Near Unit Root in the Spatial Autoregressive Model

Lung-fei Lee*
Department of Economics
The Ohio State University

Jihai Yu
Department of Economics
University of Kentucky

December 20, 2007

Abstract

This paper studies the spatial autoregressive (SAR) model for cross sectional data when the true spatial effect parameter is near unity. We decompose the data generating process (DGP) into an unstable component and a stable component and then establish asymptotic properties of QMLE, 2SLSE and linearized QMLE. The spatial effect estimator has a higher rate of convergence and other parameters have the regular $\sqrt{n}$ rate. The higher rate of convergence reflects how fast the spatial root converges to unity. In contrast to near unit root time series, the estimators are all asymptotically normal. Similarly to the regular SAR model, QMLE and linearized QMLE are more efficient than 2SLSE.

JEL classification: C13; C23

Keywords: Spatial autoregressive model, Spatial unit root, Near unit root, Two stage least square, Quasi-maximum likelihood estimation

*Lee appreciates having financial support for his research from NSF grant SES-0519204.

1 Introduction

Spatial autoregressive (SAR) models study autocorrelation among cross sectional units and have wide applications[1]. To estimate the SAR model, Kelejian and Prucha (1998) proposed a two stage least squares (2SLS) method which uses instrumental variables (IVs) constructed from exogenous variables and the spatial weights matrix . Lee (2003) chose some specific IVs and got the best two stage least square (B2SLS) estimators. Lee (2004) studies the asymptotic properties of the maximum likelihood (ML) and quasi-maximum likelihood (QML) estimators of the SAR model. For the SAR model, the spatial correlation can provide nonlinear moment conditions in addition to linear moments of IV's in the general method of moment (GMM) setting. Lee (2007) established asymptotic properties of GMM estimators, which can be more efficient than the 2SLS estimators. The best GMM can be as efficient as the ML estimators (MLEs) when the true disturbances are normal. Liu et al. (2006) showed that carefully designed linear and quadratic moment functions can generate a GMM estimator which is more efficient relative to the QML estimate when the disturbances are not normal.

For the above methods, there are restrictions on the spatial weights matrix and the spatial effect coefficient such that the spatial dependence across units are controlled to a limited degree. With a spatial weights matrix being row-normalized, the spatial effect can not be equal or larger than one (Ord 1975). If the true spatial effect is near unity, the spatial dependence across units will be very strong and the variances of the dependent variables can become very large even though we assume row sum and column sum boundedness of the spatial weights matrix. In this paper, we derive the asymptotic properties of the IV and QML estimators[2] when the true spatial effect can go to unity at any rate. We call this a "near unit root" case because a SAR model has sometimes been regarded as a generalization of an autoregressive model in time series to the spatial setting.

The near unit root model in time series has been extensively used in the univariate literature to study the test statistics under local alternatives, starting with Chan and Wei (1987) and Phillips (1987, 1988). Also, the estimation of the autoregressive roots near unity are studied both in time series (Phillips et al. 2001) and in the panel data setting (Moon and Phillips 2000, 2004). For the near unit root in time series, the basic idea is to study an intermediate case between unit root and stationary processes. In the SAR model, there is a spatial weights matrix which represents the interactions of different spatial units. In practice, the spatial weighs matrix is usually row-normalized so that the influence of neighbors can be represented in terms of averages. The SAR model is an equilibrium model. The unit root case where the spatial coefficient is equal

[1] Early development in estimation and testing for cross section data can be found in Anselin (1988, 1992), Kelejian and Robinson (1993), Cressie (1993), Anselin and Florax (1995), Anselin and Rey (1997), and Anselin and Bera (1998), among others.

[2] We expect that similar features of the estimates from a GMM approach will hold.

to 1 is not allowed (otherwise, it will not be an equilibrium model). Hence, we can only study the near unit root case in SAR model by local alternative, but not the unit root case itself[3]. Therefore, in the near unit root for spatial data, the story may be different from time series. What's more, the nonlinearity nature in the reduced form equation of the SAR model distinguishes the near unit root case here from the time series counterpart, where only the linear regression is considered in typical settings.

The paper is organized as follows. Section 2 introduces the model and decomposes the DGP into a stable part and an unstable part when the true spatial effect is near unity. Section 3 establishes the identification, consistency and asymptotic distribution of QML estimates. Section 4 gives asymptotic properties of 2SLSE, and of linearized QMLE where the initial estimates from 2SLS are used. Monte Carlo results are provided in Section 5 to compare the finite sample performance of various estimators. Conclusion is made in Section 6 and proofs are provided in Appendices.

2 Model

Consider the cross sectional (first order) SAR model

$$Y_n = \lambda_{n0} W_n Y_n + X_n \beta_0 + \epsilon_n, \tag{1}$$

where Y_n is $n \times 1$ vector of dependent variables, X_n is an $n \times k$ nonstochastic exogenous variables, W_n is a nonstochastic spatial weights matrix and the disturbance ϵ_{ni}, $i = 1, 2, \cdots, n$, of the n-dimensional vector ϵ_n are *i.i.d.* $(0, \sigma_0^2)$. Here, $W_n Y_n$ is usually referred to as a spatial lag of Y_n.

Suppose that W_n is diagonalizable[4] with eigenvalues d_{in} such that either $d_{in} = 1$ or $|d_{in}| < 1$[5]. Furthermore, suppose that there are m_n unit eigenvalues, and the remaining $(n - m_n)$ eigenvalues are all bounded away from 1 in absolute value for all n. The near unit root case refers to the situation that for the true spatial effect parameter,

$$\lambda_{n0} = 1 - \frac{1}{\psi_n}, \tag{2}$$

where ψ_n goes to infinity as n goes to infinity. Thus, as n goes to infinity, λ_{n0} tends to 1. The ψ_n specifies, in its general form, how fast that λ_{n0} tends to unity as the sample size n increases[6]. Note that because the

[3] Fingleton (1999) has investigated implications of unit root on the SAR model via Monte Carlo simulation. In order to generate unit root, he introduces an unconnected central unit such that the spatial weights matrix has a zero row. In that setting, unit root can be present and the system is an equilibrium one. This setting is analogous to having an initial observation in the time seris autoregressive model. However, this is a rather limiting setting for spatial models.

[4] When W_n is row normalized from a symmetric matrix, W_n is diagonalizable. See Lemma A.1 in Yu et al., (2007).

[5] A weights matrix row normalized has real eigenvalues, with all its eigenvalues less than or equal to one in absolute value and its largest eigenvalue always 1 (see Ord (1975)).

[6] In the time series literature, a specific rate for ψ_n is specified; for example, in Phillips (1987), ψ_n is proportional to n, where n is the time series length. For our analysis, we do not need to have a specific rate in terms of powers of n and ψ_n can be an arbitrary rate in its generality, which needs not to be specified in empirical applications.

SAR model is specified to be an equilibrium model, it is not meaningful to assume that $\lambda_{n0} = 1$. With our specification of the near unit root case, as $\lambda_{n0} < 1$, this model is still an equilibrium model for any finite n. This is so because $(I_n - \lambda_{n0}W_n)$ is invertible. With W_n being diagonalizable, let R_n be the eigenvector matrix. Then, we have

$$W_n = R_n D_n R_n^{-1}, \tag{3}$$

where $D_n = diag\{\mathbf{1}_{m_n}, d_{m_n+1}, \cdots, d_n\}$ is the diagonal eigenvalue matrix and $\mathbf{1}_{m_n}$ is $1 \times m_n$ vector of ones. This implies that $(I_n - \lambda_{n0}W_n) = R_n(I_n - \lambda_{n0}D_n)R_n^{-1}$. As $|\lambda_{n0}| < 1$ and $|d_{in}| \leq 1$ for all i, $(I_n - \lambda_{n0}D_n)$ is invertible and, so is $(I_n - \lambda_{n0}W_n)$.

2.1 Decomposition of Y_n

For the eigenvalue matrix D_n, it can be decomposed into two parts as $D_n = J_n + \tilde{D}_n$ where $J_n = diag\{\mathbf{1}_{m_n}, 0, \cdots, 0\}$ and $\tilde{D}_n = diag\{0, \cdots, 0, d_{m_n+1}, \cdots, d_n\}$. The J_n consists of all the m_n unit eigenvalues and $\tilde{D}_n$ consists of all the eigenvalues with their absolute values less than one. Denote $S_n(\lambda) = I_n - \lambda W_n$, from (1), the equilibrium vector Y_n is

$$Y_n = S_n^{-1}(\lambda_{n0})(X_n\beta_0 + \epsilon_n). \tag{4}$$

For $S_n(\lambda) = I_n - \lambda W_n = R_n(I_n - \lambda D_n)R_n^{-1}$, we have $S_n^{-1}(\lambda) = R_n(I_n - \lambda D_n)^{-1}R_n^{-1}$. When $\lambda_{n0} = 1 - 1/\psi_n$, we have

$$I_n - \lambda_{n0}D_n = diag\{(1 - \lambda_{n0}), \cdots, (1 - \lambda_{n0}), (1 - \lambda_{n0}d_{m_n+1}), \cdots, (1 - \lambda_{n0}d_n)\},$$

and

$$(I_n - \lambda_{n0}D_n)^{-1} = \psi_n(\lambda_{n0}J_n + \frac{1}{\psi_n}(I_n - \lambda_{n0}\tilde{D}_n)^{-1}).$$

Hence,

$$S_n^{-1}(\lambda_{n0}) = \psi_n R_n \Gamma_n R_n^{-1} \tag{5}$$

and

$$W_n S_n^{-1}(\lambda_{n0}) = \psi_n W_n R_n \Gamma_n R_n^{-1} \tag{6}$$

where

$$\Gamma_n = \Gamma_n^u + \Gamma_n^s,\ \Gamma_n^u = \lambda_{n0}J_n,\ \Gamma_n^s = \frac{1}{\psi_n}(I_n - \lambda_{n0}\tilde{D}_n)^{-1}, \tag{7}$$

the eigenvector matrix R_n is determined by W_n from (3) and Γ_n depends on λ_{n0} and W_n. Hence, when λ_{n0} is near unity, $S_n(\lambda_{n0})$ is ill conditioned and its inverse has the large factor ψ_n.

Therefore, from (4) and (5), we have,

$$\begin{aligned} Y_n &= \psi_n R_n \Gamma_n R_n^{-1}(X_n\beta_0 + \epsilon_n) \\ &= \psi_n R_n \Gamma_n^u R_n^{-1}(X_n\beta_0 + \epsilon_n) + \psi_n R_n \Gamma_n^s R_n^{-1}(X_n\beta_0 + \epsilon_n). \end{aligned} \tag{8}$$

This equation is revealing in that the first term on the right hand side is an unstable component and the second term is a stable component. The implied variance of Y_n can be large because the first component has ψ_n, which is explosive. Also, because $W_n R_n = R_n D_n$ and $D_n J_n = J_n$, one has

$$W_n R_n \Gamma_n R_n^{-1} = W_n R_n(\Gamma_n^u + \Gamma_n^s) R_n^{-1} = R_n(\Gamma_n^u + D_n \Gamma_n^s) R_n^{-1}. \tag{9}$$

Comparing (8) and (9), we can see that for Y_n and $W_n Y_n$, they have the same unstable component such that $W_n Y_n^u = Y_n^u$, where $Y_n^u = \psi_n R_n \Gamma_n^u R_n^{-1}(X_n\beta_0 + \epsilon_n)$.

2.2 Assumptions of the Model

To analyze the model, we make the following assumptions.

Assumption 1 *$\lambda_{n0} = 1 - \frac{1}{\psi_n}$ where ψ_n is an increasing function of n.*

Assumption 2 *W_n is diagonalizable, row sum and column sum bounded[7], and has m_n unit roots, and the remaining eigenvalues d_{nj}, $j = m_n + 1, \cdots, n$, are bounded away from 1 in absolute value for all n and j.*

Assumption 3 *The disturbances $\{\epsilon_{ni}\}$, $i = 1, 2, ..., n$ are i.i.d with zero mean, variance σ_0^2 and $E|\epsilon_{ni}|^{4+\eta} < \infty$ for some $\eta > 0$.*

Assumption 4 *The elements of X_n are nonstochastic and bounded, uniformly in n, and $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular.*

Assumption 5 *$R_n(I_n - \lambda_{n0}\tilde{D}_n)^{-1} R_n^{-1}$ and $R_n J_n R_n^{-1}$ are uniformly bounded in row sum and column sum norms.*

Assumption 1 specifies that the true spatial effect is near unity and how close it is near unity. Assumption 2 says that W_n is diagonalizable so that we can decompose the process into a stable part and an unstable part under Assumption 1. As an example, a weights matrix row-normalized from a symmetric matrix is diagonalizable. All its eigenvalues are real, less than or equal to one in absolute value and its largest

[7] We say a (sequence of $n \times n$) matrix P_n is uniformly bounded in row and column sum norms if $\sup_{n\geq 1} \|P_n\|_\infty < \infty$ and $\sup_{n\geq 1} \|P_n\|_1 < \infty$, where $\|P_n\|_\infty \equiv \sup_{1\leq i\leq n} \sum_{j=1}^n |p_{ij,n}|$ is the row sum norm and $\|P_n\|_1 = \sup_{1\leq j\leq n} \sum_{i=1}^n |p_{ij,n}|$ is the column sum norm.

eigenvalue always 1 (see Ord (1975)). Assumption 3 and 4 are the standard assumptions. The higher than the fourth moment condition in Assumption 3 is needed to apply the central limit theorem in Kelejian and Prucha (2001). The nonstochastic X_n and its uniform boundedness are assumed for convenience. If X_n is stochastic, Assumption 4 can be replaced by some moment restrictions. Assumption 5 is to guarantee that the stable and unstable parts of $S_n^{-1}(\lambda_{n0})$ in (5), after being rescaled, are row sum and column sum bounded. In the standard case with the true λ_0 strictly less than one, an important assumption due to Kelejian and Prucha (1998) is that $S_n^{-1}(\lambda_0)$ is uniformly bounded in both row and column sum norms. For the near unit root case, $S_n^{-1}(\lambda_{n0})$ will not be uniformly bounded in either row or column sum norms due to the explosive factor ψ_n in (5) for the unstable component. By taking out the ψ_n factor, the remaining matrix will not grow with n. Assumption 5 maintains the remaining components to be spatially stable.

3 QMLE

3.1 Consistency

Denote $\theta = (\delta', \sigma^2)'$ where $\delta = (\beta', \lambda)'$. For (1), we can write the likelihood function as

$$\ln L_n(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\epsilon_n'(\delta)\epsilon_n(\delta), \tag{10}$$

where $\epsilon_n(\delta) = Y_n - \lambda W_n Y_n - X_n\beta$. As we might have different rates of convergence for different components of the vector of estimates, it is desirable to consider the concentrated likelihood function at λ. From (10), given λ, the QMLE of β is

$$\hat{\beta}_n(\lambda) = (X_n'X_n)^{-1}X_n'S_n(\lambda)Y_n \tag{11}$$

and the QMLE of σ^2 is

$$\begin{aligned}\hat{\sigma}_n^2(\lambda) &= \frac{1}{n}[S_n(\lambda)Y_n - X_n\hat{\beta}_n(\lambda)]'[S_n(\lambda)Y_n - X_n\hat{\beta}_n(\lambda)] \\ &= \frac{1}{n}Y_n'S_n'(\lambda)M_nS_n(\lambda)Y_n,\end{aligned} \tag{12}$$

where $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$. Hence, the concentrated likelihood function is

$$\ln L_n(\lambda) = -\frac{n}{2}(\ln(2\pi)+1) - \frac{n}{2}\ln\hat{\sigma}_n^2(\lambda) + \ln|S_n(\lambda)|. \tag{13}$$

The range of parameter values of λ is known to be the interval $(\frac{1}{d_{min,n}}, 1)$ in order that the determinant of $S_n(\lambda)$ is positive (Anselin 1988), where $d_{min,n} = \min\{d_{ni}|i = 1, \cdots, n\}$ is the smallest eigenvalue of W_n. Note that, for the case that W_n is row-normalized and has zero diagonal, the trace of W_n is zero and, hence, $d_{min,n} < 0$ and the largest eigenvalue is 1; also, $|d_{ni}| \leq 1$ for all i, which implies that d_{ni} are bounded from

both below and above. Hence, we will consider the parameter spaces of λ_{n0} as $\Lambda = (-1,\ 1)$. The lower bound is set to -1 for convenience which is often specified in the spatial literature. This is so because with a row-normalized spatial weights matrix, $|\lambda| < 1$ implies that $S_n^{-1}(\lambda)$ can be expanded as a power series of W_n (Horn and Johnson 1985). The QMLE $\hat{\lambda}_n$ maximizes (13) on the parameter space Λ and the QMLE of β and σ^2 are $\hat{\beta}_n(\hat{\lambda}_n)$ and $\hat{\sigma}_n^2(\hat{\lambda}_n)$.

