Near Unit Root in the Spatial Autoregressive Model*

Lung-fei Lee
Department of Economics
Ohio State University
Columbus, OH 43210, USA
Email: lflee@econ.ohio-state.edu

Jihai Yu
Department of Economics
University of Kentucky
Lexington, KY 40506, USA
Email: jihai.yu@uky.edu

December 2007; revised: November, 2008

*We would like to thank two anonymous referees and the coeditor, Professor Guido Kuersteiner, for their comments and suggestions for improving this paper. Lee acknowledges financial support from NSF under Grant No. SES-0519204.

Running Head: Near Unit Root in the Spatial Autoregressive Model

Mailing Address:

Lung-fei Lee, Department of Economics, Ohio State University, Columbus, OH 43210, USA. Email: lflee@econ.ohio-state.edu

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 into an unstable component and a stable one, and establish asymptotic properties of QMLE, 2SLSE and linearized QMLE of the parameters. The estimator for the spatial effect has a higher rate of convergence, and the estimators for 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 in time series, the estimators are all asymptotically normal. Similarly to the regular SAR model, QMLE and linearized QMLE are more efficient than 2SLSE.

Keywords:

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

JEL Classification: C13, C23, R15

1. INTRODUCTION

Spatial econometrics consists of econometric techniques dealing with the interactions of economic units in space, where the space can be of physical or economic nature. For a cross sectional model, the spatial autoregressive (SAR) model by Cliff and Ord (1973) has received the most attention in economics[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. By choosing some specific IVs, Lee (2003) derived the best two stage least square (B2SLS) estimators. Lee (2004) studied the asymptotic properties of the maximum likelihood (ML) and quasi-maximum likelihood (QML) methods for the SAR model. In 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 the GMM estimator (GMME), which can be more efficient than the 2SLS estimator (2SLSE). The best GMME can be as efficient as the ML estimator (MLE) when the true disturbances are normal. Liu et al. (2006) showed that carefully designed linear and quadratic moment functions can generate a GMME which is more efficient relative to the QML estimate (QMLE) 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 is controlled to a limited degree. With a spatial weights matrix being row-normalized, the spatial effect cannot be equal to or larger than one (Ord 1975). If the true spatial effect is near unity, the spatial dependence across units will be strong and the variances of the dependent variables will be 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 2SLSE and QMLE[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 the time series setting to the spatial setting.

In empirical applications, either for the spatial lag model or the spatial error model, there are cases in which the spatial effect coefficient is close to one. Pinkse et al. (2002) investigated the local competition in U.S. wholesale gasoline markets, and found that the competition is highly localized. Keller and Shiue (2007) studied the spatial interaction of the interregional trade on the rice price; they found that the spatial effect is very strong under the SAR model specification. Parys and Verbeke (2007) investigated the tax competition among Belgian municipalities; using the contiguity weights matrix, the spatial effect estimate is close to unity. Also, Tsutsumi et al. (2007) studied the spatial characteristics in the Tokyo apartment market; they found that the autoregressive coefficient of the spatial errors is very high for some weights matrix specification. These findings motivate our investigation into the possible near unit root model in the spatial setting[3].

The near unit root model has been extensively used in the time series literature to study the test statistics under local alternatives, starting with Chan and Wei (1987) and Phillips (1987b, 1988). Also, the estimation of the autoregressive roots near unity is studied both in the time series (Phillips et al. 2001) and in the panel data setting (Moon and Phillips 2000, 2004). For the near unit root model 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 weights matrix is usually row-normalized so that the influence of neighbors can be represented in terms of averages. As the SAR model is an equilibrium model, the unit root case where the spatial coefficient is equal to 1 is not allowed (otherwise, it will not be an equilibrium model). Hence, we study the near unit root case in SAR model through local alternatives, but not the unit root case itself[4]. Therefore, in the near unit root model for spatial data, the story may be different from the one in time series[5].

The paper is organized as follows. Section 2 introduces the model and decomposes the data generating process (DGP) into a stable part and an unstable part when the true spatial effect is near unity. Section 3 establishes the consistency and asymptotic distribution of the QMLE. Section 4 gives asymptotic properties of the 2SLSE, and of the linearized QMLE where the asymptotic efficiency of any initial estimates can be improved. Monte Carlo results are provided in Section 5 to compare the finite sample performance of various estimators. Conclusions are made in Section 6 and proofs for theorems are provided in Appendices.

2. THE MODEL

We 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[6] with eigenvalues d_{ni} such that either $d_{ni} = 1$ or $|d_{ni}| < 1$[7]. 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 where for the true spatial effect parameter,

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

where ψ_n goes to positive 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[8]. Note that because the SAR model is specified to be an equilibrium model, it is not relevant to assume that $\lambda_{n0} = 1$. With W_n being diagonalizable, let R_n be the eigenvector matrix, we have

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

where $D_n = diag\{\mathbf{1}_{m_n}, d_{n,m_n+1}, \cdots, d_{nn}\}$ 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 $|d_{ni}| \leq 1$ for all i, $(I_n - \lambda_{n0}D_n)$ is invertible and, so is $(I_n - \lambda_{n0}W_n)$. Hence, with our near unit root specification, for any finite n, this model is still a well-defined equilibrium model.

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 $\tilde{D}_n = diag\{0, \cdots, 0, d_{n,m_n+1}, \cdots, d_{nn}\}$. The J_n consists of all the m_n unit eigenvalues and $\tilde{D}_n$ consists of all the eigenvalues with an absolute value less than one. It follows that W_n can also be decomposed accordingly into two parts: $W_n = W_n^u + \tilde{W}_n$, where $W_n^u = R_nJ_nR_n^{-1}$ and $\tilde{W}_n = R_n\tilde{D}_nR_n^{-1}$. Denote $S_n(\lambda) = I_n - \lambda W_n$, from (1), the equilibrium vector Y_n is $Y_n = S_n^{-1}(\lambda_{n0})(X_n\beta_0 + \epsilon_n)$. As $S_n(\lambda) = I_n - \lambda W_n = R_n(I_n - \lambda D_n)R_n^{-1}$, $S_n^{-1}(\lambda) = R_n(I_n - \lambda D_n)^{-1}R_n^{-1}$. From (2), $I_n - \lambda_{n0}D_n = diag\{(1-\lambda_{n0}), \cdots, (1-\lambda_{n0}), (1-\lambda_{n0}d_{n,m_n+1}), \cdots, (1-\lambda_{n0}d_{nn})\}$, and hence, $(I_n - \lambda_{n0}D_n)^{-1} = \psi_n\lambda_{n0}J_n + (I_n - \lambda_{n0}\tilde{D}_n)^{-1}$. Denote $G_n = W_nS_n^{-1}(\lambda_{n0})$, we have

$$S_n^{-1}(\lambda_{n0}) = \psi_n\lambda_{n0}W_n^u + (I_n - \lambda_{n0}\tilde{W}_n)^{-1} \text{ and } G_n = \psi_n\lambda_{n0}W_n^u + W_n(I_n - \lambda_{n0}\tilde{W}_n)^{-1}, \tag{4}$$

because $W_nW_n^u = R_nD_nJ_nR_n^{-1} = W_n^u$. 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), we have,

$$Y_n = \psi_nY_n^u + \tilde{Y}_n \text{ where } Y_n^u = \lambda_{n0}W_n^u(X_n\beta_0 + \epsilon_n) \text{ and } \tilde{Y}_n = (I_n - \lambda_{n0}\tilde{W}_n)^{-1}(X_n\beta_0 + \epsilon_n). \tag{5}$$

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 one. The implied variance of Y_n can be large because the first component has ψ_n, which is explosive. Also, Y_n and W_nY_n have the same unstable component because $W_nY_n^u = Y_n^u$. Thus, Y_n and W_nY_n feature spatial cointegration in the cross sectional setting.

