}'M_n\varsigma_n^*)\right] b_1(\theta_0^G-\theta^G).
\end{aligned}$$

From these moments, we see $\frac{1}{n}g_n(\theta^G)\xrightarrow{p} g(\theta^G)$ with $g(\theta_0^G)=0$. As all parameters in θ^G enter $g_n(\theta^G)$ polynomially, pointwise convergence gives the uniform convergence that $\sup_{\theta^G}\frac{1}{n}\left\|a_n g_n(\theta^G) - a_n g(\theta^G)\right\|\xrightarrow{p} 0$. With the identification conditions from Assumption 7, the consistency of GMM $\widehat{\theta}_n^G$ follows.

For the asymptotic distribution of $\widehat{\theta}_n^G$, by Taylor's expansion of $\frac{\partial g_n'(\widehat{\theta}_n^G)}{\partial\theta^G}a_n' a_n g_n(\widehat{\theta}_n^G)=0$ at θ_0^G,

$$\sqrt{n}(\widehat{\theta}_n^G-\theta_0^G) = -\left(\frac{1}{n}\frac{\partial g_n'(\widehat{\theta}_n^G)}{\partial\theta^G}a_n' a_n\frac{1}{n}\frac{\partial g_n'(\overline{\theta}_n^G)}{\partial\theta^G}\right)^{-1}\times\frac{1}{n}\frac{\partial g_n'(\widehat{\theta}_n^G)}{\partial\theta^G}a_n'\frac{1}{\sqrt{n}}a_n g_n(\theta_0^G),$$

where $\overline{\theta}_n^G$ is between $\widehat{\theta}_n^G$ and θ_0^G. Denote $A^s = A+A'$ as the sum of A and its transpose, then $\frac{1}{n}\frac{\partial g_n(\theta^G)}{\partial\theta^{G\prime}}$ is given in Box II. It is easy to check $\sup_{\theta^G}\frac{1}{n}\left\|\frac{\partial g_n(\theta^G)}{\partial\theta^G} - \left(\frac{\partial g(\theta^G)}{\partial\theta^G}\right)\right\|\xrightarrow{p} 0$. Thus $\sqrt{n}(\widehat{\theta}_n^G-\theta_0^G) = -\left(D_n' a_n' a_n D_n\right)^{-1}D_n' a_n'\frac{1}{\sqrt{n}}a_n g_n(\theta_0^G)+o_p(1)$. As $\frac{1}{\sqrt{n}}g_n(\theta_0^G)$ involves $\frac{1}{\sqrt{n}}X_n'\varepsilon_n$, $\frac{1}{\sqrt{n}}Q_n'\xi_n$, and

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\xi_n' M_{jn}\xi_n - \frac{1}{\sqrt{n}}\xi_n'\xi_n\frac{\mathrm{tr}(M_{jn})}{n} \\
&= \frac{1}{\sqrt{n}}\xi_n' M_{jn}\xi_n - \frac{1}{\sqrt{n}}\xi_n'\xi_n\frac{E[\mathrm{tr}(M_{jn})]}{n}+o_p(1),
\end{aligned}$$

the asymptotic distribution of $\sqrt{n}(\widehat{\theta}_n^G-\theta_0^G)$ is of the form of R_n in Proposition 2. Under Assumptions 1–4, and 7, $\sqrt{n}(\widehat{\theta}_n^G-\theta_0^G)\xrightarrow{d} N(0,\Sigma_{\mathrm{GMM}})$.

Now we give the expressions of D_n and $\Omega(\theta_0^G)$. Eq. (C.5) is given in Box III, where $g_{j\lambda} = \sigma_{\xi0}^2\left(\lim_{n\to\infty}\frac{1}{n}E[\mathrm{tr}(M_{jn}^s G_n)] - \lim_{n\to\infty}\frac{1}{n}E[\mathrm{tr}(M_{jn}^s)]\lim_{n\to\infty}\frac{1}{n}E[\mathrm{tr}(G_n)]\right)$ for $j=1,\ldots,m$, because