For $\hat{\sigma}_n^2(\lambda)$, using $S_n(\lambda)Y_n = S_n(\lambda)S_n^{-1}(\lambda_{n0})(X_n\beta_0 + \epsilon_n)$ and $S_n(\lambda)S_n^{-1}(\lambda_{n0}) = I_n - (\lambda - \lambda_{n0})G_n$ where $G_n = W_nS_n^{-1}(\lambda_{n0})$, we have

$$\begin{aligned}\hat{\sigma}_n^2(\lambda) &= \frac{1}{n}(X_n\beta_0 + \epsilon_n)'(S_n(\lambda)S_n^{-1}(\lambda_{n0}))'M_nS_n(\lambda)S_n^{-1}(\lambda_{n0})(X_n\beta_0 + \epsilon_n) \\ &= \frac{1}{n}(\lambda - \lambda_{n0})^2(G_nX_n\beta_0)'M_nG_nX_n\beta_0 + \frac{1}{n}\epsilon_n'(S_n(\lambda)S_n^{-1}(\lambda_{n0}))'M_nS_n(\lambda)S_n^{-1}(\lambda_{n0})\epsilon_n \\ &\quad -2\frac{\lambda - \lambda_{n0}}{n}\epsilon_n'(S_n(\lambda)S_n^{-1}(\lambda_{n0}))'M_nG_nX_n\beta_0. \qquad (14)\end{aligned}$$

Correspondingly, we have

$$Q_n(\lambda) = \max_{\beta,\sigma^2} E\ln L_n(\theta) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\sigma_n^{*2}(\lambda) + \ln|S_n(\lambda)|, \qquad (15)$$

where

$$\sigma_n^{*2}(\lambda) = \frac{1}{n}(\lambda - \lambda_{n0})^2(G_nX_n\beta_0)'M_nG_nX_n\beta_0 + \frac{\sigma_0^2}{n}tr[(S_n(\lambda)S_n^{-1}(\lambda_{n0}))'S_n(\lambda)S_n^{-1}(\lambda_{n0})]. \qquad (16)$$

The presence of X_n in (1) is a distinctive feature of the mixed regressive SAR model. From (1) and (4), the reduced form equation of Y_n can be represented as

$$Y_n = \lambda_{n0}(G_nX_n\beta_0) + X_n\beta_0 + \epsilon_n.$$

It is noted that while the regressors in X_n are bounded, the generated regressor $G_nX_n\beta_0$ is explosive as it has the explosive factor ψ_n from (6). This has the implication that an estimate of λ_{n0} may have a higher rate of convergence than that of β_0.

For our analysis for QMLE, the following additional assumptions are made.

Assumption 6 *(1)* $\lim_{n\to\infty}\frac{m_n}{n} > 0$*; (2)* $\lim_{n\to\infty}\frac{1}{n}tr(R_n^{'-1}J_nR_n'R_nJ_nR_n^{-1}) > 0$*; (3) for any finite constant* c, $\lim_{n\to\infty}\frac{1}{n}tr[(I_n - cR_nJ_nR_n^{-1})'(I_n - cR_nJ_nR_n^{-1})] > 0$.

Assumption 7 $\beta_0'\lim_{n\to\infty}(\frac{1}{n}X_n'R_n^{'-1}J_nR_n'M_nR_nJ_nR_n^{-1}X_n)\beta_0 > 0$ *holds.*

Assumption 6 (1) and (2) specifies that we have a significant number[8] of unit roots in W_n. Assumption

[8] When $\frac{m_n}{n} \to 0$, the unstable component might be dominated by the stable component unless ψ_n is large enough. From (7), the contribution of the unstable component might depend on $\psi_n \cdot \frac{m_n}{n}$. Here, larger value of ψ_n is to increase the contribution of the unstable component and smaller value of $\frac{m_n}{n}$ has the effect of reducing the influence of the unstable component. As we are interested in the case where the stable component is dominated, we make the assumption that $\psi_n \to \infty$ and $\lim_{n\to\infty}\frac{m_n}{n} \neq 0$. For the case $\lim_{n\to\infty}\psi_n \cdot \frac{m_n}{n}$ is finite or zero, the rate of convergence for estimators would be the same as the regular case. The case with $\lim_{n\to\infty}\frac{m_n}{n} \neq 0$ is expected to provide the relevant approximation in practice with a finite sample with size n.

6 (3) might not be satisfied if $R_n J_n R_n^{-1}$ were approximately equal to I_n up to a scalar for large n. If $J_n = I_n$ for large enough n, the limit in Assumption 6 (3) would be zero by taking $c = 1$. As the nonzero elements of J_n are the diagonal elements consisting of unit eigenvalues of W_n, this assumption has implicitly ruled out the possibility that $\lim_{n\to\infty} \frac{m_n}{n} = 1$. This becomes clearer for the special case when W_n is symmetric. In that case, R_n is an orthonormal matrix and we have $\frac{1}{n} tr[(I_n - cR_n J_n R_n^{-1})'(I_n - cR_n J_n R_n^{-1})] = \frac{m_n}{n}(1-c)^2 + (1 - \frac{m_n}{n})$, which is positive when $0 < \frac{m_n}{n} < 1$ for any c. If $\frac{m_n}{n}$ goes to 1, the limit will be zero at $c = 1$. Assumption 7 is an identification condition, which specifies that for the generated spatial regressor $G_n X_n \beta_0$, its unstable component is not multicollinear with X_n.

Theorem 1 *Under Assumptions 1-5, 6 and 7, the QMLE $\hat{\lambda}_n$ is ψ_n-consistent such that $\psi_n(\hat{\lambda}_n - \lambda_{n0}) = o_p(1)$. Also, $\hat{\beta}_n - \beta_0 = o_p(1)$ and $\hat{\sigma}_n^2 - \sigma_0^2 = o_p(1)$.*

Proof. *See Appendix B.1.* ■

As ψ_n goes to infinity, the ψ_n-consistency of $\hat{\lambda}_n$ implies the consistency of $\hat{\lambda}_n$. Establishing ψ_n-consistency for $\hat{\lambda}_n$ is needed as the first step in order to derive the asymptotic distribution of $\hat{\lambda}_n$ and the consistency and asymptotic distribution of other estimates. As we shall see, the $\hat{\lambda}_n$ is not only ψ_n-consistent, it is $\sqrt{n}\psi_n$-consistent.

Our analysis in the preceding theorem relies on the identification condition in Assumption 7. This condition would obviously not be satisfied when $\beta_0 = 0$, i.e., no relevant exogenous variables in the model. More generally, Assumption 7 will fail if $R_n J_n R_n^{-1} X_n \beta_0$ is multicollinear with X_n in the limit. If this happens, the identification uniqueness condition can rely on the following assumption. This situation is corresponding to the model identification via the reduced form disturbances of Y_n, which forms a pure SAR process.

Assumption 8 *Either* $0 < \lim_{n\to\infty} \frac{m_n}{n} \leq 1/2$ *or* $\lim_{n\to\infty} \frac{m_n}{n} \leq \lim_{n\to\infty} \frac{1}{n} tr(R_n^{'-1} J_n R_n' R_n J_n R_n^{-1})$.

It is noted that when W_n is symmetric, $\frac{m_n}{n} = \frac{1}{n} tr(R_n^{'-1} J_n R_n' R_n J_n R_n^{-1})$. This occurs because, when W_n is symmetric, R_n is an orthonormal matrix and $\frac{1}{n} tr(R_n^{'-1} J_n R_n' R_n J_n R_n^{-1}) = \frac{m_n}{n}$. Identification condition in the second part of Assumption 8 will be satisfied regardless whether the value of $\frac{m_n}{n}$ is less than 1/2 or not. The symmetric W_n occurs in some social interactions models with group interactions.

Theorem 2 *Under Assumptions 1-5, 6 and 8, the QMLE $\hat{\lambda}_n$ is ψ_n-consistent such that $\psi_n(\hat{\lambda}_n - \lambda_{n0}) = o_p(1)$. Also, $\hat{\beta}_n - \beta_0 = o_p(1)$ and $\hat{\sigma}_n^2 - \sigma_0^2 = o_p(1)$.*

Proof. *See Appendix B.2.* ■

Hence, from either Theorem 1 or Theorem 2, $\hat{\lambda}_n$ has a higher rate of convergence than other parameters. This is so due to the unstable component in W_nY_n, which has an explosive factor ψ_n in its mean equation $G_nX_n\beta_0$; and has also an explosive variance of $G_n\epsilon_n$. As compared with a unit root time series with a drift term, the time series will have a time trend as its mean function and random walk disturbances. As time increases, the mean value of the sample observation increases and the variance will also increase. For the SAR model, spatial units are not ordered, and the mean values of W_nY_n and its variances for each spatial unit have the same order. Their magnitudes are larger for populations with larger spatial units. These are apparently the differences between our near unit scenario and the time series scenario. They have some common but also different features. The asymptotics for $\hat{\lambda}_n$ in Theorems 2 and 3, which include the results for a pure SAR process (with $\beta_0 = 0$), is observed to be different from the non-normal weak convergence results in Phillips (1987).

3.2 Asymptotic Distribution

From (10), the first order derivative are $\frac{\partial \ln L_n(\theta)}{\partial \beta} = \frac{1}{\sigma^2} X_n'\epsilon_n(\delta)$, $\frac{\partial \ln L_n(\theta)}{\partial \lambda} = \frac{1}{\sigma^2}(W_nY_n)'\epsilon_n(\delta) - tr(G(\lambda))$, $\frac{\partial \ln L_n(\theta)}{\partial \sigma^2} = \frac{1}{2\sigma^4}(\epsilon_n'(\delta)\epsilon_n(\delta) - n\sigma^2)$. The variance of the $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_{n0})}{\partial \theta}$ is $E\left(\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_{n0})}{\partial \theta}\cdot\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_{n0})}{\partial \theta'}\right) = -\left(E\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_{n0})}{\partial\theta\partial\theta'}\right) + \Omega_n$ and

$$-\left(E\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_{n0})}{\partial\theta\partial\theta'}\right) = E\begin{pmatrix} \frac{1}{\sigma_0^2 n}X_n'X_n & \frac{1}{\sigma_0^2 n}X_n'W_nY_n & 0 \\ * & \frac{1}{\sigma_0^2 n}(W_nY_n)'W_nY_n - \frac{1}{n}tr(G_n^2) & \frac{1}{\sigma_0^2 n}trG_n \\ * & * & \frac{1}{2\sigma_0^4} \end{pmatrix}, \tag{17}$$

$$\Omega_n = \begin{pmatrix} 0 & * & * \\ \frac{\mu_3}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}x_{i,n} & \frac{2\mu_3}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}(G_nX_n\beta_0)_i + \frac{\mu_4-3\sigma_0^4}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}^2 & * \\ \frac{\mu_3}{2\sigma_0^6 n}l_n'X_n & \frac{1}{2\sigma_0^6 n}\mu_3 l_n'G_nX_n\beta_0 + \frac{\mu_4-3\sigma_0^4}{2\sigma_0^6 n}trG_n & \frac{\mu_4-3\sigma_0^4}{4\sigma_0^8} \end{pmatrix}, \tag{18}$$

where Ω_n is symmetric and μ_3, μ_4 are, respectively, the third and fourth moments of ϵ_{ni}.

Using Lemma 2, we have $\frac{1}{n}X_n'X_n = O(1)$, $\frac{1}{n\psi_n}X_n'W_nY_n = O(1) + o_p(1)$ and $\frac{1}{n\psi_n^2}(W_nY_n)'W_nY_n = O(1) + o_p(1)$. Hence, the elements of $-\left(E\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_{n0})}{\partial\theta\partial\theta'}\right)$ and Ω_n have different orders because G_n has a higher order of ψ_n specified in (6). This implies that we need a rescaling matrix

$$\Upsilon_{\theta,n} = \begin{pmatrix} I_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \psi_n & 0 \\ \mathbf{0} & 0 & 1 \end{pmatrix},$$

such that

$$\begin{aligned} \lim_{n\to\infty} \Upsilon_{\theta,n}^{-1}\left(-\frac{1}{n}E\frac{\partial^2 \ln L_n(\theta_{n0})}{\partial\theta\partial\theta'}\right)\Upsilon_{\theta,n}^{-1} &= \Sigma, \\ \lim_{n\to\infty} \Upsilon_{\theta,n}^{-1}\Omega_n\Upsilon_{\theta,n}^{-1} &= \Omega. \end{aligned}$$

Assumption 9 Σ *is nonsingular.*

Assumption 9 is a local identification condition. The global identification condition in Assumption 7 implies apparently that Σ is nonsingular. The Assumption 9 is added to supplement the identification situation of Assumption 8. If θ_{n0} is a regular point (Rothenberg 1971), as Assumption 8 is a global identification condition which implies local identification, the limiting average Hessian matrix will be nonsingular. In the presence of Assumption 8, Assumption 9 can be guaranteed by the condition that $\lim_{n\to\infty}\frac{1}{n\psi_n^2}[trG_n'G_n + tr(G_n^2) - 2\frac{(trG_n)^2}{n}] \neq 0$, which is similar to the regular SAR case but with rescaling (see Lee (2004) for the regular SAR model).

By Taylor expansion, $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0}) = -\left(\Upsilon_{\theta,n}^{-1}\frac{1}{n}\frac{\partial^2 \ln L_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1}\right)^{-1}\Upsilon_{\theta,n}^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta}$ where $\tilde{\theta}_n$ lies between $\hat{\theta}_n$ and θ_{n0}. Hence, using the central limit theorem for linear and quadratic form of ϵ_n and convergence of $\Upsilon_{\theta,n}^{-1}\frac{1}{n}\frac{\partial^2 \ln L_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1}$ to Σ by using $\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0}) = o_p(1)$, we have the following theorem.

Theorem 3 *Under Assumptions 1-5, 6 and 9,*

$$\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0}) \xrightarrow{d} N(0, \Sigma^{-1} + \Sigma^{-1}\Omega\Sigma^{-1}).$$

When ϵ_{ni}'s are normally distributed, $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0}) \xrightarrow{d} N(0, \Sigma^{-1})$.

Proof. *See Appendix B.3* ■

4 Two Stage Least Squares Estimators

To estimate the SAR model (1), we can use IVs. Denote $\delta_n = (\beta', \lambda_n)'$ and $Z_n = (X_n, W_nY_n)$, (1) can be rewritten as

$$Y_n = Z_n\delta_{n0} + \epsilon_n. \tag{19}$$

Denote the $n \times p$ matrix H_n where $p \geq k+1$ as the possible instruments and we can use H_n to instrument (X_n, W_nY_n). The ideal instruments would be $E(Z_n|X_n) = (X_n, W_n(I_n - \lambda_{n0}W_n)^{-1}X_n\beta_0)$. Kelejian and Prucha (1998) use H_n that is generated from W_n and X_n, for example, $H_n = (X_n, W_nX_n, W_n^2X_n)$. Lee (2003) use $(X_n, W_n(I_n - \tilde{\lambda}_nW_n)^{-1}X_n\tilde{\beta}_n)$ as the feasible optimal instruments where $\tilde{\lambda}_n$ and $\tilde{\beta}_n$ are initial consistent estimators.

From (8) and (9), elements of W_nY_n have the order of ψ_n. Thus, denote

$$\Upsilon_{\delta,n} = \begin{pmatrix} I_k & \mathbf{0} \\ \mathbf{0} & \psi_n \end{pmatrix}. \tag{20}$$

4.1 2SLS

The SAR model may be estimated by the 2SLS method. Let H_n be an IV matrix generated from X_n and W_n, e.g., $H_n = (X_n, W_nX_n, W_n^2X_n)$, and $P_{H_n} = H_n(H_n'H_n)^{-1}H_n'$, a 2SLS estimate can be

$$\hat{\delta}_{n,2sls} = \{Z_n'P_{H_n}Z_n\}^{-1}Z_n'P_{H_n}Y_n. \tag{21}$$

Assumption 10 *(1) The instrument matrices H_n have full column rank $p \geq k+1$ (for all n large enough); (2) $\lim_{n\to\infty}\frac{1}{n}H_n'H_n$ is finite and of full rank.; (3) plim$\frac{1}{n}H_n'Z_n\Upsilon_{\delta,n}^{-1}$ is finite and of full rank.*

Since the number of variables is equal to $k+1$, the number of IVs must be at least $k+1$, as is specified in Assumption 10 (1). Assumption 10 (2) is standard for IVs and Assumption 10 (3) is similar to the rank condition for identification in a linear simultaneous equation system. As the IV matrix is generated as functions of W_n and X_n, it rules out the case when the exogenous variables are irrelevant.

From (21),

$$\Upsilon_{\delta,n}\sqrt{n}(\hat{\delta}_{n,2sls} - \delta_{n0}) = \begin{pmatrix} \sqrt{n}(\hat{\beta}_{n,2sls} - \beta_0) \\ \sqrt{n}\psi_n(\hat{\lambda}_{n,2sls} - \lambda_{n0}) \end{pmatrix} = \left[\frac{1}{n}\Upsilon_{\delta,n}^{-1}Z_n'P_{H_n}Z_n\Upsilon_{\delta,n}^{-1}\right]^{-1}\left[\Upsilon_{\delta,n}^{-1}Z_n'P_{H_n}\frac{1}{\sqrt{n}}\epsilon_n\right].$$

As $P_{H_n} = H_n(H_n'H_n)^{-1}H_n'$, we have

$$\begin{aligned} &\begin{pmatrix} \sqrt{n}(\hat{\beta}_{n,2sls} - \beta_0) \\ \sqrt{n}\psi_n(\hat{\lambda}_{n,2sls} - \lambda_{n0}) \end{pmatrix} \\ = &\left[\left(\frac{1}{n}\Upsilon_{\delta,n}^{-1}Z_n'H_n\right)\left(\frac{1}{n}H_n'H_n\right)^{-1}\left(\frac{1}{n}H_n'Z_n\Upsilon_{\delta,n}^{-1}\right)\right]^{-1}\left[\left(\frac{1}{n}\Upsilon_{\delta,n}^{-1}Z_n'H_n\right)\left(\frac{1}{n}H_n'H_n\right)^{-1}\left(\frac{1}{\sqrt{n}}H_n'\epsilon_n\right)\right]. \end{aligned}$$

Using Lemma 2, we will get the asymptotic properties of 2SLS estimators of λ_{n0} and β_0. With $\hat{\lambda}_{n,2sls}$ and $\hat{\beta}_{n,2sls}$ available, we can also get the estimator $\hat{\sigma}^2_{n,2sls} = \frac{1}{n}\sum_{i=1}^n(Y_n - Z_n\hat{\delta}_{n,2sls})'(Y_n - Z_n\hat{\delta}_{n,2sls})$.