2.2 Assumptions of the Model

Our analysis of the asymptotic properties of estimators are based on the following assumptions.

Assumption 1. $\lambda_{n0} = 1 - \frac{1}{\psi_n}$ where ψ_n is an increasing positive sequence which tends to infinity as n goes to infinity.

Assumption 2. W_n is diagonalizable, row sum and column sum bounded[9](for short, UB). It has m_n unit roots, and the remaining eigenvalues d_{ni}, $i = m_n + 1, \cdots, n$, are uniformly bounded away from 1 in absolute value for all n and i.

Assumption 3. The disturbances $\{\epsilon_{ni}\}$, $i = 1, 2, ..., n$ are $i.i.d$ with zero mean, variance σ_0^2 and $\mathrm{E}\,|\epsilon_{ni}|^{4+\eta} < \infty$ 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. $(I_n - \lambda_{n0}\tilde{W}_n)^{-1}$ and W_n^u are UB.

Assumption 1 specifies how close the true spatial effect 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, and all of its eigenvalues are real, less than or equal to one in absolute value and its largest eigenvalue is always equal to 1 (see Ord (1975)). If the W_n is a block diagonal matrix with row normalization, m_n would be equal or greater than the number of blocks. The diagonalizability of W_n allows us to characterize W_n by its eigenvalues and eigenvector decomposition. In time series, the use of eigenvalues of relevant matrices to characterize stability of dynamics is popular (see, e.g., Fountis and Dickey 1989, Hamilton 1994). It is known that a square matrix can be approximated arbitrarily well by a diagonalizable matrix (Dhrymes 1978, p.46); hence, Assumption 2 might be practical. Assumption 3 and 4 are standard assumptions. The higher than the fourth moment condition in Assumption 3 is needed to apply the central limit theorem (CLT) 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 higher order moment conditions. Assumption 5 is to guarantee that the stable part and the unstable part, after being rescaled, of $S_n^{-1}(\lambda_{n0})$ in (4) are UB. In the standard case with the true λ_0 being strictly less than one, an important assumption due to Kelejian and Prucha (1998) is that $S_n^{-1}(\lambda_0)$ is UB. For the near unit root case, $S_n^{-1}(\lambda_{n0})$ will not be UB due to the explosive factor ψ_n in (4) for the unstable component. By taking out the ψ_n factor in $S_n^{-1}(\lambda_{n0})$, the elements of the remaining matrix and its row and column sum norms will not grow with n.

3. QMLE

3.1 Consistency

Denote $\theta = (\delta', \sigma^2)'$ where $\delta = (\beta', \lambda)'$. For (1), the log likelihood function is

$$\ln L_n(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\epsilon_n'(\delta)\epsilon_n(\delta), \tag{6}$$

where $\epsilon_n(\delta) = S_n(\lambda)Y_n - X_n\beta$. From (6), given λ, the QMLE of β and σ^2 are

$$\hat{\beta}_n(\lambda) = (X_n'X_n)^{-1}X_n'S_n(\lambda)Y_n, \tag{7}$$

$$\hat{\sigma}_n^2(\lambda) = \frac{1}{n}[S_n(\lambda)Y_n - X_n\hat{\beta}_n(\lambda)]'[S_n(\lambda)Y_n - X_n\hat{\beta}_n(\lambda)] = \frac{1}{n}Y_n'S_n'(\lambda)M_nS_n(\lambda)Y_n, \tag{8}$$

where $M_n = I_n - P_{X_n}$ with $P_{X_n} = X_n(X_n'X_n)^{-1}X_n'$. Hence, the log likelihood function can concentrate at λ as

$$\ln L_n(\lambda) = -\frac{n}{2}(\ln(2\pi)+1) - \frac{n}{2}\ln\hat{\sigma}_n^2(\lambda) + \ln|S_n(\lambda)|. \tag{9}$$

When W_n is row normalized, the parameter value of λ is known to be in the interval $(\frac{1}{d_{min,n}}, 1)$ to assure 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 due to the fact that 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 (9) 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 nonlinear estimation via an extremum estimator, it is conventional to assume a parameter space to be compact. With a compact parameter space and a continuous objective function, the extremum estimator is guaranteed to exist. For our model on hand, Λ will not be compact. However, from the proof of Theorem 1 for consistency of the QMLE, with large enough n, the estimator exists in a neighborhood of λ_{n0} in probability because the log likelihood values with λ far away from λ_{n0} are smaller than that at λ_{n0}.

The presence of X_n in (1) is a distinctive feature of the mixed regressive SAR model. From (1), the reduced form equation of Y_n is $Y_n = \lambda_{n0}(G_n X_n \beta_0) + X_n\beta_0 + S_n^{-1}\epsilon_n$. It is noted that while the regressors in X_n are bounded, the generated regressor $G_n X_n \beta_0$ is explosive as it has the explosive factor ψ_n from (4). 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) $\lim_{n\to\infty} \frac{1}{n} tr(W_n'^u W_n^u) > 0$; (3) for any finite constant c, $\lim_{n\to\infty} \frac{1}{n} tr[(I_n - cW_n^u)'(I_n - cW_n^u)] > 0$.

Assumption 7. $\beta_0' \lim_{n\to\infty}(\frac{1}{n} X_n' W_n'^u M_n W_n^u X_n)\beta_0 > 0$ holds.

Assumption 6 (1) and (2) specify that the unit roots in W_n play a significant role[10]. Assumption 6 (3) might not be satisfied if W_n^u 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[11] $\lim_{n\to\infty} \frac{m_n}{n} = 1$. 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 1-7, the QMLE $\hat{\lambda}_n$ is ψ_n-consistent, i.e., $\psi_n(\hat{\lambda}_n - \lambda_{n0}) = o_p(1)$. Also, $\hat{\beta}_n - \beta_0 = o_p(1)$ and $\hat{\sigma}_n^2 - \sigma_0^2 = o_p(1)$.*

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 but $\sqrt{n}\psi_n$-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 $W_n^u X_n \beta_0$ is multicollinear with X_n in the limit. If this happens, the identification uniqueness condition can rely on the following Assumption 8. This situation corresponds to the model identification via the disturbances of the reduced form 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 $\lim_{n\to\infty} \frac{m_n}{n} \leq \lim_{n\to\infty} \frac{1}{n} tr(W_n'^u W_n^u)$.

It is noted that when W_n is symmetric, $\frac{m_n}{n} = \frac{1}{n} tr(W_n'^u W_n^u)$.[12] In this case, identification condition in the second part of Assumption 8 will be satisfied regardless of whether $\frac{m_n}{n}$ is less than 1/2 or not. The symmetric W_n occurs in some social interaction models with group interactions.

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

Hence, from either Theorem 1 or Theorem 2, $\hat{\lambda}_n$ has a higher rate of convergence than other parameters. This is due to the unstable component in $W_n Y_n$, which has an explosive factor ψ_n in its mean equation $G_n X_n \beta_0$ 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. The mean values of $W_n Y_n$ and its variances for each spatial unit are large when n is large, but they have the same order. These are the differences between the near unit root in the SAR scenario and that in the time series scenario, although they do have some common features. The asymptotics for $\hat{\lambda}_n$ in Theorems 2 (and the following Theorem 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 (1987b).