$$\begin{aligned}
\frac{1}{n}\xi_n' P_{jn}^s W_n Y_n &= \frac{1}{n}\xi_n' P_{jn}^s G_n\xi_n + \frac{1}{n}[\xi_n' P_{jn}^s G_n(X_{1n}\beta_0+\varepsilon_n\delta_0)] \\
&\xrightarrow{p} \lim_{n\to\infty}\frac{1}{n}\xi_n' P_{jn}^s G_n\xi_n = \lim_{n\to\infty}\frac{1}{n}\xi_n' M_{jn}^s G_n\xi_n - \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(M_{jn}^s)\frac{1}{n}\xi_n' G_n\xi_n \\
&= \sigma_{\xi0}^2\left(\lim_{n\to\infty}\frac{1}{n}E[\mathrm{tr}(M_{jn}^s G_n)] - \lim_{n\to\infty}\frac{1}{n}E[\mathrm{tr}(M_{jn}^s)]\lim_{n\to\infty}\frac{1}{n}E[\mathrm{tr}(G_n)]\right).
\end{aligned}$$

For the variance, Eq. (C.6) is given in Box IV, where $\Omega_{jk} = \mathrm{Var}(\xi_n' P_{jn}\xi_n\xi_n' P_{kn}\xi_n) = \sum_{i=1}^n E[(\xi_{i,n}^4 - 3\sigma_{\xi0}^4)P_{jn}(i,i)P_{kn}(i,i)] + \sigma_{\xi0}^4\mathrm{tr}(P_{jn}P_{kn}^s)$ for $j,k=1,\ldots,m$. ∎

Proof of Claim 1. As $\widehat{\xi}_n = S_n(\widehat{\lambda})Y_n - X_{1n}\widehat{\beta} - (Z_n - X_{2n}\widehat{\Gamma})\widehat{\delta} = \widehat{\xi}_n^* + X_{2n}(\Gamma_0-\widehat{\Gamma})(\delta_0-\widehat{\delta})$, where

$$\begin{aligned}
\widehat{\xi}_n^* &= S_n(\widehat{\lambda})Y_n - X_{1n}\widehat{\beta} - (Z_n - X_{2n}\widehat{\Gamma})\delta_0 + \varepsilon_n(\delta_0-\widehat{\delta}) \\
&= (\lambda_0-\widehat{\lambda})G_n(X_{1n}\beta_0+\varepsilon_n\delta_0) + X_{1n}(\beta_0-\widehat{\beta}) \\
&\quad - X_{2n}(\Gamma_0-\widehat{\Gamma})\delta_0 + \varepsilon_n(\delta_0-\widehat{\delta}) - (\widehat{\lambda}-\lambda_0)G_n\xi_n + \xi_n,
\end{aligned}$$

we can express $\frac{1}{n}\widehat{\xi}_n^{*\prime}\widehat{\xi}_n^*$ in the following form as in Proposition 1 so that:

$$\begin{aligned}
\frac{1}{n}\widehat{\xi}_n^{*\prime}\widehat{\xi}_n^* &= (\theta_0-\widehat{\theta})'\frac{1}{n}a_1\varsigma_n^{*\prime}M_n\varsigma_n^* b_1(\theta_0-\widehat{\theta}) \\
&\quad + \frac{1}{n}a_2\varsigma_n^{*\prime}M_n\varsigma_n^* b_2(\theta_0-\widehat{\theta}) + \frac{1}{n}\xi_n'\xi_n \xrightarrow{p} \sigma_{\xi0}^2.
\end{aligned}$$

Similarly, $\frac{1}{n}\widehat{\xi}_n^* X_{2n}(\Gamma_0-\widehat{\Gamma})(\delta_0-\widehat{\delta}) = o_p(1)$ and $\frac{1}{n}(\delta_0-\widehat{\delta})'(\Gamma_0-\widehat{\Gamma})' X_{2n}' X_{2n}(\Gamma_0-\widehat{\Gamma})(\delta_0-\widehat{\delta}) = o_p(1)$. Thus, $\frac{1}{n}\widehat{\xi}_n'\widehat{\xi}_n \xrightarrow{p} \sigma_{\xi0}^2$.