Theorem 4 *Under Assumptions 1-5 and 10, $\hat{\delta}_{n,2sls}$ is consistent and*

$$\Upsilon_{\delta,n}\sqrt{n}(\hat{\delta}_{n,2sls} - \delta_{n0}) \xrightarrow{d} N(0, \Phi_{2sls})$$

where $\Phi_{2sls} = \sigma_0^2\left(\lim_{n\to\infty}\frac{1}{n}\Upsilon_{\delta,n}^{-1}Z_n'P_{H_n}Z_n\Upsilon_{\delta,n}^{-1}\right)^{-1}$. Also, $\sqrt{n}(\hat{\sigma}^2_{n,2sls} - \sigma_0^2) \xrightarrow{d} N(0, 2\sigma_0^4 + (\mu_4 - 3\sigma_0^4))$.

Proof. *See Appendix B.4.* ■

This shows that the 2SLS estimate $\hat{\lambda}_{n,2sls}$ of λ_{n0} is $\sqrt{n}\psi_n$-consistent. The rate is higher than the $\sqrt{n}$ rate in Kelejian and Prucha (1998) and Lee (2003). For the 2SLS estimate $\hat{\beta}_{n,2sls}$ of β_0, it is at the usual $\sqrt{n}$ rate of convergence.

4.2 Best 2SLS

The above 2SLS estimate might not be asymptotically optimal. The optimal IV is $H_n^* = E(X_n, W_nY_n|X_n) = (X_n, G_nX_n\beta_0)$ where $G_n = W_nS_n^{-1}(\lambda_{n0})$. To construct the optimal IV, Lee (2003) uses an initial estimator $\tilde{\delta}_n = (\tilde{\beta}_n', \tilde{\lambda}_n)'$ to approximate H_n^*. Denote $\tilde{H}_n^* = (X_n, G_n(\tilde{\lambda}_n)X_n\tilde{\beta}_n)$.

Assumption 11 $\lim_{n\to\infty} \Upsilon_{\delta,n}^{-1} \frac{1}{n} H_n^{*\prime} H_n^* \Upsilon_{\delta,n}^{-1}$ *exists and is nonsingular.*

Assumption 12 $\Upsilon_{\delta,n}(\tilde{\delta}_n - \delta_{n0}) = o_p(1)$.

Assumption 11 is similar to Assumption 10 (2). The nonsingular matrix properties in Assumption 4 and 7 will imply Assumption 11 from the partition matrix inversion formulae. As $G_n = W_nS_n^{-1}(\lambda_{n0})$ in H_n^* has the higher order of ψ_n, we need to standardize H_n^* by $\Upsilon_{\delta,n}^{-1}$. Assumption 12 requires that we need initial $\Upsilon_{\delta,n}$-consistent estimators $\tilde{\delta}_n$ to approximate H_n^*. A natural candidate would be the 2SLS from (21). Let

$$\hat{\delta}_{n,b2sls} = (\tilde{H}_n^{*\prime} Z_n)^{-1} \tilde{H}_n^{*\prime} Y_n. \tag{22}$$

Then, $\Upsilon_{\delta,n}\sqrt{n}\hat{\delta}_{n,b2sls} = (\Upsilon_{\delta,n}^{-1}\frac{1}{n}\tilde{H}_n^{*\prime}Z_n\Upsilon_{\delta,n}^{-1})^{-1}\Upsilon_{\delta,n}^{-1}\frac{1}{\sqrt{n}}\tilde{H}_n^{*\prime}Y_n$, which implies that

$$\Upsilon_{\delta,n}\sqrt{n}(\hat{\delta}_{n,b2sls} - \delta_{n0}) = (\Upsilon_{\delta,n}^{-1}\frac{1}{n}\tilde{H}_n^{*\prime}Z_n\Upsilon_{\delta,n}^{-1})^{-1}\Upsilon_{\delta,n}^{-1}\frac{1}{\sqrt{n}}\tilde{H}_n^{*\prime}\epsilon_n.$$

We can show that $\text{plim}_{n\to\infty}\Upsilon_{\delta,n}^{-1}\frac{1}{n}\tilde{H}_n^{*\prime}Z_n\Upsilon_{\delta,n}^{-1}$=$\text{plim}_{n\to\infty}\Upsilon_{\delta,n}^{-1}\frac{1}{n}H_n^{*\prime}H_n^*\Upsilon_{\delta,n}^{-1}$ and $\text{plim}_{n\to\infty}\frac{1}{n}\Upsilon_{\delta,n}^{-1}\tilde{H}_n^{*\prime}\epsilon_n = 0$, which will give us consistency. After showing $\Upsilon_{\delta,n}^{-1}\frac{1}{\sqrt{n}}\tilde{H}_n^{*\prime}\epsilon_n \xrightarrow{d} N(0, \sigma_0^2 \lim_{n\to\infty}\frac{1}{n}\Upsilon_{\delta,n}^{-1}H_n^{*\prime}H_n^*\Upsilon_{\delta,n}^{-1})$, we can get the following theorem.

Theorem 5 *Under Assumptions 1-5, 11 and 12,* $\hat{\delta}_{n,b2sls}$ *is consistent and*

$$\Upsilon_{\delta,n}\sqrt{n}(\hat{\delta}_{n,b2sls} - \delta_{n0}) \xrightarrow{d} N(0, \Psi_{b2sls})$$

where $\Psi_{b2sls} = \sigma_0^2\left(\lim_{n\to\infty}\frac{1}{n}\Upsilon_{\delta,n}^{-1}H_n^{*\prime}H_n^*\Upsilon_{\delta,n}^{-1}\right)^{-1}$.

Proof. *See Appendix B.5* ■

The variance of the best 2SLS estimate is $\Psi_{b2sls} = \sigma_0^2\left(\lim_{n\to\infty}\frac{1}{n}\Upsilon_{\delta,n}^{-1}H_n^{*\prime}H_n^*\Upsilon_{\delta,n}^{-1}\right)^{-1}$ and the variance of 2SLS estimate is $\Phi_{2sls} = \sigma_0^2\left(\lim_{n\to\infty}\frac{1}{n}\Upsilon_{\delta,n}^{-1}Z_n'P_{H_n}Z_n\Upsilon_{\delta,n}^{-1}\right)^{-1}$. As $\lim_{n\to\infty}\frac{1}{n}\Upsilon_{\delta,n}^{-1}Z_n' = \lim_{n\to\infty}\frac{1}{n}\Upsilon_{\delta,n}^{-1}H_n^{*\prime}$, we can see that $\Psi_{b2sls} - \Phi_{2sls}$ is negative semi-definite using generalized Schwartz inequality.

4.3 Linearized QMLE

A computationally convenient way to increase the efficiency is to use the Newton-Raphson method to get the linearized QMLE. Assume that we have an initial estimator available (e.g., the two stage least square estimators in Theorem 4) where $\Upsilon_{\theta,n}(\hat{\theta}_{n1} - \theta_{n0}) = o_p(1)$. The linearized QMLE is to set

$$\hat{\theta}_{n2} = \hat{\theta}_{n1} - \left(\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'} \right)^{-1} \frac{\partial \ln L_n(\hat{\theta}_{n1})}{\partial\theta}, \tag{23}$$

where $\ln L_n(\theta)$ is the log likelihood function in (10).

Theorem 6 *Under Assumptions 1-5, 9 and* $\Upsilon_{\theta,n}(\hat{\theta}_{n1} - \theta_{n0}) = o_p(1)$,

$$\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_{n2} - \theta_{n0}) \xrightarrow{d} N(0, \Sigma^{-1} + \Sigma^{-1}\Omega\Sigma^{-1}).$$

When ϵ_{ni}*'s are normally distributed,* $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_{n2} - \theta_{n0}) \xrightarrow{d} N(0, \Sigma^{-1})$.

Proof. *See Appendix B.6.* ■

From Theorems 3 and 6, we see that the linearized QMLE has the same limiting distribution as the QMLE. For a regular likelihood function, this is a familiar result in econometrics (Ruud 2000). The linearized QMLE is a second round estimator, which can be computationally simpler than the iterative maximum likelihood approach. But, there is evidence that its finite sample properties might not be better than the iterative one. However, it may have an improvement on the asymptotic efficiency of the 2SLSE when the disturbances are normally distributed. For the initial 2SLS estimators, the variance of $\hat{\delta}_{n,2sls}$ is equal to $\Phi_{2sls} = \sigma_0^2 \left(\lim_{n\to\infty} \frac{1}{n} \Upsilon_{\delta,n}^{-1} Z_n' P_{H_n} Z_n \Upsilon_{\delta,n}^{-1} \right)^{-1}$. For the linearized MLE when the disturbance is normally distributed, we have the variance of the estimates of δ_{n0} as $\sigma_0^2 \left(\lim_{n\to\infty} \frac{1}{n} \Upsilon_{\delta,n}^{-1} (\Delta_n + Z_n' Z_n) \Upsilon_{\delta,n}^{-1} \right)^{-1}$ where $\Delta_n = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & tr((G_n - \frac{trG_n}{n} I_n)^s G_n \end{pmatrix}$ and $(G_n - \frac{trG_n}{n} I_n)^s = (G_n - \frac{trG_n}{n} I_n) + (G_n - \frac{trG_n}{n} I_n)'$. As $\Delta_n + (Z_n' Z_n - Z_n' P_{H_n} Z_n)$ is sum of two positive semi-definite matrices, the variance of linearized MLE is smaller than that of 2SLSE.

5 Monte Carlo Results

The model in the Monte Carlo study is specified as

$$Y_n = \lambda_{n0} W_n Y_n + X_{n1}\beta_{01} + X_{n2}\beta_{02} + X_{n3}\beta_{03} + \epsilon_n, \tag{24}$$

where x_{i1}, x_{i2} and x_{i3} are three independently generated standard normal variables and are *i.i.d.* for all i, and ϵ_{ni}'s are also drawn independently from the standard normal distribution. We choose $n = 49$ and

Table 1: 2SLSEs and MLEs, with exogenous variables: $n = 49$

Method	λ	β_1	β_2	β_3
	Mean (SD)[RMSE]	Mean (SD)[RMSE]	Mean (SD)[RMSE]	Mean (SD)[RMSE]
λ_{n0}= 0.6				
2SLS	0.6472 (0.1529)[0.1599]	-0.9918 (0.1547)[0.1548]	0.0112 (0.1524)[0.1528]	0.9949 (0.1505)[0.1505]
B2SLS	0.5520 (0.2068)[0.2122]	-1.0009 (0.1562)[0.1561]	0.0017 (0.1535)[0.1539]	1.0030 (0.1535)[0.1535]
MLE	0.5705 (0.1117)[0.1155]	-1.0050 (0.1523)[0.1523]	0.0118 (0.1530)[0.1533]	1.0063 (0.1492)[0.1493]
LMLE	0.5885 (0.1198)[0.1203]	-1.0030 (0.1528)[0.1528]	0.0116 (0.1528)[0.1532]	1.0005 (0.1491)[0.1491]
λ_{n0}= 0.9				
2SLS	0.9106 (0.0632)[0.0640]	-0.9980 (0.1543)[0.1542]	0.0128 (0.1516)[0.1521]	1.0009 (0.1527)[0.1526]
B2SLS	0.8595 (0.1405)[0.1462]	-1.0158 (0.1609)[0.1616]	0.0104 (0.1566)[0.1568]	1.0154 (0.1579)[0.1585]
MLE	0.8775 (0.0549)[0.0593]	-1.0105 (0.1536)[0.1539]	0.0117 (0.1541)[0.1544]	1.0115 (0.1513)[0.1517]
LMLE	0.8890 (0.0569)[0.0579]	-1.0094 (0.1541)[0.1543]	0.0128 (0.1528)[0.1533]	1.0114 (0.1534)[0.1537]
λ_{n0}= 0.99				
2SLS	0.9850 (0.0212)[0.0217]	-1.0048 (0.1548)[0.1548]	0.0131 (0.1534)[0.1538]	1.0146 (0.1532)[0.1538]
B2SLS	0.9802 (0.0632)[0.0639]	-1.0089 (0.1565)[0.1566]	0.0139 (0.1604)[0.1609]	1.0192 (0.1567)[0.1578]
MLE	0.9833 (0.0201)[0.0211]	-1.0084 (0.1526)[0.1528]	0.0121 (0.1542)[0.1546]	1.0113 (0.1501)[0.1504]
LMLE	0.9827 (0.0194)[0.0207]	-1.0103 (0.1534)[0.1545]	0.0130 (0.1537)[0.1541]	1.0190 (0.1539)[0.1550]
λ_{n0}= 0.999				
2SLS	0.9980 (0.0041)[0.0043]	-1.0079 (0.1540)[0.1541]	0.0153 (0.1560)[0.1566]	1.0150 (0.1521)[0.1527]
B2SLS	0.9978 (0.0140)[0.0140]	-1.0059 (0.1523)[0.1523]	0.0150 (0.1606)[0.1612]	1.0132 (0.1524)[0.1529]
MLE	0.9982 (0.0049)[0.0050]	-1.0045 (0.1521)[0.1521]	0.0119 (0.1531)[0.1534]	1.0072 (0.1488)[0.1489]
LMLE	0.9979 (0.0038)[0.0039]	-1.0115 (0.1532)[0.1535]	0.0152 (0.1557)[0.1563]	1.0173 (0.1523)[0.1523]

Note:

1. $\beta_{01} = -1$, $\beta_{02} = 0$ and $\beta_{03} = 1$, rep=1000.
2. We only report the estimates where the estimated $\hat{\lambda}_n$ is smaller than 1 in absolute value. For the B2SLS and LMLE which use initial estimates from 2SLS, we use only those initial estimates that are smaller than 1.
3. For $\lambda_{n0} = 0.6$, the numbers of estimates $\hat{\lambda}_n$ that are smaller than 1 are, respectively, 992, 983, 1000, 992 for 2SLS, B2SLS, MLE and LMLE.
4. For $\lambda_{n0} = 0.9$, the numbers of estimates $\hat{\lambda}_n$ that are smaller than 1 are, respectively, 884, 849, 1000, 884 for 2SLS, B2SLS, MLE and LMLE.
5. For $\lambda_{n0} = 0.99$, the numbers of estimates $\hat{\lambda}_n$ that are smaller than 1 are, respectively, 729, 684, 998, 729 for 2SLS, B2SLS, MLE and LMLE.
6. For $\lambda_{n0} = 0.999$, the numbers of estimates $\hat{\lambda}_n$ that are smaller than 1 are, respectively, 749, 704, 989, 949 for 2SLS, B2SLS, MLE and LMLE