3.2 Asymptotic Distribution

By using the formula for variance of the linear and quadratic form of the disturbances, the variance of the $\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_{n0})}{\partial \theta}$ is (see Appendix B) $\mathrm{E}\left(\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_{n0})}{\partial \theta} \cdot \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_{n0})}{\partial \theta'}\right) = \Sigma_n + \Omega_n$ where Σ_n and Ω_n are, respectively, in (22) and (23). The Ω_n vanishes if the disturbances are normally distributed. Using Lemma 2, the elements of Σ_n and Ω_n may have different orders because G_n has a higher order of ψ_n specified in (4). This implies that we need a rescaling matrix $\Upsilon_{\theta,n} = \begin{pmatrix} \Upsilon_{\delta,n} & \mathbf{0}_{(k+1)\times 1} \\ \mathbf{0}_{1\times(k+1)} & 1 \end{pmatrix}$ where $\Upsilon_{\delta,n} = \begin{pmatrix} I_k & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times k} & \psi_n \end{pmatrix}$, in order the limits of $\Upsilon_{\theta,n}^{-1} \Sigma_n \Upsilon_{\theta,n}^{-1}$ and $\Upsilon_{\theta,n}^{-1} \Omega_n \Upsilon_{\theta,n}^{-1}$ can be well defined. Let $\Sigma = \lim_{n\to\infty} \Upsilon_{\theta,n}^{-1} \Sigma_n \Upsilon_{\theta,n}^{-1}$ and $\Omega = \lim_{n\to\infty} \Upsilon_{\theta,n}^{-1} \Omega_n \Upsilon_{\theta,n}^{-1}$ be assumed to exist.

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. This assumption 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 Σ will be nonsingular. 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, $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0}) = -(\Upsilon_{\theta,n}^{-1} \frac{1}{n} \frac{\partial^2 \ln L_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'} \Upsilon_{\theta,n}^{-1})^{-1} \Upsilon_{\theta,n}^{-1} \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_{n0})}{\partial\theta}$ where $\tilde{\theta}_n$ lies between $\hat{\theta}_n$ and θ_{n0}. Hence, using the CLT for the linear and quadratic form of ϵ_n and the 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 1-6, and 7 (or, 8 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,* $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0}) \xrightarrow{d} N(0, \Sigma^{-1})$.

In the time series literature, the asymptotic distribution of the least square estimator of the AR(1) coefficient is normal, when it is strictly smaller than 1. It follows a nonstandard distribution (in terms of functionals of the standard Brownian motion) when it is equal to or near 1 (Chan and Wei (1987) and Phillips (1987a, 1987b, 1988)). In this paper, the asymptotic distribution of $\hat{\lambda}_n$ is still normal, which is different from the time series counterpart.

Another difference between the SAR and time series models is the testing for the near unit root estimator. In the time series literature, the testing is focused on the hypothesis testing of unit root *vs.* stationarity process, and the discriminatory power of statistical tests for unit root is rather low under local alternatives. In the SAR model, as the unit root model is not allowed, it is not apparent how the null hypothesis on the stability of SAR process can be formulated. As an alternative, it is tractable to construct confidence intervals or sets with the asymptotic distribution of $\hat{\theta}_n - \theta_{n0}$ in Theorem 3. Denote $\hat{\sigma}_{jj}^2$ the (j,j) element of an estimated $\hat{\Sigma}_n^{-1} + \hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1}$ (see (24) and (25) in Appendix B for variance estimators $\hat{\Sigma}_n$ and $\hat{\Omega}_n$). The interval estimate for the jth element of θ_{n0} is $(\hat{\theta}_{n,j} - z_{\alpha/2}\frac{\hat{\sigma}_{jj}}{\sqrt{n}}, \hat{\theta}_{n,j} + z_{\alpha/2}\frac{\hat{\sigma}_{jj}}{\sqrt{n}})$, where $z_{\alpha/2}$ is the critical value of the standard normal distribution associated with the significance level α. As the elements of $\hat{\Sigma}_n$ and $\hat{\Omega}_n$ have different orders (see (26) in Appendix B), $\hat{\sigma}_{jj}$ will be $O_p(1)$ for β and σ^2, but $O_p(\psi_n^{-1})$ for[13] λ. Hence, the confidence interval for λ_{n0} will be much narrower. Also, the confidence set of the whole parameter vector can be specified. From Theorem 3, $n(\hat{\theta}_n - \theta_{n0})(\Sigma_n^{-1}(\Sigma_n + \Omega_n)\Sigma_n^{-1})^{-1}(\hat{\theta}_n - \theta_{n0})$ will be asymptotically χ_{k+2}^2 distributed. Therefore, the confidence set Θ_c is the set of θ such that $n(\hat{\theta}_n - \theta)(\hat{\Sigma}_n^{-1} + \hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1})^{-1}(\hat{\theta}_n - \theta) < x_\alpha$, where x_α is the critical value of the chi-square distribution associated with the significance level α.

4. TWO STAGE LEAST SQUARES ESTIMATORS

To estimate the SAR model (1), the method of IV is feasible. Denote $\delta_n = (\beta', \lambda_n)'$ and $Z_n = (X_n, W_nY_n)$, (1) can be rewritten as $Y_n = Z_n\delta_{n0} + \epsilon_n$. Denote an $n \times p$ matrix H_n, where $p \geq k+1$, as the possible instruments for (X_n, W_nY_n). Kelejian and Prucha (1998) use H_n generated from W_n and X_n, for example, $H_n = (X_n, W_nX_n, W_n^2X_n)$. The ideal instruments would be $\mathrm{E}(Z_n|X_n) = (X_n, W_n(I_n - \lambda_{n0}W_n)^{-1}X_n\beta_0)$. 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 estimates.

4.1 2SLS

With H_n as an IV matrix, denote $P_{H_n} = H_n(H_n'H_n)^{-1}H_n'$, a 2SLS estimate can be

$$\hat{\delta}_{n,2sls} = (Z_n'P_{H_n}Z_n)^{-1}Z_n'P_{H_n}Y_n. \tag{10}$$

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

Since the number of variables in Z_n is equal to $k+1$, the number of IVs must be at least $k+1$, as 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. From (10),

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

By using Lemma 2, we can get the asymptotic properties of 2SLSEs of λ_{n0} and β_0. With $\hat{\delta}_{n,2sls}$ available, σ_0^2 can be estimated by $\hat{\sigma}_{n,2sls}^2 = \frac{1}{n}\sum_{i=1}^n (Y_n - Z_n\hat{\delta}_{n,2sls})'(Y_n - Z_n\hat{\delta}_{n,2sls})$.

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

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

4.2 Best 2SLS

The above 2SLSE might not be asymptotically optimal. The optimal IV is $H_n^* = \mathrm{E}(X_n, W_n Y_n | X_n) = (X_n, G_n X_n \beta_0)$.

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

Assumption 11 is similar to Assumption 10 (2). The nonsingular matrix properties in Assumptions 4 and 7 imply Assumption 11 from the partition matrix inversion formulae. As $G_n = W_n S_n^{-1}(\lambda_{n0})$ in H_n^* has the higher order of ψ_n, we need to standardize H_n^* by $\Upsilon_{\delta,n}^{-1}$. To construct a feasible optimum IV, we need initial $\Upsilon_{\delta,n}$-consistent estimators $\tilde{\delta}_n$ so as to approximate H_n^* by $\tilde{H}_n^* = (X_n, G_n(\tilde{\lambda}_n) X_n \tilde{\beta}_n)$. A natural candidate would be the 2SLS from (10). In general, any $\Upsilon_{\delta,n}$-consistent initial estimate will do.