Terms in Σ_{IV} and Σ_{BGIV} have some common features, but the most complicated term we need to show is

$$\frac{1}{n}[a'\widehat{\varepsilon}_n' G_n(\widehat{\lambda})' G_n(\widehat{\lambda})\widehat{\varepsilon}_n b - E(a'\varepsilon_n' G_n' G_n\varepsilon_n b)] = o_p(1).$$

As $\widehat{\varepsilon}_n = \varepsilon_n + X_{2n}(\Gamma_0-\widehat{\Gamma})$, we have $\frac{1}{n}[a'\widehat{\varepsilon}_n' G_n(\widehat{\lambda})' G_n(\widehat{\lambda})\widehat{\varepsilon}_n b - E(a'\widehat{\varepsilon}_n' G_n(\widehat{\lambda})' G_n(\widehat{\lambda})\widehat{\varepsilon}_n b)] = o_p(1)$[14] from the ULLN in Corollary 1 and $\frac{1}{n}[E(a'\widehat{\varepsilon}_n' G_n(\widehat{\lambda})' G_n(\widehat{\lambda})\widehat{\varepsilon}_n b) - E(a'\varepsilon_n' G_n' G_n\varepsilon_n b)] = o_p(1)$ from the equicontinuity of $\frac{1}{n}E[a'\varepsilon_n'(\theta)G_n(\lambda)' G_n(\lambda)\varepsilon_n'(\theta)b]$. These together complete the proof of $\frac{1}{n}[\widehat{\varepsilon}_n' G_n(\widehat{\lambda})' G_n(\widehat{\lambda})\widehat{\varepsilon}_n - E(\varepsilon_n' G_n' G_n\varepsilon_n)] = o_p(1)$. ∎

Proof of Claim 2. Consider the moments in Σ_{QMLE} and Σ_{GMM}. The most complicated term we need to show is

$$\frac{1}{n}\sum_{i=1}^n\widehat{\xi}_{i,n}^3 G_{ii,n}(\widehat{\lambda})G_{i,n}(\widehat{\lambda})\widehat{\varepsilon}_n b - \frac{1}{n}\sum_{i=1}^n E[\xi_{i,n}^3 G_{ii,n}G_{i,n}\varepsilon_n b] = o_p(1).$$

As we can express $\widehat{\varepsilon}_n = \varepsilon_n + X_{2n}(\Gamma_0-\widehat{\Gamma})$ and

$$\begin{aligned}
\widehat{\xi}_n &= (\lambda_0-\widehat{\lambda})G_n(X_{1n}\beta_0+\varepsilon_n\delta_0+\xi_n) \\
&\quad + X_{1n}(\beta_0-\widehat{\beta}) - X_{2n}(\Gamma_0-\widehat{\Gamma})\widehat{\delta} + \varepsilon_n(\delta_0-\widehat{\delta}) + \xi_n \\
&= M_{1n}\varsigma_n^* b_1(\theta_0^*-\widehat{\theta}^*) + \xi_n
\end{aligned}$$

with $\theta^* = (\theta', \delta'\Gamma')'$, it is sufficient to show

$$\frac{1}{n}\sum_{i=1}^n[e_{i,n}' M_{1n}\varsigma_n^* b_1(\theta_0^*-\widehat{\theta}^*)]^3 G_{ii,n}(\widehat{\lambda})G_{i,n}(\widehat{\lambda})\varsigma_n^* b_2 = o_p(1); \tag{C.7}$$

$$\frac{1}{n}\sum_{i=1}^n[e_{i,n}' M_{1n}\varsigma_n^* b_1(\theta_0^*-\widehat{\theta}^*)]^2\xi_{i,n}$$

[14] The expectation is with respect to ε_n only but not with respect to estimated parameters, such as $\widehat{\lambda}$. The expectation function is then evaluated at the estimated parameters.