Table 2: 2SLSEs and MLEs, with exogenous variables: $n = 245$

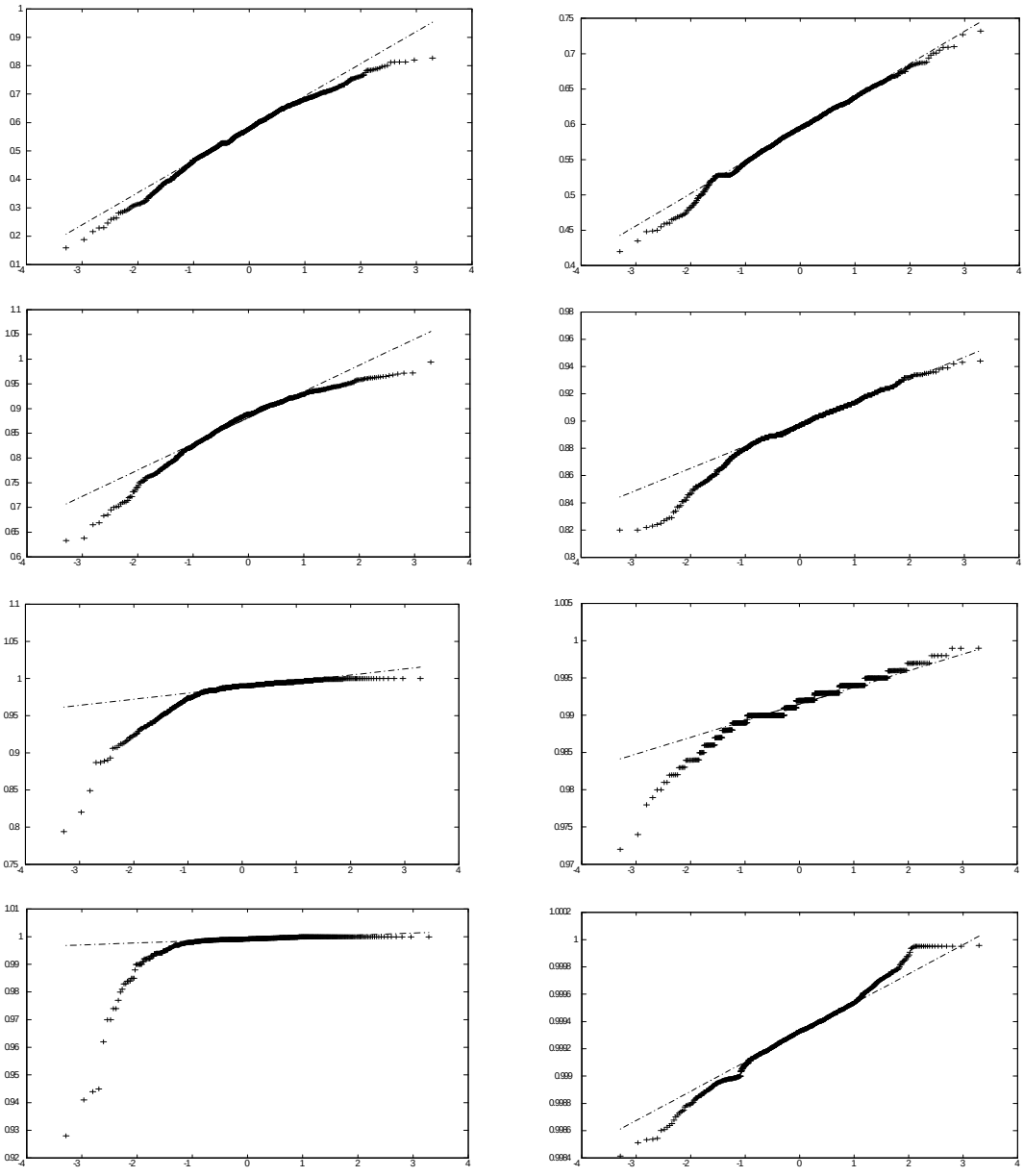
Method	λ	β_1	β_2	β_3
	Mean (SD)[RMSE]	Mean (SD)[RMSE]	Mean (SD)[RMSE]	Mean (SD)[RMSE]
λ_{n0}= 0.6				
2SLS	0.6120 (0.0623)[0.0635]	-0.9982 (0.0653)[0.0652]	-0.0024 (0.0665)[0.0665]	0.9977 (0.0650)[0.0650]
B2SLS	0.5950 (0.0650)[0.0652]	-1.0007 (0.0652)[0.0652]	-0.0025 (0.0668)[0.0668]	1.0003 (0.0649)[0.0649]
MLE	0.5924 (0.0473)[0.0478]	-1.0019 (0.0647)[0.0647]	-0.0023 (0.0667)[0.0667]	1.0015 (0.0646)[0.0645]
LMLE	0.5979 (0.0465)[0.0465]	-1.0010 (0.0647)[0.0647]	-0.0023 (0.0666)[0.0666]	1.0006 (0.0645)[0.0645]
λ_{n0}= 0.9				
2SLS	0.9053(0.0283)[0.0287]	-0.9978 (0.0659)[0.0659]	-0.0024 (0.0664)[0.0664]	0.9971 (0.0654)[0.0654]
B2SLS	0.8929 (0.0353)[0.0360]	-1.0031 (0.0655)[0.0655]	-0.0025 (0.0669)[0.0670]	1.0025 (0.0651)[0.0651]
MLE	0.8958 (0.0198)[0.0202]	-1.0028 (0.0649)[0.0650]	-0.0023 (0.0668)[0.0668]	1.0023 (0.0646)[0.0647]
LMLE	0.8981 (0.0200)[0.0201]	-1.0017 (0.0650)[0.0650]	-0.0023 (0.0667)[0.0667]	1.0012 (0.0647)[0.0646]
λ_{n0}= 0.99				
2SLS	0.9903 (0.0053)[0.0053]	-1.0008 (0.0654)[0.0653]	-0.0026 (0.0663)[0.0663]	0.9995 (0.0645)[0.0645]
B2SLS	0.9867 (0.0490)[0.0491]	-1.0045 (0.0736)[0.0737]	-0.0035 (0.0849)[0.0850]	1.0023 (0.0673)[0.0674]
MLE	0.9914 (0.0030)[0.0033]	-0.9980 (0.0640)[0.0640]	0.0020 (0.0664)[0.0664]	0.9974 (0.0640)[0.0640]
LMLE	0.9899 (0.0033)[0.0033]	-1.0025 (0.0645)[0.0645]	-0.0024 (0.0665)[0.0665]	1.0012 (0.0640)[0.0640]
λ_{n0}= 0.999				
2SLS	0.9990 (0.0008)[0.0008]	-1.0015 (0.0650)[0.0650]	-0.0026 (0.0672)[0.0672]	0.9999 (0.0643)[0.0643]
B2SLS	0.9988 (0.0011)[0.0011]	-1.0028 (0.0642)[0.0642]	-0.0016 (0.0675)[0.0674]	1.0020 (0.0643)[0.0643]
MLE	0.9993 (0.0003)[0.0004]	-0.9961 (0.0639)[0.0640]	-0.0019 (0.0663)[0.0663]	0.9954 (0.0640)[0.0641]
LMLE	0.9990 (0.0004)[0.0004]	-1.0025 (0.0641)[0.0641]	-0.0024 (0.0663)[0.0663]	1.0009 (0.0639)[0.0639]

Note:

1. $\beta_{01} = -1$, $\beta_{02} = 0$ and $\beta_{03} = 1$, rep=1000.
2. We only report the estimates where the estimated $\hat{\lambda}_n$ is smaller than 1. For the B2SLS and LMLE which use initial estimates from 2SLS, we use only those initial estimates that are smaller than 1.
3. For $\lambda_{n0} = 0.6$, all the 1000 estimate of $\hat{\lambda}_n$ for each case are smaller than 1.
4. For $\lambda_{n0} = 0.9$, all the 1000 estimate of $\hat{\lambda}_n$ for each case are smaller than 1 except for one estimate in B2SLS.
5. For $\lambda_{n0} = 0.99$, the numbers of estimates $\hat{\lambda}_n$ that are smaller than 1 are, respectively, 956, 950, 1000, 956 for 2SLS, B2SLS, MLE and LMLE.
6. For $\lambda_{n0} = 0.999$, the numbers of estimates $\hat{\lambda}_n$ that are smaller than 1 are, respectively, 937, 931, 1000, 937 for 2SLS, B2SLS, MLE and LMLE.

Table 3: Q-Q Plot of $\hat{\lambda}_n$: Exogenous variable included–MLE

$n = 49$ $n = 245$

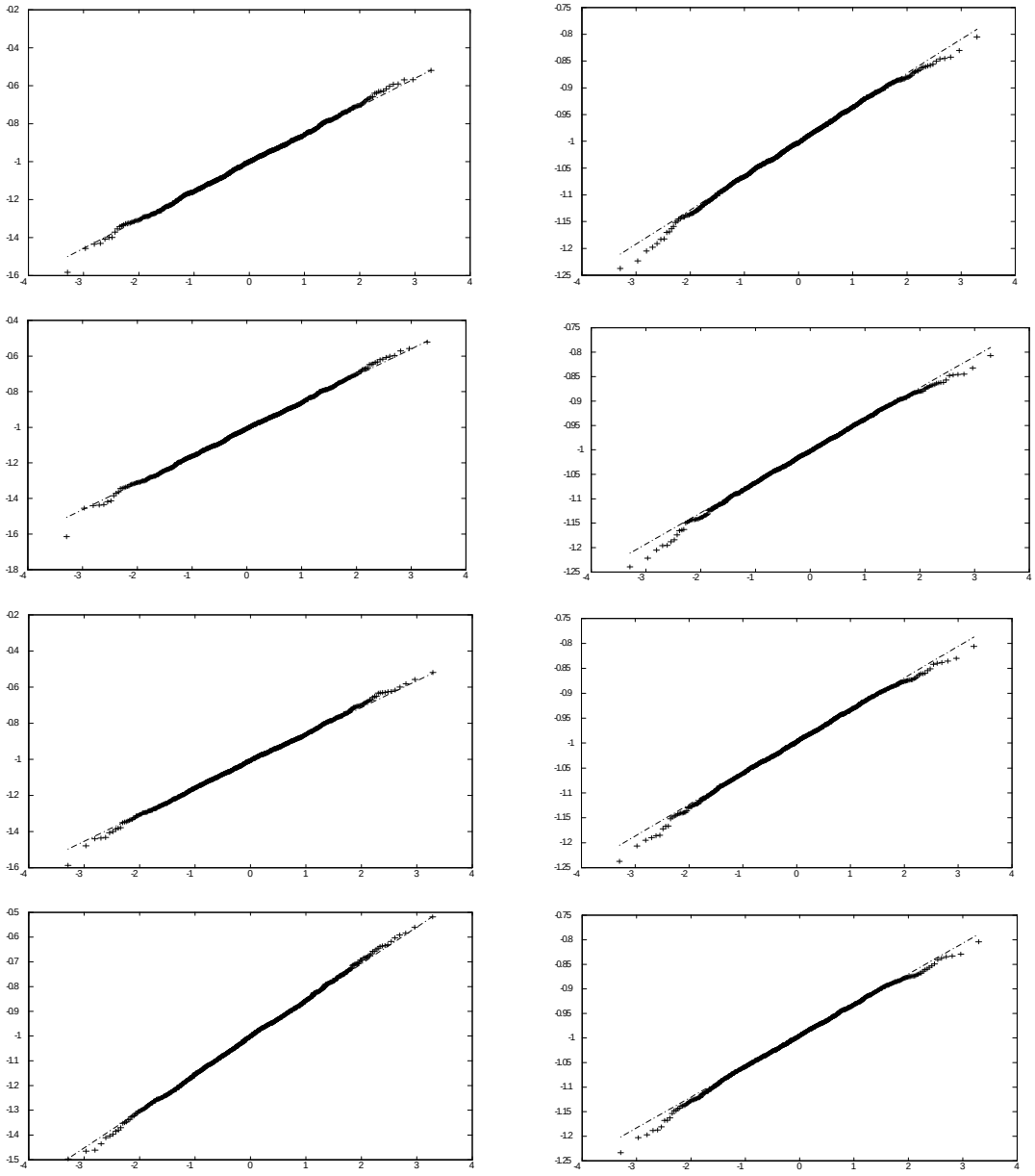


Note:

1. From 1st row to fourth row are $\lambda_{n0} = 0.6, 0.9, 0.99, 0.999$.
2. Only the estimates with $\hat{\lambda}_n$ smaller than 1 are plotted.

Table 4: Q-Q Plot of $\hat{\beta}_{1n}$: Exogenous variable included–MLE

$n = 49$ $n = 245$



Note:

1. From 1st row to fourth row are $\lambda_{n0} = 0.6, 0.9, 0.99, 0.999$.
2. Only the estimates with $\hat{\lambda}_n$ smaller than 1 are plotted.

Table 5: MLE $\hat{\lambda}_n$ without exogenous variables

	$n = 49$	$n = 245$
	Mean (SD)[RMSE]	Mean (SD)[RMSE]
λ_{n0}= 0.6	0.5585 (0.1603)[0.1655]	0.5816 (0.0667)[0.0692]
λ_{n0}= 0.9	0.8660 (0.0841)[0.0907]	0.8907 (0.0290)[0.0304]
λ_{n0}= 0.99	0.9803 (0.0317)[0.0331]	0.9937 (0.0045)[0.0058]
λ_{n0}= 0.999	0.9979 (0.0083)[0.0083]	0.9999 (0.0001)[0.0009]

Note:
1. $n = 49$ and $n = 245$, no exogenous variables included , rep=1000.
2. For $\lambda_{n0} = 0.6, 0.9, 0.99$, all the 1000 estimates for each case are smaller than 1.
3. For $\lambda_{n0} = 0.999$, there is one estimate that exceeds 1 when $n = 49$.

$n = 245$. When $n = 49$, the spatial weights matrix we choose is the neighborhood matrix across 49 districts in Columbus, Ohio in Anselin (1988) for crime study. For $n = 245$, the corresponding spatial weights matrix is a block diagonal matrix with the preceding 49×49 matrix as their diagonal blocks. This corresponds to the pooling of five separate districts with similar neighboring structures in each district. The estimation methods we consider are:

(1) 2SLS–the 2SLS method with IV's X_n, W_nX_n and $W_n^2X_n$;

(2) B2SLS–the best 2SLS method with IV's $G_n(\tilde{\lambda}_n)X_n\tilde{\beta}_n$ and X_n where $\tilde{\lambda}_n$ and $\tilde{\beta}_n$ are initial consistent estimators. In the simulation, we use the 2SLS estimators are the initial estimators;

(3) MLE–the ML approach;

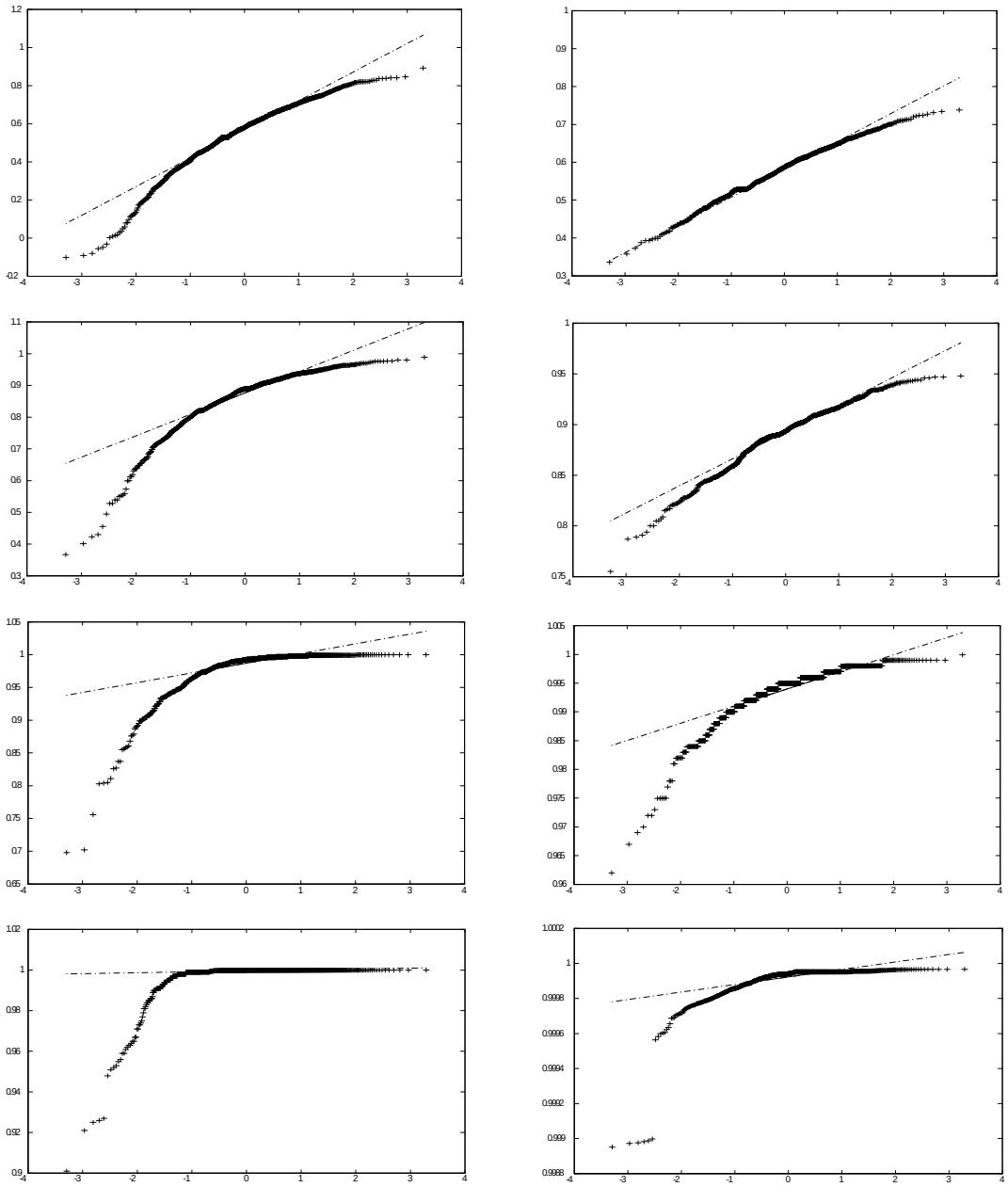
(4) LMLE–the Newton-Raphson method using (23). In the simulation, we use the 2SLS estimates as the initial estimates.

The number of repetition is 1000 for each case. In each case, we report the mean (Mean), empirical standard deviation (SD) and empirical root mean square error (RMSE) of the estimates. The true λ_{n0} takes the value $0.6, 0.9, 0.99, 0.999$ and the true $\beta_{01}, \beta_{02}, \beta_{03}$ are $-1, 0, 1$ respectively. Table 1 is the result when $n = 49$ and Table 2 is the result when $n = 245$. To check whether the estimates under (24) are normally distributed, Table 3 and Table 4 plot the quantile-quantile diagram of MLE of λ_{n0} and β_{01} under different values of λ_{n0} and n.

We can see that all the estimates have small biases under different values of λ_{n0}. When λ_{n0} is closer to 1, the SD of $\hat{\lambda}_n$ decreases and the SD of β does not change in the order of magnitude, which is consistent with our theoretical prediction. In terms of finite sample performance, the ML estimates are better than the IV estimates in terms of smaller variances. The B2SLSEs turn out worse than the simpler 2SLSEs in terms of variances for these two sample sizes. Numerically, the ML approach is also more stable than the IV

Table 6: Q-Q Plot of $\hat{\lambda}_n$: Exogenous variable excluded–MLE

$n = 49$ $n = 245$



Note:

1. From 1st row to fourth row are $\lambda_{n0} = 0.6, 0.9, 0.99, 0.999$.
2. Only the estimates with $\hat{\lambda}_n$ smaller than 1 are plotted.

approach as it rarely produces estimates of λ_{n0} greater than 1. From the quantile-quantile plot, we can see that the estimators have large proportions lying in a straight line. This is more obvious for the estimates of β_0 than the estimates of λ_{n0}. These indicate that the distribution of the estimates of β can be better approximated than that of the estimates of λ_{n0} by the normal distribution.