Assumption 12. The initial estimator $\tilde{\delta}_n$ is $\Upsilon_{\delta,n}$-consistent, i.e.,$\Upsilon_{\delta,n}(\tilde{\delta}_n - \delta_{n0}) = o_p(1)$.

Let

$$\hat{\delta}_{n,b2sls} = (\tilde{H}_n^{*\prime} Z_n)^{-1} \tilde{H}_n^{*\prime} Y_n. \tag{11}$$

Theorem 5 *Under Assumptions 1-5, 11 and 12, $\hat{\delta}_{n,b2sls}$ is consistent and $\Upsilon_{\delta,n}\sqrt{n}(\hat{\delta}_{n,b2sls} - \delta_{n0}) \xrightarrow{d} N(0, \Phi_{b2sls})$ where $\Phi_{b2sls} = \sigma_0^2(\lim_{n\to\infty} \frac{1}{n}\Upsilon_{\delta,n}^{-1} H_n^{*\prime} H_n^* \Upsilon_{\delta,n}^{-1})^{-1}$.*

By comparing Φ_{b2sls} in Theorem 5 with Φ_{2sls} in Theorem 4, 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 see that $\Phi_{b2sls} - \Phi_{2sls}$ is negative semi-definite using the Cauchy-Schwarz inequality. Hence, the $\hat{\delta}_{n,b2sls}$ is indeed the best IV estimator.

4.3 Linearized QMLE

A computationally convenient way to increase the efficiency of a $\Upsilon_{\delta,n}$-consistent estimate is to use the Newton-Raphson method to obtain a linearized QMLE. Assume that we have an initial estimator $\hat{\theta}_{n1}$ available such that $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_{n1}-\theta_{n0}) = O_p(1)$. The linearized QMLE is

$$\hat{\theta}_{n2} = \hat{\theta}_{n1} - \left(\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\right)^{-1} \frac{\partial \ln L_n(\hat{\theta}_{n1})}{\partial\theta}, \tag{12}$$

where $\ln L_n(\theta)$ is the log likelihood function in (6).

Theorem 6 *Under Assumptions 1-5, 9 and* $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_{n1}-\theta_{n0}) = O_p(1)$, $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_{n2}-\theta_{n0}) \xrightarrow{d} N(0, \Sigma^{-1} + \Sigma^{-1}\Omega\Sigma^{-1})$. *When* ϵ_{ni}*'s are normally distributed,* $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_{n2}-\theta_{n0}) \xrightarrow{d} N(0, \Sigma^{-1})$.

From Theorems 3 and 6, the linearized QMLE has the same limiting distribution as the QMLE. For a regular log 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. Although there is evidence that its finite sample properties might not be better than the iterative one, it may improve the asymptotic efficiency of the 2SLSE when the disturbances are normally distributed. For an initial 2SLSE $\hat{\delta}_{n,2sls}$, its variance is $\Phi_{2sls} = \sigma_0^2(\lim_{n\to\infty} \frac{1}{n}\Upsilon_{\delta,n}^{-1} Z_n' P_{H_n} Z_n \Upsilon_{\delta,n}^{-1})^{-1}$. Denote $A^s = A + A'$ for any square matrix A. For the linearized MLE when the disturbance is normally distributed, its limiting variance of the estimates of δ_{n0} is $\sigma_0^2(\lim_{n\to\infty} \frac{1}{n}\Upsilon_{\delta,n}^{-1}(\Delta_n + Z_n'Z_n)\Upsilon_{\delta,n}^{-1})^{-1}$ where $\Delta_n = \begin{pmatrix} \mathbf{0}_{k\times k} & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times k} & tr((G_n - \frac{trG_n}{n}I_n)^s G_n) \end{pmatrix}$. 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[14]

$$Y_n = \lambda_{n0}W_nY_n + X_{n1}\beta_{01} + X_{n2}\beta_{02} + X_{n3}\beta_{03} + \epsilon_n, \tag{13}$$

where x_{i1}, x_{i2} and x_{i3} are three independently generated standard normal (scalar) variables[15] and are *i.i.d.* for all i. Also, ϵ_{ni}'s are drawn independently from the standard normal distribution. We choose $n = 49$ and $n = 245$. When $n = 49$, the spatial weights matrix we choose is the neighborhood matrix across 49 districts in Columbus, Ohio in Anselin (1988) for a 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 estimates; (3) ML–the maximum likelihood method; and (4) LML–the linearized maximum likelihood method using (12). In the simulation, we use the 2SLSEs as the initial estimates for B2SLS and LML.

The number of repetitions is 1000 for each case. We report the mean (Mean), empirical standard deviation (SD) and empirical root mean square error (RMSE) of the estimates. λ_{n0} takes the value $0.6, 0.9, 0.99, 0.999$

and $\beta_{01}, \beta_{02}, \beta_{03}$ are $-1, 0, 1$ respectively. We also report the CI, the percentage that each true parameter is covered by the estimated 95% confidence interval, and the CS, the percentage that the true parameter vector is covered by the 95% confidence set. Both the CI and CS are based on the relevant asymptotic distributions of the estimates. Table 1 is the result when $n = 49$ and Table 2 is the result when $n = 245$. To check whether the estimates 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.[16]

Tables 1-4 here

We see that all the estimates have small biases under different values of λ_{n0}. The CS and CI values are reasonably close to the specified 95% for many cases but, for some cases, the coverage percentages are lower. When λ_{n0} is closer to 1, the SD of $\hat{\lambda}_n$ decreases and that of β does not change in the order of magnitude, which is consistent with the theoretical prediction. In terms of finite sample performance, the MLEs are better than the 2SLSEs in terms of smaller variances. The B2SLSEs turn out worse than the simpler 2SLSEs in terms of variances for these two sample sizes[17]. From the quantile-quantile plot, we 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_0 can be better approximated by the normal distribution than that of the estimates of λ_{n0}.

To check the robustness of the estimates under non-normality of ϵ_n, we also run the simulation when ϵ_n is generated from independent exponential distributions with unit variance (demeaned by the population mean). The disturbances are skewed. To reduce unnecessary tables, we only present the case for $n = 245$ in Table 5. From Table 5, we can see that the CS and CI values are mostly similar to those of Table 2, but can be slightly smaller as well, while the estimates have similar asymptotic properties (as those with ϵ_n being normal).

Yu et al. (2007, 2008) investigate, respectively, the asymptotics of the QMLE for spatial dynamic panel data (SDPD) with spatial cointegration, and with stationarity. For the model $Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + V_{nt}$ where Y_{nt} is $n \times 1$ column vector and $\mathbf{c}_{n0}$ is the fixed effects, it is of interest to investigate the dynamic behavior via the matrix $A_n = (I_n - \lambda_0 W_n)^{-1}(\gamma_0 I_n + \rho_0 W_n)$. When $\gamma_0 + \lambda_0 + \rho_0 < 1$, the SDPD can be stationary; but when $\gamma_0 + \lambda_0 + \rho_0 = 1$ with $\gamma_0 \neq 1$, some eigenvalues of A_n are equal to the unit and the SDPD process is termed as spatial cointegration. It is interesting to investigate the property of the estimates of the cross SAR model, when the true DGP is an SDPD model with spatial cointegration or with $\gamma_0 + \lambda_0 + \rho_0$ less than one. In Table 6, we have different DGPs of the SDPD model and estimate the cross sectional SAR model using the last period cross sectional observations. We notice that when the DGP is an SDPD process with spatial cointegration or $\gamma_0 + \lambda_0 + \rho_0$ is close to one, the estimate of λ_0 will be close to 1. The estimate of λ_0 in the cross section data may capture dynamic effects via the spatial time lag[18]. This simulation result might explain the empirical findings in Keller and Shiue (2007), where both the estimates from cross sectional and panel data are presented. In Keller and Shiue (2007), the cross section estimate of λ_0 is close to 1, while the panel estimates of γ_0, λ_0 and ρ_0 has a sum close to 1.