$$\frac{1}{n}\frac{\partial g_n(\theta^G)}{\partial \theta^{G\prime}} = \frac{1}{n}\begin{pmatrix} -\xi_n'(\theta^G)P_{1n}^sW_nY_n & -\xi_n'(\theta^G)P_{1n}^sX_{1n} & -\delta'\otimes[\xi_n'(\theta^G)P_{1n}^sX_{2n}] & \xi_n'(\theta^G)P_{1n}^s(Z_n - X_{2n}\Gamma) \\ \vdots & \vdots & \vdots & \vdots \\ -\xi_n'(\theta^G)P_{mn}^sW_nY_n & -\xi_n'(\theta^G)P_{mn}^sX_{1n} & -\delta'\otimes[\xi_n'(\theta^G)P_{mn}^sX_{2n}] & \xi_n'(\theta^G)P_{mn}^s(Z_n - X_{2n}\Gamma) \\ -Q_n'W_nY_n & -Q_n'X_{1n} & \delta'\otimes(Q_n'X_{2n}) & -Q_n'(Z_n - X_{2n}\Gamma) \\ 0 & 0 & -I_{p_2}\otimes(X_n'X_{2n}) & 0 \end{pmatrix}$$

Box II.

$$\begin{aligned} D_n &= -\operatorname*{plim}_{n\to\infty}\frac{1}{n}\frac{\partial(g_n(\theta_0^G))}{\partial\theta^{G\prime}} \\ &= \lim_{n\to\infty}\begin{pmatrix} g_{1\lambda} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ g_{m\lambda} & 0 & 0 & 0 \\ \frac{1}{n}E[Q_n'G_n(X_{1n}\beta_0+\varepsilon_n\delta_0)] & E\left(\frac{Q_n'X_{1n}}{n}\right) & -\delta'\otimes E\left(\frac{Q_n'X_{2n}}{n}\right) & E\left(\frac{Q_n'\varepsilon_n}{n}\right) \\ 0 & 0 & I_{p_2}\otimes\left(\frac{X_n'X_{2n}}{n}\right) & 0 \end{pmatrix} \end{aligned} \tag{C.5}$$

Box III.

$$\Omega(\theta_0^G) = \mathrm{Var}(g_n(\theta_0^G)) = \lim_{n\to\infty}\frac{1}{n}\begin{pmatrix} \Omega_{11} & \cdots & \Omega_{1m} & \sum_{i=1}^n E[(\xi_{i,n}^3P_{1n}(i,i)Q_{i.,n})] & 0 \\ \Omega_{21} & \cdots & \Omega_{2m} & \sum_{i=1}^n E[(\xi_{i,n}^3P_{2n}(i,i)Q_{i.,n})] & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ \Omega_{m1} & \cdots & \Omega_{mm} & \sum_{i=1}^n E[(\xi_{i,n}^3P_{mn}(i,i)Q_{i.,n})] & 0 \\ 0 & \cdots & 0 & \sigma_{\xi 0}^2Q_n'Q_n & 0 \\ 0 & \cdots & 0 & 0 & \Sigma_{\varepsilon 0}\otimes(X_n'X_n) \end{pmatrix} \tag{C.6}$$

Box IV.

$$\times G_{ii,n}(\widehat{\lambda})G_{i,n}(\widehat{\lambda})\varsigma_n^*b_2 = o_p(1); \tag{C.8}$$

$$\frac{1}{n}\sum_{i=1}^n e_{i,n}'M_{1n}\varsigma_n^*b_1(\theta_0^*-\widehat{\theta}^*)\xi_{i,n}^2$$
$$\times G_{ii,n}(\widehat{\lambda})G_{i,n}(\widehat{\lambda})\varsigma_n^*b_2 = o_p(1); \tag{C.9}$$

$$\frac{1}{n}\sum_{i=1}^n \xi_{i,n}^3G_{ii,n}(\widehat{\lambda})G_{i,n}(\widehat{\lambda})\varepsilon_nb$$
$$-\frac{1}{n}\sum_{i=1}^n E[\xi_{i,n}^3G_{ii,n}G_{i,n}\varepsilon_nb] = o_p(1). \tag{C.10}$$