We also report the result of the pure SAR model without exogenous variables

$$Y_n = \lambda_{n0} W_n Y_n + \epsilon_n. \tag{25}$$

The weights matrices are the same and we have $n = 49$, $n = 245$. The results are reported in Table 5. We can see that the estimates of λ_{n0} have some downward biases for $\lambda_{n0} = 0.6$, or 0.9 when $n = 49$. The biases become small with $n = 245$. When λ_{n0} is closer to 1, the SD of $\hat{\lambda}_n$ decreases, which is consistent with our theoretical prediction. In time series, when there is no exogenous variables included, the near unit root case is analyzed in terms of the weak convergence. According to Theorem 3, for the near unit root case in SAR model when there is no exogenous variables included, the estimate is still normally distributed. To check that, we plot the quantile-quantile diagram for this case in Table 6. From the quantile-quantile plot, we can see that the estimator of λ_{n0} can be approximated by the normal distribution similar to those in Table 3.

6 Conclusion

In this paper, asymptotic properties of estimates from various estimation methods (namely, QML, 2SLS and linearized QML) for the near unit root case in the SAR model are investigated. The estimates of the spatial effect parameter have a higher rate of convergence and other parameters have regular $\sqrt{n}$ rate. Similarly to the regular SAR model, QMLE and linearized QMLE are more efficient than 2SLSE.

The SAR model considered in this paper has a row-normalized spatial weights matrix W_n. The near unit root for this SAR model refers to the case where the spatial effect λ of the spatial lag $W_n Y_n$ is close to unity. Similarly to the time series literature (Phillips 1987), this concept is modeled with a sequence λ_n, which converges to the unity as the sample increases. This situation generates an irregularity in that the inverse of the spatial transformation matrix $(I_n - \lambda_n W_n)$ will not satisfy the uniform boundedness in both row and column sum norms property. The uniform boundedness property originated in Kelejian and Prucha (1998, 2001) is one of the essential regularity conditions in order to establish the asymptotic properties of various estimators, e.g., IV estimates in Kelejian and Prucha (1998) and Lee (2004) for the QMLE. This paper investigates the possible asymptotic implications on those estimators when the uniform boundedness property on $(I_n - \lambda_n W_n)^{-1}$ is violated.

With a near unit root λ_n, the outcome of Y_n for each spatial unit can be decomposed into two components

– one represents an unstable component which contains explosive elements for large n, and the remaining component is stable. Accordingly, the spatial lag W_nY_n will also contain an explosive component, which turns out to be exactly the same unstable component of Y_n. The consequence is that the IV and QMLE of λ_n have a faster rate of convergence, while the estimates of the other remaining regression coefficients and the variance of disturbances have the usual rate of convergence. The fast rate of convergence of λ_n reflects the rate of convergence of λ_n to the unity.

Contrary to the near unit root model of an autoregressive process in time series, the asymptotic distributions of the estimates remain to be asymptotically normal. This difference in the asymptotic distribution may be due to differences in structures of the SAR model and the autoregressive time series process. With a near unit λ_n, the variance of Y_n conditional on exogenous variables can be very large, which is a feature for a time series with a unit root or near unit root. However, there are apparent differences between a SAR model and an autoregressive model in time series. A SAR model is an equilibrium model as the dependent variables for all the spatial units are equilibria determined by the exogenous variables and the disturbances, and there is not an initial value problem which can be important in a time series or a panel dynamic model. For the typical SAR model with a row-normalized weights matrix, the coefficient of W_nY_n can not be the unity; otherwise, the model is not complete. The simplest SAR process which resembles the autoregressive process of a time series is the spatial process where spatial units lie in a circular world. In a circular world, there is neither a beginning nor an end. Thus, contrary to the time series with a starting period, the variances of all the outcomes of spatial units in the presence of a near unit root have the same large order of magnitudes. These different features may explain the different asymptotics obtained in this paper as compared with the nonstandard asymptotics via stochastic integral representation in Phillips (1987).

We have also compared the performance of the IV estimator with the ML estimator for the SAR model and the pure SAR process by asymptotic efficiency as well as finite sample performance via Monte Carlo experiments. In terms of finite sample performance, both the ML and IV type estimates have small biases, but the ML estimates are better than the IV estimates in terms of smaller variances. Numerically, the ML approach is also more stable than the IV approach.

While this paper has clarified the asymptotic implications on the near unit root case for the SAR model and its estimates, there are other related issues that have not been investigated in this paper. As in Fingleton (1999), there might be implications on spurious regression and co-integration in the spatial context. Those issues have not been addressed in this paper but they may be worthy of investigation in future research.

Appendices

A Some Lemmas and Proofs

Lemma 1 *(Elementary properties involving G_n) Under Assumptions 1-5,*

(1) $\frac{1}{n\psi_n}X_n'G_nX_n = \frac{1}{n}X_n'R_nJ_nR_n^{-1}X_n + O(\frac{1}{\psi_n})$;

(2) $\frac{1}{n\psi_n^2}X_n'G_n'G_nX_n = \frac{1}{n}X_n'(R_nJ_nR_n^{-1})'(R_nJ_nR_n^{-1})X_n + O(\frac{1}{\psi_n})$;

(3) $\frac{1}{\psi_n\sqrt{n}}X_n'G_n'\epsilon_n = \frac{1}{\sqrt{n}}X_n'R_n^{'-1}J_nR_n'\epsilon_n + O_p(\frac{1}{\psi_n})$;

(4) $\frac{1}{\psi_n^2\sqrt{n}}X_n'G_n'G_n\epsilon_n = \frac{1}{\sqrt{n}}X_n'(R_nJ_nR_n^{-1})'(R_nJ_nR_n^{-1})\epsilon_n + O_p(\frac{1}{\psi_n})$;

(5) $\frac{1}{\psi_n\sqrt{n}}(\epsilon_n'G_n\epsilon_n - \sigma_0^2tr(G_n)) = \frac{1}{\sqrt{n}}[\epsilon_n'R_nJ_nR_n^{-1}\epsilon_n - \sigma_0^2tr(J_n)] + O_p(\frac{1}{\psi_n})$;

(6) $\frac{1}{\psi_n^2\sqrt{n}}(\epsilon_n'G_n'G_n\epsilon_n - \sigma_0^2tr(G_n'G_n)) = \frac{1}{\sqrt{n}}[\epsilon_n'(R_nJ_nR_n^{-1})'(R_nJ_nR_n^{-1})\epsilon_n - \sigma_0^2tr(R_n^{'-1}J_nR_n'R_nJ_nR_n^{-1})] + O_p(\frac{1}{\psi_n})$;

(7) $\lim_{n\to\infty}\frac{1}{n\psi_n^2}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) = \beta_0'\lim_{n\to\infty}(\frac{1}{n}X_n'R_n^{'-1}J_nR_n'M_nR_nJ_nR_n^{-1}X_n)\beta_0$;

(8) $\frac{1}{n\psi_n}tr(G_n) = \frac{m_n}{n} + O(\frac{1}{\psi_n})$; *and*

(9) $\frac{1}{n\psi_n^2}tr(G_n'G_n) = \frac{1}{n}tr(R_n^{'-1}J_nR_n'R_nJ_nR_n^{-1}) + O(\frac{1}{\psi_n})$.

Lemma 2 *Under Assumptions 1-5, for any $n\times n$ nonstochastic matrix B_n which is uniformly bounded in row sum and column sum norms,*

$$\frac{1}{\sqrt{n}}X_n'B_n\epsilon_n \xrightarrow{d} N(0,\sigma_0^2\lim_{n\to\infty}\frac{1}{n}X_n'B_nB_n'X_n). \tag{26}$$

Hence,

$$\frac{1}{n\cdot\psi_n}X_n'B_nY_n - E\frac{1}{n\cdot\psi_n}X_n'B_nY_n = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{27}$$

$$\frac{1}{n\cdot\psi_n^2}Y_n'B_nY_n - E\frac{1}{n\cdot\psi_n^2}Y_n'B_nY_n = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{28}$$

where $E\frac{1}{n\cdot\psi_n}X_n'B_nY_n$ and $E\frac{1}{n\cdot\psi^2(n)}Y_n'B_nY_n$ are $O(1)$. Also,

$$\frac{1}{n\cdot\psi_n}X_n'B_nY_n - \frac{1}{n}X_n'B_nR_nJ_nR_n^{-1}X_n\beta_0 = o_p(1), \tag{29}$$

$$\frac{1}{n\cdot\psi_n^2}Y_n'B_nY_n - \frac{1}{n}[(R_nJ_nR_n^{-1}X_n\beta_0)'B_nR_nJ_nR_n^{-1}X_n\beta_0 + \sigma_0^2tr((R_nJ_nR_n^{-1})'B_nR_nJ_nR_n^{-1})] = o_p(1). \tag{30}$$

Consider the parameter space of λ_{n0}: $\Lambda = (-1,\ 1)$. Also, let $\mathfrak{M}$ be a finite constant such that $\mathfrak{M}$ is greater than 1 and ϵ. Denote

$$\Lambda_n^{(\mathfrak{M}+)} = \{\lambda|\lambda\in\Lambda,\ \psi_n(\lambda_{n0}-\lambda) > \mathfrak{M}\}, \tag{31}$$

$$\Lambda_n^{(\mathfrak{M})} = \{\lambda|\lambda\in\Lambda,\ |\psi_n(\lambda_{n0}-\lambda)| \le \mathfrak{M}\}. \tag{32}$$

Note that as $\lambda_{n0} = 1 - \frac{1}{\psi_n}$, $\Lambda_n^{(\mathfrak{M}+)} = \{\lambda | \lambda \in \Lambda,\ |\psi_n(\lambda_{n0} - \lambda)| > \mathfrak{M}\}$ (This is so, because $\mathfrak{M} > 1$ and $\lambda < 1$ in Λ. If $\psi_n(\lambda - \lambda_{n0}) > \mathfrak{M}$, then $\psi_n(\lambda - 1) > \mathfrak{M} - 1$, which implies that $\lambda > 1$, a contradiction).

Lemma 3 *(Uniform convergence of variance) Under Assumptions 1-5, for $\hat{\sigma}_n^2(\lambda)$ in (12) and $\sigma_n^{*2}(\lambda)$ in (16), we have*

(i) $[\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda)] = O_p(\frac{1}{\sqrt{n}})$ uniformly in $\lambda \in \Lambda_n^{(\mathfrak{M})}$;

(ii) $\frac{1}{[\psi_n(\lambda_{n0} - \lambda)]^2}[\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda)] = O_p(\frac{1}{\sqrt{n}})$ uniformly in $\lambda \in \Lambda_n^{(\mathfrak{M}+)}$.

Lemma 4 *(Boundedness of variance) Denote $\sigma_n^2(\lambda) = \frac{\sigma_0^2}{n} tr[(S_n(\lambda)S_n^{-1}(\lambda_{n0}))' S_n(\lambda) S_n^{-1}(\lambda_{n0})]$. Under Assumptions 1-5, for large enough n,*

(i) $\sigma_n^2(\lambda)$ is bounded away from zero and bounded from above, uniformly in $\lambda \in \Lambda_n^{(\mathfrak{M})}$ under Assumption 6 (3);

(ii) $\frac{1}{[\psi_n(\lambda_{n0} - \lambda)]^2}\sigma_n^2(\lambda)$ is bounded away from zero and bounded from above, uniformly in $\lambda \in \Lambda_n^{(\mathfrak{M}+)}$.

Denote $\Lambda_\kappa = (-1, 1 - \kappa]$, where κ is a small positive number. For any $\kappa > 0$, Λ_κ consists of parameters bounded away from 1 with a fixed distance κ.

Lemma 5 *Under Assumptions 1-5,*

$$\begin{aligned}\frac{1}{\psi_n^2}\sigma_n^2(\lambda) &= \sigma_0^2[\frac{1}{\psi_n^2} + 2(\lambda_{n0} - \lambda)\frac{tr(G_n)}{n\psi_n^2} + (\lambda_{n0} - \lambda)^2\frac{tr(G_n' G_n)}{n\psi_n^2}] \\ &\xrightarrow{p} \sigma_0^2(1 - \lambda)^2 \lim_{n\to\infty}\frac{1}{n} tr(R_n^{'-1} J_n R_n' R_n J_n R_n^{-1}),\end{aligned}$$

where the convergence is uniformly in λ and the limit is bounded away from zero on λ in Λ_κ under Assumption 6 (2).

Lemma 6 *Under Assumptions 1-5,*

$$\begin{aligned}&\frac{1}{n}\ln|\sigma_0^2 S_n^{-1}(\lambda_{n0}) S_n^{'-1}(\lambda_{n0})| - \frac{1}{n}\ln|\sigma_n^2(\lambda) S_n^{-1}(\lambda) S_n^{-1}(\lambda)| \\ =\ &\ln\sigma_0^2 - \frac{2}{n}\ln|S_n(\lambda_{n0})| - \ln\sigma_n^2(\lambda) + \frac{2}{n}\ln|S_n(\lambda)| \\ =\ &2(\frac{m_n}{n} - 1)\ln\psi_n + O(1),\end{aligned}$$

uniformly in $\lambda \in \Lambda_\kappa$.

Proof of Lemma 1: We note that Assumptions 2 and 5 on the uniform boundedness conditions will guarantee that $\frac{1}{\psi_n} S_n^{-1}(\lambda_{n0})$ and $\frac{1}{\psi_n} G_n$ are uniformly bounded in both row and column sum norms. The

results follow from some basic law of large numbers and central limit theorems for linear and quadratic forms (e.g., Kelejian and Prucha (1998) and Lee (2004)). The remaining order follows from the order of G_n, and $(\lambda_{n0}-1)$ and (λ_{n0}^2-1) are also of order $O(\frac{1}{\psi_n})$. The items (1)-(6) are those consequences. The item (7) is also straightforward. It follows from

$$\begin{aligned}(G_nX_n)'M_n(G_nX_n) &= \psi_n^2X_n'R_n^{'-1}[\lambda_{n0}J_n+\frac{1}{\psi_n}(I_n-\lambda_n\tilde{D}_n)^{-1}D_n]R_n' \\ &\quad \times M_nR_n[\lambda_{n0}J_n+\frac{1}{\psi_n}D_n(I_n-\lambda_n\tilde{D}_n)^{-1}]R_n^{-1}X_n.\end{aligned}$$

For (8) and (9), as $G_n=\psi_n[\lambda_nR_nJ_nR_n^{-1}+\frac{1}{\psi_n}R_nD_n(I_n-\lambda_n\tilde{D}_n)^{-1}R_n^{-1}]$ and

$$G_n'G_n=\psi_n^2R_n^{'-1}[\lambda_nJ_n+\frac{1}{\psi_n}(I_n-\lambda_n\tilde{D}_n)^{-1}D_n)R_n'R_nD_n(\lambda_nJ_n+\frac{1}{\psi_n}(I_n-\lambda_n\tilde{D}_n)^{-1})R_n^{-1},$$

we have $\frac{1}{n\psi_n}tr(G_n)=\lambda_n\frac{tr(J_n)}{n}+\frac{1}{n\psi_n}tr(R_nD_n(I_n-\lambda_n\tilde{D}_n)^{-1}R_n^{-1})=\frac{m_n}{n}+O(\frac{1}{\psi_n})$, because the eigenvalues in $\tilde{D}_n$ is bounded away from 1 in absolute value (by Assumption 2). Also, $tr(G_n'G_n)=\frac{1}{n}tr(R_n^{'-1}J_nR_n'R_nJ_nR_n^{-1})+O(\frac{1}{\psi_n})$. ■

Proof of Lemma 2: For (26), using the central limit theorem for linear and quadratic forms in Kelejian and Prucha (2001), we can get the result.

For (27), as $Y_n=\psi_nR_n\Gamma_nR_n^{-1}(X_n\beta_0+\epsilon_n)$ from (8), we have $\frac{1}{n\cdot\psi_n}X_n'B_nY_n=\frac{1}{n}X_n'B_nR_n\Gamma_nR_n^{-1}(X_n\beta_0+\epsilon_n)$ where $\Gamma_n=\Gamma_n^u+\Gamma_n^s$, $\Gamma_n^u=\lambda_{n0}J_n$ and $\Gamma_n^s=\frac{1}{\psi_n}(I_n-\lambda_{n0}\tilde{D}_n)^{-1}$ from (7). By Assumption 5, Γ_n is row sum and column sum bounded. Hence, $\frac{1}{n}X_n'B_nR_n\Gamma_nR_n^{-1}\epsilon_n=O_p\left(\frac{1}{\sqrt{n}}\right)$ from (26), which proves (27).

For (29), as $\frac{1}{n\cdot\psi_n}X_n'B_nY_n=\frac{1}{n}X_n'B_nR_n\Gamma_nR_n^{-1}(X_n\beta_0+\epsilon_n)$ where $\frac{1}{n}X_n'B_nR_n\Gamma_nR_n^{-1}\epsilon_n=O_p\left(\frac{1}{\sqrt{n}}\right)$, we have $\lim_{n\to\infty}\frac{1}{n\cdot\psi_n}X_n'B_nY_n=\lim_{n\to\infty}\frac{1}{n}X_n'B_nR_n\Gamma_nR_n^{-1}X_n\beta_0$. As $\Gamma_n^s=\frac{1}{\psi_n}(I_n-\lambda_{n0}\tilde{D}_n)^{-1}$ and $R_n(I_n-\lambda_{n0}\tilde{D}_n)^{-1}R_n^{-1}$ is row sum and column sum bounded from Assumption 5, $\lim_{n\to\infty}\frac{1}{n}X_n'B_nR_n\Gamma_n^sR_n^{-1}X_n\beta_0=0$ as ψ_n is an increasing function of n. Hence, $\lim_{n\to\infty}\frac{1}{n\cdot\psi_n}X_n'B_nY_n=\lim_{n\to\infty}\frac{1}{n}X_n'B_nR_n\Gamma_n^uR_n^{-1}X_n\beta_0=\lim_{n\to\infty}\frac{1}{n}X_n'B_nR_nJ_nR_n^{-1}X_n\beta_0$.