Tables 5-6 here

6. CONCLUSION

In this paper, asymptotic properties of 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 while other parameters have the usual $\sqrt{n}$ rate. Similarly to the regular SAR model, QMLE and linearized QMLE are more efficient than 2SLSE.

The spatial weights matrix W_n in the SAR model considered in this paper has some unit eigenvalues while the remaining eigenvalues are less than one in absolute value. This includes the popular case where the spatial weights matrix is row-normalized. The near unit root for this SAR model refers to the case where the spatial effect λ_{n0} of the spatial lag W_nY_n is close to unity. Similarly to the time series literature (Phillips 1987b), this concept is modeled with a sequence of λ_{n0}, which converges to the unity as the sample size increases. This situation generates an irregularity in that the inverse of the spatial transformation matrix $(I_n - \lambda_{n0}W_n)$ will not satisfy the UB property. The UB property originated in Kelejian and Prucha (1998, 2001) is one of the essential regularity conditions used in order to establish the asymptotic properties of various estimators in the literature, such as the 2SLSE in Kelejian and Prucha (1998) and QMLE in Lee (2004). This paper investigates the possible asymptotic implications on those estimators when the UB property on $(I_n - \lambda_{n0}W_n)^{-1}$ is violated.

With a near unit root λ_{n0}, Y_n can be decomposed into two components – an unstable one which contains explosive elements for large n, and a stable one. Accordingly, the spatial lag W_nY_n will also contain an unstable component, which turns out to be exactly the same unstable component of Y_n. This feature presents spatial cointegration in the cross sectional setting. The consequence is that the 2SLSE and QMLE of λ_{n0} have a higher rate of convergence, while the estimates of the other remaining regression coefficients and the variance parameter have the usual rate of convergence. The fast rate of convergence of $\hat{\lambda}_n$ reflects the rate of convergence of λ_{n0} to the unity.

Contrary to the near unit root model of an autoregressive process in the time series literature, the asymptotic distributions of the estimates remain normal. This difference in the asymptotic distribution may be due to differences in structures of the SAR model from those of the autoregressive time series process. With a near unit λ_{n0}, the variance of the dependent variable conditional on exogenous variables can become very large over time, which is a feature of a time series with a unit root or near unit root. However, the 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. Also, there is no 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 $W_n Y_n$ cannot be the unity; otherwise, the model is not complete. The simplest SAR process, which resembles the autoregressive process of a time series, is the case 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 (1987b).

We also compare the performance of the 2SLSE with the MLE, 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 MLEs are better than the 2SLSEs in terms of smaller variances.

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), spurious regression and cointegration in the spatial context may have implications in empirical analysis. Those issues have not been addressed in this paper but they may be worthy of investigation in future research.

Notes

[1]Early development in estimation and testing for spatial dependence in cross sectional 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.

[3]The estimated spatial effect coefficients are 0.921 in Pinkse et al. (2002), 0.945 for Keller and Shiue (2007), 0.987 in Parys and Verbeke (2007) and 0.9577 in Tsutsumi et al. (2007).

[4]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, the system is an equilibrium one and a unit coefficient of the spatial lag is possible. This setting is analogous to having an initial observation in the time series autoregressive model. However, this is a rather limiting setting for spatial models.

[5]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.

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

[7]For a weights matrix that is row normalized from a symmetric matrix, it has real eigenvalues, with all its eigenvalues less than or equal to one in absolute value and its largest eigenvalue always equal to 1 (see Ord (1975)).

[8]In a time series near unit root model, the deviation from the unit root is measured through a noncentrality parameter, where the AR(1) coefficient is usually specified as $\exp(c/n)$ or $1 - \frac{c}{n}$ with c being the noncentrality parameter. For the near unit root in the SAR model, ψ_n can take a general form as long as it is increasing in n, which does not need to be specified in empirical applications.

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

[10]When $\frac{m_n}{n} \to 0$, the unstable component might be dominated by the stable component unless ψ_n is large enough. From (4), the contribution of the unstable component will depend on $\psi_n \cdot \frac{m_n}{n}$. Here, a larger value of ψ_n increases the contribution of the unstable component and a 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 size n.

[11]This becomes clearer for the special case when W_n is symmetric. In that case, R_n is an orthonormal matrix and we have $\frac{1}{n} tr[(I_n - cW_n^u)'(I_n - cW_n^u)] = \frac{m_n}{n}(1-c)^2 + (1 - \frac{m_n}{n})$, 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 $c = 1$.

[12]This occurs because, when W_n is symmetric, R_n is an orthonormal matrix and $\frac{1}{n} tr(W_n'^u W_n^u) = \frac{m_n}{n}$.

[13] We note that $\Upsilon_{\theta,n}$ appears neither in the confidence interval nor in the confidence set because $\Upsilon_{\theta,n}$ as the normalization factor for $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0})$ has been cancelled with that in $\Upsilon_{\theta,n}\hat{\Sigma}_n^{-1}\Upsilon_{\theta,n}$, which is a consistent estimate of the limiting variance Σ^{-1} in Theorem 3.

[14] For the simulations, we have also investigated the finite sample performance of the QML approach for the pure SAR model without exogenous variables: $Y_n = \lambda_{n0} W_n Y_n + \epsilon_n$. The results are similar. To save space, those tables are not presented here.

[15] One possible problem associated with the normal assumption is the uniform boundedness of X_n (with normal distribution specification, X_n is not bounded). Hence, we also use truncated standard normal distribution with the range $[-1.5, 1.5]$ and $[-3, 3]$. It turns out that the estimates where X_n is truncated have similar properties to the case where X_n is not truncated. The results are not reported in the paper. This similarity is not surprising both numerically and theoretically. As we have indicated, the boundedness assumption for X_n may be replaced by some moment conditions on X_n.

[16] To save space, we do not present the plots for β_{02} and β_{03}. The conclusions are similar to that of β_{01}.

[17] This might be due to the sensitivity in the computation of the inverse of $S_n(\tilde{\lambda}_n)$ when the initial value $\tilde{\lambda}_n$ is close to the unity. The large SD values of the B2SLS are caused by several outliers. Without those outliers, the B2SLSE has the same magnitude of SDs as 2SLSE.

[18] If $Y_{nt} = Y_{n,t-1}$ (more accurately, from Yu et al. (2007), $Y_{nt} = Y_{n,t-1}$+linear functions of (X_{nt}, V_{nt} and all their time lags), where the additional linear sum of $V_{n,t-s}$, $s = 0, 1, ..$, are stable), then the SDPD model would imply that $Y_{nt} = \frac{\lambda_0+\rho_0}{1-\gamma_0} W_n Y_{nt} + X_{nt}\frac{\beta_0}{1-\gamma_0} + \frac{1}{1-\gamma_0}\mathbf{c}_{n0} + \frac{1}{1-\gamma_0} V_{nt}$. The composite coefficient $\frac{\lambda_0+\rho_0}{1-\gamma_0}$ is close (equal) to 1 if $\gamma_0 + \lambda_0 + \rho_0$ is close (equal) to 1.