Eqs. (C.7), (C.8), and (C.9) have some common features, so we will show (C.7) as an example. As $\sup_{i,n}\sup_{\lambda\in\Lambda}|G_{ii,n}(\lambda)| = O(1)$ and $\theta_0^*-\widehat{\theta}^* = o_p(1)$, we only need to show

$$\sup_{\lambda\in\Lambda}\frac{1}{n}\left|\sum_{i=1}^n(e_{i,n}'M_{1n}\varsigma_n^*a_1)^3G_{i,n}(\lambda)\varsigma_n^*b_2\right| = O_p(1).$$

It is sufficient to show

$$E\left|\sup_{\lambda\in\Lambda}\frac{1}{n}\sum_{i=1}^n(e_{i,n}'M_{1n}\varsigma_n^*a_1)^3G_{i,n}(\lambda)\varsigma_n^*b_2\right|$$
$$\le \sup_{i,n}E\left||e_{i,n}'M_{1n}\varsigma_n^*a_1|^3\sup_{\lambda\in\Lambda}|G_{i,n}(\lambda)\varsigma_n^*b_2|\right|$$
$$\le (\sup_{i,n}\|e_{i,n}'M_{1n}\varsigma_n^*a_1\|_4)^3\sup_{i,n}\|\sup_{\lambda\in\Lambda}|G_{i,n}(\lambda)\varsigma_n^*b_2|\,\|_4 = O(1).$$

The second inequality is from Hölder's inequality. For the equality, $\sup_{i,n}\|e_{i,n}'M_{1n}\varsigma_n^*a_1\|_4 = O(1)$ is directly from Claims C.1.6, C.1.7, C.2.5, and C.2.6, so we need to show $\sup_{i,n}\|\sup_{\lambda\in\Lambda}|G_{i,n}(\lambda)\varsigma_n^*b_2|\,\|_4 = O(1)$. As $|G_{i,n}(\lambda)\varsigma_n^*b_2| = |\sum_{l=0}^\infty\lambda^lW_{i,n}^{l+1}\varsigma_n^*b_2| \le \sum_{l=0}^\infty|\lambda|^l|W_{i,n}^{l+1}\varsigma_n^*b_2|$,

$$\|\sup_{\lambda\in\Lambda}|G_{i,n}(\lambda)\varsigma_n^*b_2|\,\|_4 \le \left\|\sup_{\lambda\in\Lambda}\sum_{l=0}^\infty|\lambda|^l|W_{i,n}^{l+1}\varsigma_n^*b_2|\right\|_4$$
$$\le \sup_{\lambda\in\Lambda}\sum_{l=0}^\infty|\lambda|^l\|t_{i,n}(l+1)\|_4.$$

As $\|t_{i,n}(m)\|_p \le m^{c_3d_0+2}c_w^mC_{ap}$ under Assumption (4.1) and $\|t_{i,n}(m)\|_p \le C_{ap}c_w^mm^{d_0}$ under Assumption (4.2) from Claims C.1.6 and C.2.5, together with $\sup_{\lambda\in\Lambda}|\lambda|c_w < 1$ from Assumption (3.2), we have $\|\sup_{\lambda\in\Lambda}|G_{i,n}(\lambda)\varsigma_n^*b_2|\,\|_4 < C$, where C does not depend on i or n. Therefore, $\sup_{i,n}\|\sup_{\lambda\in\Lambda}|G_{i,n}(\lambda)\varsigma_n^*b_2|\,\|_4 = O(1)$.

To show Eq. (C.10), using similar arguments as those in Corollary 1, Claims C.1.6 and C.2.5, we have the uniform convergence that

$$\sup_{\lambda\in\Lambda}\frac{1}{n}\left|\sum_{i=1}^n\xi_{i,n}^3G_{ii,n}(\lambda)G_{i,n}(\lambda)\varepsilon_nb\right.$$

$$- \sum_{i=1}^{n} E[\xi_{i,n}^3 G_{ii,n}(\lambda) G_{i,n}(\lambda)\varepsilon_n b] \Bigg| = o_p(1)$$

and by the equicontinuity of $\frac{1}{n}\sum_{i=1}^{n} E[\xi_{i,n}^3 G_{ii,n}(\lambda) G_{i,n}(\lambda)\varepsilon_n b]$,

$$\frac{1}{n}\sum_{i=1}^{n} E[\xi_{i,n}^3 G_{ii,n}(\widehat{\lambda}) G_{i,n}(\widehat{\lambda})\varepsilon_n b] - \frac{1}{n}\sum_{i=1}^{n} E[\xi_{i,n}^3 G_{ii,n} G_{i,n}\varepsilon_n b] = o_p(1).$$

Thus Eq. (C.10) is proved.