Similarly, we can establish (28) and (30). ■

Proof of Lemma 3: By expansion,

$$\hat{\sigma}_n^2(\lambda)=\frac{1}{n}Y_n'S_n'(\lambda)M_nS_n(\lambda)Y_n=(\lambda_{n0}-\lambda)^2\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)+2(\lambda_{n0}-\lambda)H_{1n}(\lambda)+H_{2n}(\lambda),$$

where $H_{1n}(\lambda)=\frac{1}{n}(G_nX_n\beta_0)'M_nS_n(\lambda)S_n^{-1}(\lambda_{n0})\epsilon_n$ and $H_{2n}(\lambda)=\frac{1}{n}\epsilon_n'S_n^{'-1}(\lambda_{n0})S_n'(\lambda)M_nS_n(\lambda)S_n^{-1}(\lambda_{n0})\epsilon_n$. As $S_n(\lambda)S_n^{-1}(\lambda_{n0})=I_n+(\lambda_{n0}-\lambda)G_n$, it follows that

$$(\lambda_{n0}-\lambda)H_{1n}(\lambda)=(\lambda_{n0}-\lambda)\frac{1}{n}(G_nX_n\beta_0)'M_n\epsilon_n+(\lambda_{n0}-\lambda)^2\frac{1}{n}(G_nX_n\beta_0)'M_nG_n\epsilon_n.$$

Also, $H_{2n}(\lambda) - \sigma_n^2(\lambda) = T_{2n,1}(\lambda) + T_{2n,2}(\lambda)$, where

$$\begin{aligned} T_{2n,1}(\lambda) &= \frac{1}{n}\epsilon_n' S_n^{'-1}(\lambda_{n0}) S_n'(\lambda) S_n(\lambda) S_n^{-1}(\lambda_{n0})\epsilon_n - \sigma_n^2(\lambda) \\ &= (\frac{1}{n}\epsilon_n'\epsilon_n - \sigma_0^2) + (\lambda_{n0}-\lambda)\frac{1}{n}(\epsilon_n'(G_n+G_n')\epsilon_n - \sigma_0^2 tr(G_n+G_n')) \\ &\quad + (\lambda_{n0}-\lambda)^2\frac{1}{n}(\epsilon_n' G_n' G_n \epsilon_n - \sigma_0^2 tr(G_n' G_n)), \end{aligned}$$

and

$$\begin{aligned} T_{2n,2}(\lambda) &= \frac{1}{n}\epsilon_n' S_n^{'-1}(\lambda_{n0}) S_n'(\lambda) X_n (X_n'X_n)^{-1} X_n' S_n(\lambda) S_n^{-1}(\lambda_{n0})\epsilon_n \\ &= \frac{1}{n}\epsilon_n' X_n (X_n'X_n)^{-1} X_n'\epsilon_n + (\lambda_{n0}-\lambda)\frac{2}{n}\epsilon_n' G_n' X_n (X_n'X_n)^{-1} X_n'\epsilon_n \\ &\quad + (\lambda_{n0}-\lambda)^2\frac{1}{n}\epsilon_n' G_n' X_n (X_n'X_n)^{-1} X_n' G_n \epsilon_n. \end{aligned}$$

(i) On $\Lambda_n^{(\mathfrak{M})}$: By Lemma 1, we have

$$(\lambda_{n0}-\lambda)H_{1n}(\lambda) = [\psi_n(\lambda_{n0}-\lambda)]\frac{1}{n\psi_n}(G_nX_n\beta_0)'M_n\epsilon_n + [\psi_n(\lambda_{n0}-\lambda)]^2\frac{1}{n\psi_n^2}(G_nX_n\beta_0)'M_nG_n\epsilon_n = O_p(\frac{1}{\sqrt{n}}),$$

$$\begin{aligned} T_{2n,1}(\lambda) &= (\frac{1}{n}\epsilon_n'\epsilon_n - \sigma_0^2) + [\psi_n(\lambda_{n0}-\lambda)]\frac{1}{n\psi_n}(\epsilon_n'(G_n+G_n')\epsilon_n - \sigma_0^2 tr(G_n+G_n')) \\ &\quad + [\psi_n(\lambda_{n0}-\lambda)]^2\frac{1}{n\psi_n^2}(\epsilon_n' G_n' G_n\epsilon_n - \sigma_0^2 tr(G_n' G_n)) \\ &= O_p(\frac{1}{\sqrt{n}}) \end{aligned}$$

and

$$\begin{aligned} T_{2n,2}(\lambda) &= \frac{1}{n}\epsilon_n' X_n (X_n'X_n)^{-1} X_n'\epsilon_n + \psi_n(\lambda_{n0}-\lambda)\frac{2}{n\psi_n}\epsilon_n' G_n' X_n (X_n'X_n)^{-1} X_n'\epsilon_n \\ &\quad + [\psi_n(\lambda_{n0}-\lambda)]^2\frac{1}{n\psi_n^2}\epsilon_n' G_n' X_n (X_n'X_n)^{-1} X_n' G_n\epsilon_n \\ &= O_p(\frac{1}{n}), \end{aligned}$$

uniformly in λ when $|\psi_n(\lambda_{n0}-\lambda)| \le \mathfrak{M}$. Hence, we conclude $\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda) = 2(\lambda_{n0}-\lambda)H_{1n}(\lambda) + (H_{2n}(\lambda) - \sigma_n^2(\lambda)) = O_p(\frac{1}{\sqrt{n}})$ uniformly in $\lambda \in \Lambda_n^{(\mathfrak{M})}$.

(ii) On $\Lambda_n^{(\mathfrak{M}+)}$: In this case,

$$\frac{(\lambda_{n0}-\lambda)H_{1n}(\lambda)}{[\psi_n(\lambda_{n0}-\lambda)]^2} = \frac{1}{[\psi_n(\lambda_{n0}-\lambda)]}\frac{1}{n\psi_n}(G_nX_n\beta_0)'M_n\epsilon_n + \frac{1}{n\psi_n^2}(G_nX_n\beta_0)'M_nG_n\epsilon_n = O_p(\frac{1}{\sqrt{n}});$$

$$
\begin{aligned}
& \frac{T_{2n,1}(\lambda)}{[\psi_n(\lambda_{n0}-\lambda)]^2} \\
= & \frac{1}{[\psi_n(\lambda_{n0}-\lambda)]^2}(\frac{1}{n}\epsilon_n'\epsilon_n - \sigma_0^2) + \frac{1}{\psi_n(\lambda_{n0}-\lambda)}\frac{1}{n\psi_n}[\epsilon_n'(G_n+G_n')\epsilon_n - \sigma_0^2 tr(G_n+G_n')] \\
& + \frac{1}{n\psi_n^2}(\epsilon_n' G_n' G_n \epsilon_n - \sigma_0^2 tr(G_n' G_n)) \\
= & O_p(\frac{1}{\sqrt{n}}),
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{T_{2n,2}(\lambda)}{[\psi_n(\lambda_{n0}-\lambda)]^2} \\
= & \frac{1}{[\psi_n(\lambda_{n0}-\lambda)]^2}\frac{1}{n}\epsilon_n' X_n(X_n'X_n)^{-1}X_n'\epsilon_n + \frac{1}{\psi_n(\lambda_{n0}-\lambda)}\frac{2}{n\psi_n}\epsilon_n' G_n' X_n(X_n'X_n)^{-1}X_n'\epsilon_n \\
& + \frac{1}{n\psi_n^2}\epsilon_n' G_n' X_n(X_n'X_n)^{-1}X_n' G_n\epsilon_n \\
= & O_p(\frac{1}{n}),
\end{aligned}
$$

by Lemma 1, uniformly in $\lambda \in \Lambda_n^{(\mathfrak{M}+)}$, because $\frac{1}{[\psi_n(\lambda_{n0}-\lambda)]} \leq \frac{1}{\mathfrak{M}} < \infty$. We conclude $\frac{\hat{\sigma}_n^2(\lambda)-\sigma_n^{*2}(\lambda)}{[\psi_n(\lambda_{n0}-\lambda)]^2} = O_p(\frac{1}{\sqrt{n}})$ uniformly in $\lambda \in \Lambda_n^{(\mathfrak{M}+)}$. ■

Proof of Lemma 4: By Lemma 1 and Assumption 6, the limits of $\frac{1}{n\psi_n}tr(G_n) = \frac{m_n}{n} + O(\frac{1}{\psi_n})$ and $\frac{1}{n\psi_n^2}tr(G_n'G_n) = \frac{1}{n}tr(R_n^{'-1}J_nR_n'R_nJ_nR_n^{-1}) + O(\frac{1}{\psi_n})$ are strictly positive (see (33) and (34)). Recall that $\sigma_n^2(\lambda) = \sigma_0^2[1 + 2(\lambda_{n0}-\lambda)\frac{tr(G_n)}{n} + (\lambda_{n0}-\lambda)^2\frac{tr(G_n'G_n)}{n}]$.

(i) On $\Lambda_n^{(\mathfrak{M})}$: It is apparent that, when $|\psi_n(\lambda_{n0}-\lambda)| \leq \mathfrak{M}$,

$$
\sigma_n^2(\lambda) = \sigma_0^2\{1 + 2[\psi_n(\lambda_{n0}-\lambda)]\frac{tr(G_n)}{n\psi_n} + [\psi_n(\lambda_{n0}-\lambda)]^2\frac{tr(G_n'G_n)}{n\psi_n^2}\},
$$

is bounded from above on $\Lambda_n^{(\mathfrak{M})}$.

The $\sigma_n^2(\lambda)$ can be bounded away from zero on $\Lambda_n^{(\mathfrak{M})}$ as follows. Let λ_n be a sequence in $\Lambda_n^{(\mathfrak{M})}$ such that $\lambda_n = 1 - \frac{1}{\psi_n^*}$ (which is a reparameterization). Because $(\lambda_{n0}-\lambda_n)\psi_n = (\frac{\psi_n}{\psi_n^*}-1)$ and $\lambda_n \in \Lambda_n^{(\mathfrak{M})}$, $|\frac{\psi_n}{\psi_n^*}-1| \leq \mathfrak{M}$. Because $\frac{\psi_n}{\psi_n^*}$ is bounded, without loss of generality, suppose that $\frac{\psi_n}{\psi_n^*} \to c^*$.

As

$$
\begin{aligned}
\sigma_n^2(\lambda) = & \sigma_0^2[1 + 2(\lambda_{n0}-\lambda)\frac{trG_n}{n} + (\lambda_{n0}-\lambda)^2\frac{tr(G_n'G_n)}{n}] \\
= & \frac{\sigma_0^2}{n}tr[(I_n - (\lambda_{n0}-\lambda)\psi_n\cdot\frac{G_n}{\psi_n})(I_n - (\lambda_{n0}-\lambda)\psi_n\cdot\frac{G_n'}{\psi_n})],
\end{aligned}
$$

we have,

$$
\lim_{n\to\infty}\sigma_n^2(\lambda_n) = \sigma_0^2\lim_{n\to\infty}\frac{1}{n}tr[(I_n - (c^*-1)R_nJ_nR_n^{-1})(I_n - (c^*-1)R_nJ_nR_n^{-1})'],
$$

which is positive by Assumption 6 (3). That is, $\sigma_n^2(\lambda)$ is bounded away from zero on $\Lambda_n^{(\mathfrak{M})}$.

(ii) On $\Lambda_n^{(\mathfrak{M}+)}$: We have

$$\begin{aligned}\frac{\sigma_n^2(\lambda)}{[\psi_n(\lambda_{n0}-\lambda)]^2} &= \sigma_0^2\{\frac{1}{[\psi_n(\lambda_{n0}-\lambda)]^2}+2\frac{(\lambda_{n0}-\lambda)}{[\psi_n(\lambda_{n0}-\lambda)]^2}\frac{tr(G_n)}{n}+\frac{(\lambda_{n0}-\lambda)^2}{[\psi_n(\lambda_{n0}-\lambda)]^2}\frac{tr(G_n'G_n)}{n})\}\\ &= \sigma_0^2\{\frac{1}{[\psi_n(\lambda_{n0}-\lambda)]^2}+\frac{2}{\psi_n(\lambda_{n0}-\lambda)}\frac{tr(G_n)}{n\psi_n}+\frac{tr(G_n'G_n)}{n\psi_n^2})\}.\end{aligned}$$

As $(\lambda_{n0}-\lambda)>0$ on $\Lambda_n^{(\mathfrak{M}+)}$, for large enough n, $\frac{\sigma_n^2(\lambda)}{[\psi_n(\lambda_{n0}-\lambda)]^2} \geq \sigma_0^2\frac{tr(G_n'G_n)}{n\psi_n^2}$ as the limits of $\frac{tr(G_n)}{n\psi_n}$ and $\frac{tr(G_n'G_n)}{n\psi_n^2}$ are positive by Lemma 1. On the other hand, as $\psi_n(\lambda_{n0}-\lambda)\geq\mathfrak{M}$, one has, for large enough n, $\frac{\sigma_n^2(\lambda)}{[\psi_n(\lambda_{n0}-\lambda)]^2}\leq\sigma_0^2[\frac{1}{\mathfrak{M}^2}+\frac{2}{\mathfrak{M}}\frac{tr(G_n)}{n\psi_n}+\frac{tr(G_n'G_n)}{n\psi_n^2}]$, which is bounded from above. ■

Proof of Lemma 5: Because $\sigma_n^2(\lambda)=\sigma_0^2[1+2(\lambda_{n0}-\lambda)\frac{tr(G_n)}{n}+(\lambda_{n0}-\lambda)^2\frac{tr(G_n'G_n)}{n}]$, it is sufficient to investigate $tr(G_n)$ and $tr(G_n'G_n)$. As $G_n=\psi_n[\lambda_nR_nJ_nR_n^{-1}+\frac{1}{\psi_n}R_nD_n(I_n-\lambda_n\tilde{D}_n)^{-1}R_n^{-1}]$ and

$$G_n'G_n=\psi_n^2R_n^{'-1}[\lambda_nJ_n+\frac{1}{\psi_n}(I_n-\lambda_n\tilde{D}_n)^{-1}D_n']R_n'R_n[(\lambda_nJ_n+\frac{1}{\psi_n}D_n(I_n-\lambda_n\tilde{D}_n)^{-1}]R_n^{-1},$$

we have

$$\frac{1}{n\psi_n}tr(G_n)=\lambda_n\frac{tr(J_n)}{n}+\frac{1}{n\psi_n}tr(R_nD_n(I_n-\lambda_n\tilde{D}_n)^{-1}R_n^{-1})=\frac{tr(J_n)}{n}+O(\frac{1}{\psi_n})=\frac{m_n}{n}+O(\frac{1}{\psi_n}), \quad (33)$$

and

$$\frac{1}{n\psi_n^2}tr(G_n'G_n)=\frac{1}{n}tr(R_n^{'-1}J_nR_n'R_nJ_nR_n^{-1})+O(\frac{1}{\psi_n}) \quad (34)$$

because $\lambda_n=1-\frac{1}{\psi_n}$. Hence,

$$\frac{\sigma_n^2(\lambda)}{\psi_n^2}=\sigma_0^2[\frac{1}{\psi_n^2}+2(\lambda_{n0}-\lambda)\frac{tr(G_n)}{n\psi_n^2}+(\lambda_{n0}-\lambda)^2\frac{tr(G_n'G_n)}{n\psi_n^2})]\to\sigma_0^2(1-\lambda)^2\lim_{n\to\infty}\frac{1}{n}tr(R_n^{'-1}J_nR_n'R_nJ_nR_n^{-1}).$$

Proof of Lemma 6: As $S_n(\lambda)=R_n(I_n-\lambda D_n)^{-1}R_n^{-1}$ and there are m_n roots of unity, $|S_n(\lambda_{n0})|=(1-\lambda_{n0})^{m_n}\prod_{j=m_n+1}^n(1-\lambda_{n0}d_{nj})$ where d_{nj}'s are bounded away from 1. Thus,

$$\ln|S_n(\lambda_{n0})|=-m_n\ln\psi_n+\sum_{j=m_n+1}^n\ln(1-\lambda_{n0}d_{nj}). \quad (35)$$

Hence,

$$\begin{aligned}&\frac{1}{n}\ln|\sigma_0^2S_n^{-1}(\lambda_{n0})S_n^{'-1}(\lambda_{n0})|-\frac{1}{n}\ln|\sigma_n^2(\lambda)S_n^{-1}(\lambda)S_n^{'-1}(\lambda)|\\ =\;&\ln\sigma_0^2-\frac{2}{n}\ln|S_n(\lambda_{n0})|-\ln\sigma_n^2(\lambda)+\frac{2}{n}\ln|S_n(\lambda)|\\ =\;&-\frac{2}{n}\ln|S_n(\lambda_{n0})|-\ln(\frac{\sigma_n^2(\lambda)}{\psi_n^2})-2\ln\psi_n+\ln\sigma_0^2+\frac{2}{n}\ln|S_n(\lambda)|\\ =\;&2(\frac{m_n}{n}-1)\ln\psi_n+O(1),\end{aligned}$$

because $\frac{\sigma_n^2(\lambda)}{\psi_n^2}$ is bounded away from 0 on Λ_κ and $\frac{1}{n}\ln|S_n(\lambda)|$ is $O(1)$ when λ is bounded away from the unity. ■

B Proofs for Theorems

B.1 Proof for Theorem 1

We would consider the ψ_n-consistency of the QMLE as the rate ψ_n plays an important role in the analysis of its asymptotic distribution. Denote $\hat{\lambda}_n$ the QMLE. The QML approach is theoretically defined as $\max_{\lambda\in\Lambda}\ln L_n(\lambda)$. To prove that $\hat{\lambda}_n$ is ψ_n-consistent, a sufficient condition for consistency of an extremum estimator in Wu (1981, Lemma 1) can be extended. For any $\epsilon > 0$, we would like to show that

$$\lim_{n\to\infty} P(\inf_{\lambda\in\Lambda,|\psi_n(\lambda-\lambda_{n0})|\geq\epsilon}[\ln L_n(\lambda_{n0}) - \ln L_n(\lambda)] > 0) = 1. \tag{36}$$

This is a sufficient condition which guarantees that $|\psi_n(\hat{\lambda}_n - \lambda_{n0})| \xrightarrow{p} 0$.