[19] When $\frac{1}{2} < q_1 \leq 1$, $L(s)$ is also 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 $s = 0$ depends on specific values of q_1 and q_2 when $q_2 < q_1$.

References

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

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

Anselin, L. & A.K. Bera (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah and D.E.A. Giles (ed.) *Handbook of Applied Economics Statistics*. New York: Marcel Dekker, pp. 237-289.

Anselin, L. & R. Florax (1995) *New Directions in Spatial Econometrics*. Berlin: Springer-Verlag.

Anselin, L. & S. Rey (1997) Spatial econometrics. In Anselin, L. and S. Rey (ed.), Special Issue *International Regional Science Review* 20.

Cliff, A.D. & J.K. Ord (1973) *Spatial Autocorrelation*. London: Pion Ltd.

Chan, N.H. & 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*. New York: Wiley.

Dhrymes, P.J. (1976) *Mathematics for Econometrics*. New York: Springer.

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.

Fountis, N.G. & D.A. Dickey (1989) Testing for a unit root nonstationarity in multivariate autoregressive time series. *The Annals of Statistics* 17, 419-428.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.

Horn, R. & C. Johnson (1985) *Matrix Analysis*. New York: Cambride University Press.

Kelejian, H.H. & 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, 99-121.

Kelejian H.H. & 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. & 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.

Keller, W. & C.H. Shiue (2007) The origin of spatial interaction. *Journal of Econometrics* 140, 304-332.

Lee, L.F. (2003) Best spatial two-stage least squares estimator for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22, 307-335.

Lee, L.F. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica* 72, 1899-1925.

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

Liu, X., L.F. Lee & C.R. Bollinger (2006) Improved efficient quasi maximum likelihood estimator of spatial autoregressive models. Manuscript, Ohio State University.

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

Moon, H.R. & 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.

Parys, S.V. & T. Verbeke. (2007) Tax competition among Belgian municipalities: a multi-dimensional battle. Paper presented at the World Conference of Spatial Econometrics Association, 2007.

Phillips, P.C.B. (1987a) Time series regression with a unit root. *Econometrica* 55, 277-301.

Phillips, P.C.B. (1987b) 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.

Pinkse, J., M. E. Slade & C. Brett. (2002) Spatial price competition: a semiparametric approach. *Econometrica* 70, 1111-1153.

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

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

Tsutsumi, M., Y. Yoshida, H. Seya & Y. Kawaguchi. (2007) Spatial analysis of Tokyo apartment market. Paper presented at the World Conference of Spatial Econometrics Association, 2007.

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

Yu, J., R. de Jong & 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. Manuscript, Ohio State University.

Yu, J., R. de Jong & L.F. Lee (2008) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146, 118-134.

Appendix A: Some Lemmas and Proofs

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

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

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

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

(4) $\frac{1}{\psi_n^2\sqrt{n}}X_n'G_n'G_n\epsilon_n = \frac{1}{\sqrt{n}}X_n'W_n'^uW_n^u\epsilon_n + O_p(\frac{1}{\psi_n})$;
(5) $\frac{1}{\psi_n\sqrt{n}}(\epsilon_n'G_n\epsilon_n - \sigma_0^2tr(G_n)) = \frac{1}{\sqrt{n}}[\epsilon_n'W_n^u\epsilon_n - \sigma_0^2tr(J_n)] + O_p(\frac{1}{\psi_n})$;
(6) $\frac{1}{\psi_n^2\sqrt{n}}(\epsilon_n'G_n'G_n\epsilon_n - \sigma_0^2tr(G_n'G_n)) = \frac{1}{\sqrt{n}}[\epsilon_n'W_n'^uW_n^u\epsilon_n - \sigma_0^2tr(W_n'^uW_n^u)] + O_p(\frac{1}{\psi_n})$;
(7) $\lim_{n\to\infty}\frac{1}{n\psi_n^2}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) = \beta_0'\lim_{n\to\infty}(\frac{1}{n}X_n'W_n'^uM_nW_n^uX_n)\beta_0$;
(8) $\frac{1}{n\psi_n}tr(G_n) = \frac{tr(J_n)}{n} + O(\frac{1}{\psi_n})$; *and*
(9) $\frac{1}{n\psi_n^2}tr(G_n'G_n) = \frac{1}{n}tr(W_n'^uW_n^u) + O(\frac{1}{\psi_n})$.

Lemma 2 *Under Assumptions 1-5, for any $n \times n$ nonstochastic UB matrix B_n,*

$$\frac{1}{n\cdot\psi_n^2}Y_n'B_nY_n - \mathrm{E}\frac{1}{n\cdot\psi_n^2}Y_n'B_nY_n = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{14}$$

$$\frac{1}{n\cdot\psi_n}Y_n'B_n\epsilon_n - \mathrm{E}\frac{1}{n\cdot\psi_n}Y_n'B_n\epsilon_n = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{15}$$

and

$$\frac{1}{n\cdot\psi_n}X_n'B_nY_n - \mathrm{E}\frac{1}{n\cdot\psi_n}X_n'B_nY_n = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{16}$$

where $\mathrm{E}\frac{1}{n\cdot\psi^2(n)}Y_n'B_nY_n = \frac{1}{n}(W_n^uX_n\beta_0)'B_n(W_n^uX_n\beta_0) + \frac{\sigma_0^2}{n}tr(W_n'^uB_nW_n^u) + o(1)$, $\mathrm{E}\frac{1}{n\cdot\psi_n}Y_n'B_n\epsilon_n$ *is* $O(1)$ *and* $\mathrm{E}\frac{1}{n\cdot\psi_n}X_n'B_nY_n = \frac{1}{n}X_n'B_nW_n^uX_n\beta_0 + o(1)$.

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

$$\Lambda_n^{(\mathfrak{M}+)} = \{\lambda|\lambda \in \Lambda,\ \psi_n(\lambda_{n0} - \lambda) > \mathfrak{M}\}, \tag{17}$$
$$\Lambda_n^{(\mathfrak{M})} = \{\lambda|\lambda \in \Lambda,\ |\psi_n(\lambda_{n0} - \lambda)| \leq \mathfrak{M}\}. \tag{18}$$

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). Let

$$\sigma_n^2(\lambda) = \frac{\sigma_0^2}{n}tr[(S_n(\lambda)S_n^{-1}(\lambda_{n0}))'S_n(\lambda)S_n^{-1}(\lambda_{n0})], \tag{19}$$

and

$$\sigma_n^{*2}(\lambda) = \frac{1}{n}(\lambda - \lambda_{n0})^2(G_nX_n\beta_0)'M_nG_nX_n\beta_0 + \sigma_n^2(\lambda). \tag{20}$$

Lemma 3 *Under Assumptions 1-5, for $\hat{\sigma}_n^2(\lambda)$ in (8) and $\sigma_n^{*2}(\lambda)$ in (20),*
(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 *Under Assumptions 1-5, for large enough n,*
(i) $\sigma_n^2(\lambda)$ *is bounded away from zero and 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 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,*

$$\frac{1}{\psi_n^2}\sigma_n^2(\lambda) = \sigma_0^2[\frac{1}{\psi_n^2} + 2(\lambda_{n0}-\lambda)\frac{tr(G_n)}{n\psi_n^2} + (\lambda_{n0}-\lambda)^2\frac{tr(G_n'G_n)}{n\psi_n^2}] \xrightarrow{p} \sigma_0^2(1-\lambda)^2 \lim_{n\to\infty}\frac{1}{n}tr(W_n'^u W_n^u),$$