These together complete the proof

$$\frac{1}{n}\sum_{i=1}^{n} \widehat{\xi}_{i,n}^3 G_{ii,n}(\widehat{\lambda}) G_n(\widehat{\lambda})\widehat{\varepsilon}_n b - \frac{1}{n}\sum_{i=1}^{n} E[\xi_{i,n}^3 G_{ii,n} G_n \varepsilon_n b] = o_p(1).$$

Similarly, we can show $\frac{1}{n}\sum_{i=1}^{n} \widehat{\xi}_{i,n}^4 G_{ii,n}(\widehat{\lambda}) - \frac{1}{n}\sum_{i=1}^{n} E(\xi_{i,n}^4 G_{ii,n}) = o_p(1)$. Therefore, if we replace θ_0 with a consistent estimator $\widehat{\theta}$, ε_n with $\widehat{\varepsilon}_n = Z_n - X_{2n}\widehat{\Gamma}$, and ξ_{in} with $\widehat{\xi}_{in}$, then we have consistent estimators of Σ_{QMLE} and Σ_{GMM}. ■

References

Anselin, L., 1980. Estimation Methods for Spatial Autoregressive Structures. In: Regional Science Dissertation and Monograph Series, Cornell University, Ithaca, NY.

Anselin, L., Bera, A., 1997. Spatial dependence in linear regression models with an introduction to spatial econometrics. J. Public Econ. 52, 285–307.

Case, A., Rosen, H., Hines, J., 1993. Budget spillovers and fiscal policy interdependence: evidence from the states. In: Giles, D., Ullah, A. (Eds.), Handbook of Applied Economic Statistics. Marcel Dekker, NY.

Conway, K., Rork, J., 2004. Diagnosis murder: the death of state death taxes. Econ. Inquiry 42, 537–559.

Crabbé, K., Vandenbussche, H., 2008. Spatial tax competition in the EU15. Working Paper, Catholic University Leuven.

Hsieh, C., Lee, L., 2011. A social interactions model with endogenous friendship. Working Paper, The Ohio State University.

Jenish, N., Prucha, I., 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. J. Econometrics 150, 86–98.

Jenish, N., Prucha, I., 2012. On spatial processes and asymptotic inference under near-epoch dependence. J. Econometrics 170, 178–190.

Kelejian, H., Piras, G., 2014. Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. Reg. Sci. Urban Econ. 46, 140–149.

Kelejian, H., Prucha, I., 1998. A generalized spatial two stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances. J. Real Estate Finance Econ. 17, 99–121.

Kelejian, H., Prucha, I., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. Internat. Econom. Rev. 40, 509–533.

Lee, L., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.

Lee, L., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. J. Econometrics 137, 489–514.

Lee, L., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometric Theory 26, 187–230.

Lin, X., Lee, L., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. J. Econometrics 157, 34–52.

Liu, X., Lee, L., Bollinger, C., 2010. Improved efficient quasi-maximum likelihood estimator of spatial autoregressive models. J. Econometrics 159, 303–319.

Ord, J., 1975. Estimation methods for models of spatial interaction. J. Amer. Statist. Assoc. 70, 120–126.

Pinkse, J., Slade, M., 2010. The future of spatial econometrics. J. Reg. Sci. 50, 103–117.

Qu, X., Lee, L., 2012. LM tests for spatial correlation in spatial models with limited dependent variables. Reg. Sci. Urban Econ. 42, 430–445.

White, H., 1982. Maximum likelihood estimation in misspecified models. Econometrica 50, 1–26.