The concentrated log likelihood of our model is $\ln L_n(\lambda)$ in (13) and the corresponding concentrated expected log likelihood is $Q_n(\lambda)$ in (15). As

$$\frac{1}{n}[\ln L_n(\lambda_{n0}) - \ln L_n(\lambda)] = \frac{1}{n}\{[\ln L_n(\lambda_{n0}) - Q_n(\lambda_{n0})] - [\ln L_n(\lambda) - Q_n(\lambda)] + [Q_n(\lambda_{n0}) - Q_n(\lambda)]\},$$

we shall investigate the behavior of each term. We are going to show that

$$(i)\ \frac{1}{n}[\ln L_n(\lambda_{n0}) - Q_n(\lambda_{n0})] \xrightarrow{p} 0, \tag{37}$$

$$(ii)\ \sup_{\lambda\in\Lambda}\frac{1}{n}|\ln L_n(\lambda) - Q_n(\lambda)| \xrightarrow{p} 0, \tag{38}$$

$$(iii)\ \liminf_{n\to\infty}\inf_{\lambda\in\Lambda,|\psi_n(\lambda-\lambda_{n0})|\geq\epsilon}[Q_n(\lambda_{n0}) - Q_n(\lambda)] > 0. \tag{39}$$

For (i), it is trivial. At λ_{n0}, $\sigma_n^{*2}(\lambda_{n0}) = \sigma_0^2$ and $\hat{\sigma}_n^2(\lambda_{n0}) = \frac{1}{n}\epsilon_n' M_n\epsilon_n$. By the law of large numbers for independent variables and the continuity mapping theorem, $\frac{1}{n}[\ln L_n(\lambda_{n0}) - Q_n(\lambda_{n0})] = -\frac{1}{2}[\ln(\frac{1}{n}\epsilon_n' M_n\epsilon_n) - \ln\sigma_0^2] \xrightarrow{p} 0$. For (ii), $\frac{1}{n}[\ln L_n(\lambda) - Q_n(\lambda)] = -\frac{1}{2}[\ln\hat{\sigma}_n^2(\lambda) - \ln\sigma_n^{*2}(\lambda)] = -\frac{1}{2}\frac{1}{\tilde{\sigma}_n^2(\lambda)}[\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda)]$, where $\tilde{\sigma}_n^2(\lambda)$ lies between $\hat{\sigma}_n(\lambda)$ and $\sigma_n^{*2}(\lambda)$ by the mean value theorem. The analysis follows by investigating possible uniform convergence of $\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda)$ to zero on Λ and whether $\tilde{\sigma}_n^2(\lambda)$ is uniformly bounded away from zero on Λ. The (iii) is the identification uniqueness condition. From the expression of $Q_n(\lambda)$, it follows that

$$\begin{aligned}
& \frac{2}{n}[Q_n(\lambda) - Q_n(\lambda_{n0})] \\
= \ & [\ln\sigma_n^{*2}(\lambda_{n0}) - \ln\sigma_n^{*2}(\lambda)] + \frac{2}{n}[\ln|S_n(\lambda)| - \ln|S_n(\lambda_{n0})|] \\
= \ & \ln\sigma_0^2 - \ln[(\lambda_{n0}-\lambda)^2\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \sigma_n^2(\lambda)] + \frac{2}{n}[\ln|S_n(\lambda)| - \ln|S_n(\lambda_{n0})|] \\
= \ & d_{n1}(\lambda) - d_{n2}(\lambda),
\end{aligned}$$

where

$$d_{n1}(\lambda) = \ln\sigma_0^2 - \frac{2}{n}\ln|S_n(\lambda_{n0})| - \ln\sigma_n^2(\lambda) + \frac{2}{n}\ln|S_n(\lambda)|, \tag{40}$$

$$d_{n2}(\lambda) = \ln[(\lambda_{n0}-\lambda)^2\frac{1}{n\sigma_n^2(\lambda)}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)+1]. \tag{41}$$

By the information inequality for a pure SAR process as in Lee (2004), $d_{n1}(\lambda) \leq 0$ for all λ where the likelihood function is well defined. It is apparent that $d_{n2}(\lambda) > 0$. However, we need to establish that the limit does not vanish.

For (ii) and (iii), it is desirable to analyze these properties in two parts of the relevant parameter set, namely, $\Lambda_n^{(\mathfrak{M}+)}$ and $\Lambda_n^{(\mathfrak{M},\epsilon)}$ where $\Lambda_n^{(\mathfrak{M}+)}$ in defined in (31) and $\Lambda_n^{(\mathfrak{M},\epsilon)}$ is defined as

$$\Lambda_n^{(\mathfrak{M},\epsilon)} = \{\lambda|\lambda\in\Lambda,\ \epsilon\leq|\psi_n(\lambda_{n0}-\lambda)|\leq\mathfrak{M}\}. \tag{42}$$

On $\Lambda_n^{(\mathfrak{M}+)}$: We shall establish that $\lim_{n\to\infty}P(\inf_{\lambda\in\Lambda_n^{(\mathfrak{M}+)}}[\ln L_n(\lambda_{n0})-\ln L_n(\lambda)]>0)=1$.

Lemma 3 provides the uniform convergence of $\frac{1}{[\psi_n(\lambda_{n0}-\lambda)]^2}(\hat\sigma_n^2(\lambda)-\sigma_n^{*2}(\lambda))$ to zero on $\Lambda_n^{(\mathfrak{M}+)}$. As $\sigma_n^{*2}(\lambda)-\sigma_n^2(\lambda)=(\lambda_{n0}-\lambda)^2(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$, it follows that $\sigma_n^{*2}(\lambda)\geq\sigma_n^2(\lambda)$. Lemma 4 implies that, there exists a $c>0$ such that $\frac{\sigma_n^2(\lambda)}{[\psi_n(\lambda_{n0}-\lambda)]^2}\geq c$ for all n and $\lambda\in\Lambda_n^{(\mathfrak{M}+)}$. By combining this and the preceding uniform convergence, it follows that $\frac{1}{[\psi_n(\lambda_{n0}-\lambda)]^2}\tilde\sigma_n^2(\lambda)$ is bounded away from zero in probability uniformly in $\lambda\in\Lambda_n^{(\mathfrak{M}+)}$. Therefore,

$$\frac{1}{n}[\ln L_n(\lambda)-Q_n(\lambda)] = -\frac{1}{2}(\frac{[\psi_n(\lambda_{n0}-\lambda)]^2}{\tilde\sigma_n^2(\lambda)})\cdot\frac{1}{[\psi_n(\lambda_{n0}-\lambda)]^2}[\hat\sigma_n^2(\lambda)-\sigma_n^{*2}(\lambda)] = O_p(\frac{1}{\sqrt{n}}),$$

uniformly in $\lambda\in\Lambda_n^{(\mathfrak{M}+)}$. This establishes (ii) on $\Lambda_n^{(\mathfrak{M}+)}$.

For (iii), Lemma 4 and Assumption 7 imply that

$$(\lambda_{n0}-\lambda)^2\frac{1}{n\sigma_n^2(\lambda)}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) = \frac{[\psi_n(\lambda_{n0}-\lambda)]^2}{\sigma_n^2(\lambda)}\frac{(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)}{n\psi_n^2}$$

is strictly positive and bounded away from zero in the limit. It follows that $\lim_{n\to\infty}\sup_{\lambda\in\Lambda_n^{(\mathfrak{M}+)}}d_{2n}(\lambda)>0$. Hence, $\liminf_{n\to\infty}\inf_{\lambda\in\Lambda_n^{(\mathfrak{M}+)}}\frac{1}{n}(Q_n(\lambda_{n0})-Q_n(\lambda))>0$. This proves (iii) on $\Lambda_n^{(\mathfrak{M}+)}$.

On $\Lambda_n^{(\mathfrak{M},\epsilon)}$: We shall establish that $\lim_{n\to\infty}P(\inf_{\lambda\in\Lambda_n^{(\mathfrak{M},\epsilon)}}[\ln L_n(\lambda_{n0})-\ln L_n(\lambda)]>0)=1$.

Lemma 3 implies the uniform convergence of $(\hat\sigma_n^2(\lambda)-\sigma_n^{*2}(\lambda))$ to zero on $\Lambda_n^{(\mathfrak{M})}$. Lemma 4 implies that, there exists a $c>0$ such that $\sigma_n^2(\lambda)\geq c$ for all n and $\lambda\in\Lambda_n^{(\mathfrak{M})}$. As $\sigma_n^{*2}(\lambda)\geq\sigma_n^2(\lambda)$, $\tilde\sigma_n^2(\lambda)$ is bounded away from zero in probability uniformly in $\lambda\in\Lambda_n^{(\mathfrak{M})}$. Therefore, $\frac{1}{n}[\ln L_n(\lambda)-Q_n(\lambda)] = -\frac{1}{2\tilde\sigma_n^2(\lambda)}\cdot[\hat\sigma_n^2(\lambda)-\sigma_n^{*2}(\lambda)] = O_p(\frac{1}{\sqrt{n}})$, uniformly in $\lambda\in\Lambda_n^{(\mathfrak{M})}$. This establishes (ii).

For (iii), $d_{n2}(\lambda)=\ln\{[\psi_n(\lambda_{n0}-\lambda)]^2\frac{1}{\sigma_n^2(\lambda)n\psi_n^2}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)+1\}$. From Lemma 1 and under Assumption 7, $\frac{1}{n\psi_n^2}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ is greater than zero in the limit. The $\sigma_n^2(\lambda)$ is bounded from

above by Lemma 4, and, on $\Lambda_n^{(\mathfrak{M},\epsilon)}$, $|\psi_n(\lambda_{n0}-\lambda)| \geq \epsilon > 0$. Thus the identification uniqueness condition is satisfied on $\lambda \in \Lambda_n^{(\mathfrak{M},\epsilon)}$.

With all the above arguments together, we have established the property in (36), which is a sufficient condition to guarantee that $|\psi_n(\hat{\lambda}_n - \lambda_{n0})| \xrightarrow{p} 0$. This is so, if $\hat{\lambda}_n$ is not ψ_n-consistent, then there would exist a $\bar{\epsilon} > 0$ such that $\lim_{n\to\infty} P(|\psi_n(\hat{\lambda}_n - \lambda_{n0})| \geq \bar{\epsilon})$ is positive. As $\hat{\lambda}_n$ maximizes $\ln L_n(\lambda)$ in Λ, $\ln L_n(\lambda_{n0}) - \ln L_n(\hat{\lambda}_n) \leq 0$. Then, $\lim_{n\to\infty} P(\inf_{\lambda\in\Lambda, |\psi_n(\lambda-\lambda_{n0})|\geq\bar{\epsilon}}[\ln L_n(\lambda_{n0}) - \ln L_n(\lambda)] \leq 0)$ is not zero, a contradiction. Thus, we have shown that $\hat{\lambda}_n$ is a ψ_n-consistent estimate of λ_{n0}.

Using (11), we have

$$\begin{aligned}\hat{\beta}_n &= \beta_0 + (X_n'X_n)^{-1}X_n'(I_n - (\hat{\lambda}_n - \lambda_{n0})G_n)\epsilon_n \\ &= \beta_0 + (\frac{1}{n}X_n'X_n)^{-1}\left[\frac{1}{n}X_n'\epsilon_n - \psi_n(\hat{\lambda}_n - \lambda_{n0})\frac{1}{n\psi_n}X_n'G_n\epsilon_n\right].\end{aligned}$$

As $\hat{\lambda}_n$ is ψ_n-consistent, $\hat{\beta}_n$ is consistent. Also, from (12), we have $\hat{\sigma}_n^2 = \frac{1}{n}(Y_n - Z_n\hat{\delta}_n)'(Y_n - Z_n\hat{\delta}_n)$ where $\hat{\delta}_n = (\hat{\beta}_n', \hat{\lambda}_n)'$. As $Y_n = Z_n\delta_{n0} + \epsilon_n$ and $\Upsilon_{\delta,n} = \begin{pmatrix} I_k & \mathbf{0} \\ \mathbf{0} & \psi_n \end{pmatrix}$, we have,

$$\begin{aligned}\hat{\sigma}_n^2 &= \frac{1}{n}(Z_n(\delta_{n0} - \hat{\delta}_n) + \epsilon_n)'(Z_n(\delta_{n0} - \hat{\delta}_n) + \epsilon_n) \\ &= \frac{1}{n}(Z_n\Upsilon_{\delta,n}^{-1}\cdot\Upsilon_{\delta,n}(\delta_{n0} - \hat{\delta}_n) + \epsilon_n)'(Z_n\Upsilon_{\delta,n}^{-1}\cdot\Upsilon_{\delta,n}(\delta_{n0} - \hat{\delta}_n) + \epsilon_n).\end{aligned}$$

Using Lemma 2 and the consistency rates of $\hat{\delta}_n$, the consistency of $\hat{\sigma}_n^2$ follows. ■

B.2 Proof for Theorem 2

When Assumption 7 fails, (37) and (38) still hold. However, the identification uniqueness condition (39), will rely on the behavior of $d_{n1}(\lambda)$ with $\lambda \in \Lambda$ such that $|\psi_n(\lambda_{n0} - \lambda)| \geq \epsilon$ for any $\epsilon > 0$, instead of $d_{n2}(\lambda)$, where, from (40), $d_{n1}(\lambda) = \ln\sigma_0^2 - \frac{2}{n}\ln|S_n(\lambda_{n0})| - \ln\sigma_n^2(\lambda) + \frac{2}{n}\ln|S_n(\lambda)|$. In Appendix B.1, $d_{n1}(\lambda) \leq 0$. When Assumption 7 fails, we need to show that the limit of $d_{n1}(\lambda)$ is bounded away from zero. We can show that when $0 < \lim_{n\to\infty}\frac{m_n}{n} \leq 1/2$ or $\lim_{n\to\infty}\frac{m_n}{n} \leq \lim_{n\to\infty}\frac{1}{n}tr(R_n'^{-1}J_nR_n'R_nJ_nR_n^{-1})$, $\lim_{n\to\infty} d_{n1}(\lambda) < 0$.

First, we can divide the situations into two cases.

Situation 1: Consider sequence λ_n such that $\lambda_n \to \lambda^*$ where λ^* is a point less than 1, i.e., the sequence λ_n is bounded away from one.

For this case, from the proof of Lemma 6, we can show that $\frac{1}{n}\sum_{j=m_n+1}^{n}[\ln(1-\lambda_{n0}d_{nj}) - \ln(1-\lambda_n d_{nj})]$ will be dominated by the other terms in $d_{n1}(\lambda_n)$. Using Lemma 5 and Lemma 6, $d_{n1}(\lambda_n)$ is negative and bounded away from zero. These mean that no sequences which can converge to a limit value less than 1 can not be distinguished from the true sequence λ_n in terms of the identification uniqueness.

Situation 2: Consider a sequence λ_n which converges to 1 such that $|\psi_n(\lambda_{n0}-\lambda_n)| \geq \epsilon$ for some $\epsilon > 0$. We want to see under what conditions that $d_{n1}(\lambda_n)$ is bounded away from 0 (strictly negative).

From (40), $d_{n1}(\lambda_n) = \ln\sigma_0^2 - \frac{2}{n}\ln|S_n(\lambda_{n0})| - \ln(\sigma_n^2(\lambda_n)) + \frac{2}{n}\ln|S_n(\lambda_n)|$. Also, using $|S_n(\lambda_n)| = (1-\lambda_n)^{m_n}\prod_{j=m_n+1}^{n}(1-\lambda_n d_{nj})$, we have $\frac{1}{n}\ln|S_n(\lambda_{n0})| - \frac{1}{n}\ln|S_n(\lambda_n)| = \frac{m_n}{n}\ln\frac{1-\lambda_{n0}}{1-\lambda_n} + \frac{1}{n}\sum_{j=m_n+1}^{n}\ln(\frac{1-\lambda_{n0}d_{nj}}{1-\lambda_n d_{nj}})$. As $\frac{1-\lambda_{n0}}{1-\lambda_n} = \frac{1}{1+\psi_n(\lambda_{n0}-\lambda_n)}$ using $1-\lambda_{n0} = \frac{1}{\psi_n}$, we have

$$d_{n1}(\lambda_n) = \ln\sigma_0^2 - \ln\sigma_n^2(\lambda_n) + 2\frac{m_n}{n}\ln(1+\psi_n(\lambda_{n0}-\lambda_n)) - \frac{2}{n}\sum_{j=m_n+1}^{n}\ln(\frac{1-\lambda_{n0}d_{nj}}{1-\lambda_n d_{nj}}). \tag{43}$$

For this case, first we can show that $\frac{1}{n}\sum_{j=m_n+1}^{n}\ln(\frac{1-\lambda_{n0}d_{nj}}{1-\lambda_n d_{nj}})$ will go to zero, so it can be ignored. Intuitively, we can show that $\frac{1-\lambda_{n0}d_{nj}}{1-\lambda_n d_{nj}}$ converge to 1 uniformly in j. That is, for any $c > 0$, we want to show that for large enough n,

$$1 - c \leq \frac{1-\lambda_{n0}d_{nj}}{1-\lambda_n d_{nj}} \leq 1 + c \tag{44}$$

for all j. The (44) is equivalent to $-c \leq [(1-c)\lambda_n - \lambda_{n0}]d_{nj}$ and $[-\lambda_{n0} + (1+c)\lambda_n]d_{nj} \leq c$ for all j, which hold because both λ_{n0} and λ_n go to 1 and $|d_{nj}|$ are bounded away from 1.