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 UB. 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) of order $O(\frac{1}{\psi_n})$. The items (1)-(7) are those consequences. For (8) and (9), as $G_n = \psi_n[\lambda_{n0}W_n^u + \frac{1}{\psi_n}W_n(I_n-\lambda_{n0}\tilde{W}_n)^{-1}]$, we have $\frac{1}{n\psi_n}tr(G_n) = \lambda_{n0}\frac{tr(J_n)}{n} + \frac{1}{n\psi_n}tr(W_n(I_n-\lambda_{n0}\tilde{W}_n)^{-1}) = \frac{m_n}{n} + O(\frac{1}{\psi_n})$, because the eigenvalues in $\tilde{D}_n$ is bounded away from 1 in absolute value. Also, $\frac{1}{n\psi_n^2}tr(G_n'G_n) = \frac{\lambda_{n0}^2}{n}tr(W_n'^u W_n^u) + O(\frac{1}{\psi_n})$. ■

Proof of Lemma 2: For (16), we have $\frac{1}{n\cdot\psi_n}X_n'B_nY_n = \frac{1}{n}X_n'B_n(Y_n^u + \frac{1}{\psi_n}\tilde{Y}_n)$ from (5). By Assumption 5, W_n^u and $(I_n-\lambda_{n0}\tilde{W}_n)^{-1}$ are UB. Hence, $\frac{1}{n}X_n'B_nW_n^u\epsilon_n = O_p(\frac{1}{\sqrt{n}})$ and $\frac{1}{n}X_n'B_n(I_n-\lambda_{n0}\tilde{W}_n)^{-1}\epsilon_n = O_p(\frac{1}{\sqrt{n}})$. Also, we have $\mathrm{E}\frac{1}{n\cdot\psi_n}X_n'B_nY_n = \frac{1}{n}X_n'B_n[\lambda_{n0}W_n^u + \frac{1}{\psi_n}(I_n-\lambda_{n0}\tilde{W}_n)^{-1}]X_n\beta_0$. From Assumption 5, $\lim_{n\to\infty}\frac{1}{n}X_n'B_n[\frac{1}{\psi_n}(I_n-\lambda_{n0}\tilde{W}_n)^{-1}]X_n\beta_0 = 0$ as $\psi_n\to\infty$. Hence, $\mathrm{E}\frac{1}{n\cdot\psi_n}X_n'B_nY_n = \frac{1}{n}X_n'B_nW_n^uX_n\beta_0 + o(1)$. Similarly, we can establish (14) and (15). ■

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

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

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

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

and

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

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

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

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

and $T_{2n,2}(\lambda) = \frac{1}{n}\epsilon_n' P_{X_n}\epsilon_n + \psi_n(\lambda_{n0}-\lambda)\frac{2}{n\psi_n}\epsilon_n' G_n' P_{X_n}\epsilon_n + [\psi_n(\lambda_{n0}-\lambda)]^2\frac{1}{n\psi_n^2}\epsilon_n' G_n' P_{X_n} G_n\epsilon_n = O_p(\frac{1}{n})$, uniformly in λ when $|\psi_n(\lambda_{n0}-\lambda)| \leq \mathfrak{M}$. 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,

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

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

and $\frac{T_{2n,2}(\lambda)}{[\psi_n(\lambda_{n0}-\lambda)]^2} = \frac{1}{[\psi_n(\lambda_{n0}-\lambda)]^2}\frac{1}{n}\epsilon_n' P_{X_n}\epsilon_n + \frac{1}{\psi_n(\lambda_{n0}-\lambda)}\frac{2}{n\psi_n}\epsilon_n' G_n' P_{X_n}\epsilon_n + \frac{1}{n\psi_n^2}\epsilon_n' G_n' P_{X_n} G_n\epsilon_n = O_p(\frac{1}{n})$, 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 $\frac{1}{n\psi_n^2}tr(G_n'G_n) = \frac{1}{n}tr(W_n'^u W_n^u) + O(\frac{1}{\psi_n})$ are strictly positive. 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 $\sigma_n^2(\lambda) = \sigma_0^2[1+2(\lambda_{n0}-\lambda)\frac{trG_n}{n} + (\lambda_{n0}-\lambda)^2\frac{tr(G_n'G_n)}{n}] = \frac{\sigma_0^2}{n}tr[(I_n-(\lambda_{n0}-\lambda)\psi_n\cdot\frac{G_n}{\psi_n})(I_n-(\lambda_{n0}-\lambda)\psi_n\cdot\frac{G_n'}{\psi_n})]$, we have $\lim_{n\to\infty}\sigma_n^2(\lambda_n) = \sigma_0^2\lim_{n\to\infty}\frac{1}{n}tr[(I_n-(c^*-1)W_n^u)(I_n-(c^*-1)W_n^u)']$, 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 $\sigma_n^2(\lambda) = \sigma_0^2[1+2(\lambda_{n0}-\lambda)\frac{tr(G_n)}{n} + (\lambda_{n0}-\lambda)^2\frac{tr(G_n'G_n)}{n}]$, $\frac{1}{n\psi_n}tr(G_n) = \frac{m_n}{n} + O(\frac{1}{\psi_n})$ and $\frac{1}{n\psi_n^2}tr(G_n'G_n) = \frac{1}{n}tr(W_n'^uW_n^u) + O(\frac{1}{\psi_n})$, we have

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

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, $|S_n(\lambda_{n0})| = (1-\lambda_{n0})^{m_n}\prod_{i=m_n+1}^n(1-\lambda_{n0}d_{ni})$ where d_{ni}'s are bounded away from 1. Thus,

$$\ln|S_n(\lambda_{n0})| = -m_n\ln\psi_n + \textstyle\sum_{i=m_n+1}^n\ln(1-\lambda_{n0}d_{ni}). \tag{21}$$

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 $O(1)$ on Λ_κ. ■

Appendix B: Variance of the Score and Its Estimator

We have $\frac{\partial\ln L_n(\theta_0)}{\partial\beta} = \frac{1}{\sigma_0^2}X_n'\epsilon_n$, $\frac{\partial\ln L_n(\theta_0)}{\partial\lambda} = \frac{1}{\sigma_0^2}(W_nY_n)'\epsilon_n - tr(G_n)$ and $\frac{\partial\ln L_n(\theta_0)}{\partial\sigma^2} = \frac{1}{2\sigma_0^4}(\epsilon_n'\epsilon_n - n\sigma_0^2)$. Denote $G_n^s = G_n + G_n'$. For the variance part, by using the formula for variance of the linear and quadratic form of the disturbances, $\mathrm{E}\frac{1}{n}X_n'\epsilon_n\epsilon_n'X_n = \sigma_0^2\frac{1}{n}X_n'X_n$, $\mathrm{E}\frac{1}{n}(\epsilon_n'\epsilon_n - n\sigma_0^2)^2 = \mu_4 - \sigma_0^4$; using $W_nY_n = G_nX_n\beta_0 + G_n\epsilon_n$, we have $\mathrm{E}\frac{1}{n}[(W_nY_n)'\epsilon_n - \sigma_0^2tr(G_n)]^2 = \frac{\sigma_0^2}{n}(G_nX_n\beta_0)'G_nX_n\beta_0 + \frac{2\mu_3}{n}\sum_{i=1}^nG_{n,ii}(G_nX_n\beta_0)_i + \frac{\sigma_0^4}{n}tr(G_n^sG_n) + \frac{\mu_4-3\sigma_0^4}{n}\sum_{i=1}^nG_{n,ii}^2$. For the covariance part, $\mathrm{E}\frac{1}{n}X_n'\epsilon_n\epsilon_n'W_nY_n = \sigma_0^2X_n'G_nX_n\beta_0 + \frac{\mu_3}{n}\sum_{i=1}^nG_{n,ii}x_{i,n}'$,