By ignoring the vanishing terms including those in $\frac{tr(G_n)}{n\psi_n}$ and $\frac{tr(G_n'G_n)}{n\psi_n^2}$, consider the remaining terms of $d_{1n}(\lambda_n)$, which is

$$d_{n1}^*(\lambda_n) = -\ln\{1 + 2[\psi_n(\lambda_{n0}-\lambda_n)]q_{n1} + [\psi_n(\lambda_{n0}-\lambda_n)]^2 q_{n2}\} + 2q_{n1}\ln[\psi_n(\lambda_{n0}-\lambda_n)+1],$$

where $q_{n1} = \frac{m_n}{n}$ and $q_{n2} = \frac{1}{n}tr(R_n^{'-1}J_n R_n' R_n J_n R_n^{-1})$. By denoting $s_n^* = \psi_n(\lambda_{n0}-\lambda_n)$, we see that the identification uniqueness condition can be stated as "there does not exist a positive sequence s_n^* bounded away from zero such that

$$2q_{n1}\ln(1+s_n^*) - \ln(1 + 2q_{1n}s_n^* + q_{2n}s_n^*) \to 0$$

as $n \to \infty$". So, in particular, we will not have identification uniqueness if there exists an $s^* > 0$, which can be either a finite or the positive infinity value, such that $2q_1\ln(1+s^*) - \ln(1+2q_1 s^* + q_2 s^{*2}) = 0$, where $q_1 = \lim_{n\to\infty} q_{n1}$ and $q_2 = \lim_{n\to\infty} q_{n2}$, equivalently, $(1+s^*)^{2q_1} = (1+2q_1 s^* + q_2 s^{*2})$. Define the functions $R(s) = 1 + 2q_1 s + q_2 s^2$ and $L(s) = (1+s)^{2q_1}$ for $s \in [0,\infty)$. So if these two functions interact at a point different to zero or they have the same asymptotic value, then the identification uniqueness condition will not hold. Otherwise, identification uniqueness condition will be satisfied.

At $s = 0$, $R(0) = L(0) = 1$, so the two curves have the same starting value. As $\frac{dR(s)}{ds} = 2q_1$ and $\frac{dL(s)}{ds} = 2q_1(1+s)^{(2q_1-1)}$, both functions $R(s)$ and $L(s)$ are increasing functions on $s \geq 0$. They have the same slope at $s = 0$ with $\frac{dR(0)}{ds} = \frac{dL(0)}{ds} = 2q_1$. Because q_2 is positive, $R(s)$ is apparently an increasing quadratic function (strictly convex) on $s \geq 0$ as $\frac{d^2R(s)}{ds^2} = 2q_2$. The second order derivative of $L(s)$ is $\frac{d^2L(s)}{ds^2} = 4q_1\frac{(q_1-\frac{1}{2})}{(1+s)^{2(1-q_1)}}$.

The following two cases are the cases of which the identification uniqueness conditions can be satisfied for $d_{n1}(\lambda)$.

(1) $0 < q_1 \leq \frac{1}{2}$:

If $q_1 < \frac{1}{2}$, $\frac{d^2L(s)}{ds^2} < 0$ and $L(s)$ is increasing but a strictly concave function. At $q_1 = \frac{1}{2}$, $L(s)$ is linear. So, when $q_1 \leq \frac{1}{2}$, $R(s)$ will dominate $L(s)$ and they can not interact at any point of s rather than $s = 0$. So, we can conclude that if $q_1 \leq \frac{1}{2}$, the identification uniqueness condition will be satisfied[9].

(2) $q_1 \leq q_2$:

The difference $R(s)$ and $L(s)$ is $R(s) - L(s) = (1 + 2q_1 s + q_2 s^2) - (1 + s)^{2q_1}$ and its derivative is

$$\begin{aligned}\frac{d[R(s) - L(s)]}{ds} &= 2q_1 + 2q_2 s - 2q_1(1 + s)^{(2q_1 - 1)} \\ &= 2q_1[1 + s - (1 + s)^{(2q_1 - 1)}] + 2(q_2 - q_1)s \\ &= 2q_1(1 + s)^{(2q_1 - 1)}[(1 + s)^{2(1 - q_1)} - 1] + 2(q_2 - q_1)s > 0,\end{aligned}$$

for $s > 0$. Because $R(0) = L(0)$, this implies that $R(s) > L(s)$ for all $s > 0$. The difference of $R(s)$ and $L(s)$ is increasing so they are not convergent together at the infinity too.

Hence, $\lim_{n\to\infty} d_{n1}(\lambda_n) < 0$ under situation 1; and it will also hold under situation 2 if $0 < \lim_{n\to\infty} \frac{m_n}{n} \leq 1/2$ or $\lim_{n\to\infty} \frac{m_n}{n} \leq \lim_{n\to\infty} \frac{1}{n} tr(R_n'^{-1} J_n R_n' R_n J_n R_n^{-1})$. Therefore, the identification uniqueness condition (39) is proved. Combined with (37) and (38), we get the result. ■

B.3 Proof for Theorem 3

Similarly as Lee (2004), we have $\Upsilon_{\theta,n}^{-1} \frac{1}{n} \frac{\partial^2 \ln L_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'} \Upsilon_{\theta,n}^{-1} - \Upsilon_{\theta,n}^{-1} \frac{1}{n} \frac{\partial^2 \ln L_n(\theta_{n0})}{\partial\theta\partial\theta'} \Upsilon_{\theta,n}^{-1} \xrightarrow{p} 0$ and $\Upsilon_{\theta,n}^{-1} \frac{1}{n} \frac{\partial^2 \ln L_n(\theta_{n0})}{\partial\theta\partial\theta'} \Upsilon_{\theta,n}^{-1} - E\Upsilon_{\theta,n}^{-1} \frac{1}{n} \frac{\partial^2 \ln L_n(\theta_{n0})}{\partial\theta\partial\theta'} \Upsilon_{\theta,n}^{-1} \xrightarrow{p} 0$. Using the central limit theorem for linear and quadratic form of ϵ_n, we have $\Upsilon_{\theta,n}^{-1} \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_{n0})}{\partial\theta} \xrightarrow{d} N(0, \Sigma + \Omega)$ where $\Sigma = \lim_{n\to\infty} \Upsilon_{\theta,n}^{-1} \left(-\frac{1}{n} E \frac{\partial^2 \ln L_n(\theta_{n0})}{\partial\theta\partial\theta'} \right) \Upsilon_{\theta,n}^{-1}$ and $\Omega = \lim_{n\to\infty} \Upsilon_{\theta,n}^{-1} \Omega_n \Upsilon_{\theta,n}^{-1}$. Assumption 9 guarantees that Σ is nonsingular, hence, the distribution of $\hat{\theta}_n - \theta_{n0}$ comes from the Taylor expansion

$$\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0}) = -\left(\Upsilon_{\theta,n}^{-1} \frac{1}{n} \frac{\partial^2 \ln L_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'} \Upsilon_{\theta,n}^{-1} \right)^{-1} \Upsilon_{\theta,n}^{-1} \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_{n0})}{\partial\theta},$$

such that $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0}) \xrightarrow{d} N(0, \Sigma^{-1} + \Sigma^{-1}\Omega\Sigma^{-1})$. ■

[9] When $\frac{1}{2} < q_1 \leq 1$, $L(s)$ will also be an increasing convex function as $2q_1 > 1$. Even though the curvature of $L(s)$ with $2q_1 < 2$ is smaller than that of the quadratic function $R(s)$, when q_2 is smaller than q_1, these two curves might intersect at some small positive value of s. The identification uniqueness condition might fail if $\frac{1}{2} < q_1 \leq 1$. Whether $R(s)$ and $L(s)$ will intersect with each other at a point different from $s = 0$ will depend on specific values of q_1 and q_2 when $q_2 < q_1$.

B.4 Proof for Theorem 4

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_{n,2sls}-\beta_0) \\ \sqrt{n}\psi_n(\hat{\lambda}_{n,2sls}-\lambda_{n0}) \end{pmatrix}$$
$$= \left[\left(\frac{1}{n}\Upsilon_{\delta,n}^{-1}Z_n'H_n\right)\left(\frac{1}{n}H_n'H_n\right)^{-1}\left(\frac{1}{n}H_n'Z_n\Upsilon_{\delta,n}^{-1}\right)\right]^{-1}\left[\left(\frac{1}{n}\Upsilon_{\delta,n}^{-1}Z_n'H_n\right)\left(\frac{1}{n}H_n'H_n\right)^{-1}\left(\frac{1}{\sqrt{n}}H_n'\epsilon_n\right)\right].$$

Using Assumptions 10 and central limit theorem for $\frac{1}{\sqrt{n}}H_n'\epsilon_n$, we can get the result for $\hat{\delta}_{n,2sls}$. For the distribution of $\hat{\sigma}^2_{n,2sls}$, we have $\hat{\sigma}^2_{n,2sls}=\frac{1}{n}(Y_n-Z_n\hat{\delta}_n)'(Y_n-Z_n\hat{\delta}_n)$ where $\hat{\delta}_n=(Z_n'P_{H_n}Z_n)^{-1}Z_n'P_{H_n}Y_n$. As $Y_n=Z_n\delta_{n0}+\epsilon_n$, we have, $\hat{\sigma}^2_{n,2sls}=\frac{1}{n}(Z_n(\delta_{n0}-\hat{\delta}_n)+\epsilon_n)'(Z_n(\delta_{n0}-\hat{\delta}_n)+\epsilon_n)$. Hence, $\frac{1}{\sqrt{n}}(\hat{\sigma}^2_{n,2sls}-\sigma_0^2)=\frac{1}{\sqrt{n}}(\epsilon_n'\epsilon_n-n\sigma_0^2)+o_p(1)$ using $\Upsilon_{\delta,n}$-consistency of $\hat{\delta}_{n,2sls}$, where the variance of $\epsilon_n'\epsilon_n$ is $n[2\sigma_0^4+(\mu_4-3\sigma_0^4)]$. The result follows. ■

B.5 Proof for Theorem 5

This is similar as the proof of Theorem 1 in Lee (2003). ■

B.6 Proof for Theorem 6

From Taylor expansion, $\frac{\partial \ln L_n(\hat{\theta}_{n1})}{\partial\theta}=\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta}+\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'}(\hat{\theta}_{n1}-\theta_{n0})$ where $\bar{\theta}_n$ lies between $\hat{\theta}_{n1}$ and θ_{n0}. Hence, as $\hat{\theta}_{n2}=\hat{\theta}_{n1}-\left(\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial \ln L_n(\hat{\theta}_{n1})}{\partial\theta}$, we have

$$\hat{\theta}_{n2}=\hat{\theta}_{n1}-\left(\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta}+\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'}(\hat{\theta}_{n1}-\theta_{n0})\right).$$

Therefore,

$$\begin{aligned}\hat{\theta}_{n2}-\theta_{n0} &= \hat{\theta}_{n1}-\theta_{n0}-\left(\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta}+\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'}(\hat{\theta}_{n1}-\theta_{n0})\right)\\ &= \left(I_{k_\theta}-\left(\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'}\right)(\hat{\theta}_{n1}-\theta_{n0})-\left(\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta}.\end{aligned}$$

Using Proposition 2.4 in Lee (2001),

$$\Upsilon_{\theta,n}\left(\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1}=\left(\frac{1}{n}\Upsilon_{\theta,n}^{-1}\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1}\right)^{-1}\Upsilon_{\theta,n}^{-1}\frac{1}{n}\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1}\xrightarrow{p}I_{k_\theta}.$$

As $\Upsilon_{\theta,n}(\hat{\theta}_{n1}-\theta_{n0})=o_p(1)$, we have

$$\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_{n2}-\theta_{n0})=\left(-\frac{1}{n}\Upsilon_{\theta,n}^{-1}\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1}\right)^{-1}\frac{1}{\sqrt{n}}\Upsilon_{\theta,n}^{-1}\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta}+o_p(1). \tag{45}$$

As $\frac{\partial \ln L_n(\theta_{n0})}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma_0^2} X_n' V_n \\ \frac{1}{\sigma_0^2}(W_n Y_n)' V_n - tr(G_n) \\ \frac{1}{2\sigma_0^4}(V_n' V_n - n\sigma_0^2) \end{pmatrix}$, using Lemma 2 and relevant martingale difference CLT (see Theorem 1 in Kelejian and Prucha (2001)), we have $\frac{1}{\sqrt{n}} \Upsilon_{\theta,n}^{-1} \frac{\partial \ln L_n(\theta_{n0})}{\partial \theta} \xrightarrow{p} N(0, \Sigma + \Omega)$ where $\Sigma = \lim_{n\to\infty} \left(-\frac{1}{n} \Upsilon_{\theta,n}^{-1} \frac{\partial^2 \ln L_n(\theta_{n0})}{\partial \theta \partial \theta'} \Upsilon_{\theta,n}^{-1} \right)$ and $\Omega = \lim_{n\to\infty} \left(-\frac{1}{n} \Upsilon_{\theta,n}^{-1} \Omega_n \Upsilon_{\theta,n}^{-1} \right)$. Hence,

$$\sqrt{n} \Upsilon_{\theta,n} (\hat{\theta}_{n2} - \theta_{n0}) \xrightarrow{p} N(0, \Sigma^{-1}(\Sigma + \Omega)\Sigma^{-1}). \blacksquare \tag{46}$$

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic, The Netherlands.

Anselin, L., 1992. Space and applied econometrics, Anselin (ed.), Special Issue. Regional Science and Urban Economics 22.

Anselin, L. and A.K. Bera, 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics, A. Ullah and D.E.A. Giles (eds.), Handbook of Applied Economics Statistics. Marcel Dekker, New York.

Anselin, L. and R. Florax, 1995. New Directions in Spatial Econometrics, Springer-Verlag, Berlin.

Anselin, L. and S. Rey, 1997. Spatial econometrics, Anselin, L. and S. Rey (eds.), Special Issue, International Regional Science Review 20.

Cliff, A.D. and J.K. Ord, 1973. Spatial Autocorrelation. London: Pion Ltd.

Chan, N.H. and C.Z. Wei, 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Annals of Statistics 15, 1050-1063.

Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.

Fingleton, B., 1999. Spurious spatial regression: some Monte Carlo results with a spatial unit root and spatial cointegration. Journal of Regional Science 39, 1-19.

Horn, R., and C. Johnson, 1985. Matrix Analysis. Cambride University Press, New York.

Kelejian, H.H. and I.R. Prucha, 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. Journal of Real Estate Finance and Economics 17:1, 99-121.

Kelejian H.H. and I.R. Prucha, 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics, 104, 219-257.

Kelejian, H.H. and D. Robinson, 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. Papers in Regional Science 72, 297-312.

Lee, L.F., 2001. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models II: mixed regressive, spatial autoregressive models. Manuscript, Department of Economics, The Ohio State University.

Lee, L.F., 2003. Best spatial two-stage least squares estimator for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22, No.4, 307-335.

Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. Econometrica 72, No.6, 1899-1925.

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

Lee, L.F. and J. Yu, 2007. A spatial dynamic panel data model with both time and individual fixed effects. Working Paper, Department of Economics, The Ohio State University.

Liu, X., L.F. Lee, and C.R. Bollinger, 2006. Improved efficient quasi maximum likelihood estimator of spatial autoregressive models. Manuscript, Department of Economics, The Ohio State University.

Moon, H.R. and P.C.B. Phillips, 2000. Estimation of autoregressive roots near unity using panel data. Econometric Theory 16, 927-997.

Moon, H.R. and P.C.B. Phillips, 2004. GMM estimation of autoregressive roots near unity with panel data. Econometrica 72, 467-522.

Ord, J.K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120-297.

Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregressions. Biometrika 74, 535-547.

Phillips, P.C.B., 1988. Regression theory for near-integrated time series. Econometrica 56, 1021-1043.

Phillips, P.C.B., H.R. Moon, Z Xiao, 2001. How to estimate autoregressive roots near unity. Econometric Theory 17, 29-69.

Rothenberg, T.J., 1971. Identification in parametric models. Econometrica 39, 577-591.

Ruud, P.A., 2000. An Introduction to Classical Econometric Theory. Oxford University Press, New York.

Wu, C.F., 1981. Asymptotic theory of nonlinear least squares estimation. The Annals of Statistics 9, 501-513.

Yu, J., R. de Jong and L-F. Lee, 2007. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large: a nonstationary case. Working Paper, The Ohio State University.