$\mathrm{E}\frac{1}{n}X_n'\epsilon_n(\epsilon_n'\epsilon_n - n\sigma_0^2) = \frac{\mu_3}{n}X_n'l_n$ and $\mathrm{E}\frac{1}{n}(W_nY_n)'\epsilon_n(\epsilon_n'\epsilon_n - n\sigma_0^2) = \frac{1}{n}\mu_3(G_nX_n\beta_0)'l_n + \frac{\mu_4-\sigma_0^4}{n}trG_n$. On the other hand, the information matrix is equal to

$$\Sigma_n = -\left(\mathrm{E}\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_{n0})}{\partial\theta\partial\theta'}\right) = \begin{pmatrix} \frac{1}{\sigma_0^2 n}X_n'X_n & * & * \\ \frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'X_n & \frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'(G_nX_n\beta_0) + \frac{1}{n}trG_n^sG_n & * \\ \mathbf{0}_{1\times k} & \frac{1}{\sigma_0^2 n}trG_n & \frac{1}{2\sigma_0^4} \end{pmatrix}. \quad (22)$$

This implies that $\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta}\cdot\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta'}\right) = \Sigma_n + \Omega_n$ with

$$\Omega_n = \begin{pmatrix} \mathbf{0}_{k\times k} & * & * \\ \frac{\mu_3}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}x_{i,n} & \frac{2\mu_3}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}(G_nX_n\beta_0)_i + \frac{\mu_4-3\sigma_0^4}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}^2 & * \\ \frac{\mu_3}{2\sigma_0^6 n}l_n'X_n & \frac{\mu_3}{2\sigma_0^6 n}l_n'G_nX_n\beta_0 + \frac{\mu_4-3\sigma_0^4}{2\sigma_0^6 n}trG_n & \frac{\mu_4-3\sigma_0^4}{4\sigma_0^8} \end{pmatrix}, \quad (23)$$

where l_n is $n\times 1$ vector of ones and μ_3, μ_4 are, respectively, the third and fourth moments of ϵ_{ni}. When ϵ_n is normally distributed, $\Omega_n = 0$ because $\mu_4 - 3\sigma_0^4 = 0$ and $\mu_3 = 0$.

From (22) and (23), we can use

$$\hat{\Sigma}_n = \begin{pmatrix} \frac{1}{\hat{\sigma}_n^2 n}X_n'X_n & * & * \\ \frac{1}{\hat{\sigma}_n^2 n}(G_n(\hat{\lambda}_n)X_n\hat{\beta}_n)'X_n & \frac{1}{\hat{\sigma}_n^2 n}(G_n(\hat{\lambda}_n)X_n\hat{\beta}_n)'(G_n(\hat{\lambda}_n)X_n\hat{\beta}_n) + \frac{1}{n}tr[G_n^s(\hat{\lambda}_n)G_n(\hat{\lambda}_n)] & * \\ \mathbf{0}_{1\times k} & \frac{1}{\hat{\sigma}_n^2 n}trG_n(\hat{\lambda}_n) & \frac{1}{2\hat{\sigma}_n^4} \end{pmatrix}, \quad (24)$$

and

$$\hat{\Omega}_n = \begin{pmatrix} \mathbf{0}_{k\times k} & * & * \\ \frac{\hat{\mu}_3}{\hat{\sigma}_n^4 n}\sum_{i=1}^n G_{n,ii}(\hat{\lambda}_n)x_{i,n} & \frac{2\hat{\mu}_3}{\hat{\sigma}_n^4 n}\sum_{i=1}^n G_{n,ii}(\hat{\lambda}_n)(G_n(\hat{\lambda}_n)X_n\beta_0)_i + \frac{\hat{\mu}_4-3\hat{\sigma}_n^4}{\hat{\sigma}_n^4 n}\sum_{i=1}^n G_{n,ii}^2(\hat{\lambda}_n) & * \\ \frac{\hat{\mu}_3}{2\hat{\sigma}_n^6 n}l_n'X_n & \frac{\hat{\mu}_3}{2\hat{\sigma}_n^6 n}l_n'G_n(\hat{\lambda}_n)X_n\beta_0 + \frac{\hat{\mu}_4-3\hat{\sigma}_n^4}{2\hat{\sigma}_n^6 n}trG_n(\hat{\lambda}_n) & \frac{\hat{\mu}_4-3\hat{\sigma}_n^4}{4\hat{\sigma}_n^8} \end{pmatrix}, \quad (25)$$

as the estimators for Σ_n and Ω_n. As $\hat{\beta}_n - \beta_0 = o_p(1)$, $\hat{\sigma}_n^2 - \sigma_0^2 = o_p(1)$, and $\psi_n(\hat{\lambda}_n - \lambda_{n0}) = o_p(1)$, the residual $\epsilon_n(\hat{\theta}_n) = (I_n + (\lambda_{n0} - \hat{\lambda}_n)G_n)\epsilon_n + (\lambda_{n0} - \hat{\lambda}_n)G_nX_n\beta_0 + X_n(\hat{\beta}_n - \beta_0)$ can be used to construct consistent estimates for σ_0^2, μ_3 and μ_4. For the other terms in $\hat{\Sigma}_n$ and $\hat{\Omega}_n$, we need to investigate $G_n(\hat{\lambda}_n)$, with $\psi_n(\hat{\lambda}_n - \lambda_{n0}) = o_p(1)$ and $\lambda_{n0} = 1 - \frac{1}{\psi_n}$. From the analysis in Section 2.1, $G_n(\hat{\lambda}_n)$ has two components similar to those of the decomposition of $G_n(\lambda_{n0})$ in (4) with $\psi_n\lambda_{n0}W_n^u$ being the unstable component. This implies that the asymptotic results in Lemma 1 about $G_n(\lambda_{n0})$ can also be applied to $G_n(\hat{\lambda}_n)$. Hence,

$$\Upsilon_{\theta,n}^{-1}(\hat{\Sigma}_n - \Sigma_n)\Upsilon_{\theta,n}^{-1} = o_p(1), \text{ and } \Upsilon_{\theta,n}^{-1}(\hat{\Omega}_n - \Omega_n)\Upsilon_{\theta,n}^{-1} = o_p(1). \quad (26)$$

Appendix C: Proofs for Theorems

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

$$\lim_{n\to\infty} P(\inf_{\lambda\in\Lambda, |\psi_n(\lambda-\lambda_{n0})|\geq\epsilon} [\ln L_n(\lambda_{n0}) - \ln L_n(\lambda)] > 0) = 1. \tag{27}$$

This is a sufficient condition which guarantees 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 an $\bar{\epsilon} > 0$ 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 to (27). Thus, $\hat{\lambda}_n$ is a ψ_n-consistent estimate of λ_{n0} if (27) is satisfied. The subsequent sections establish (27).

The concentrated log likelihood of our model is $\ln L_n(\lambda)$ in (9). For $\hat{\sigma}_n^2(\lambda)$ in (8), using $S_n(\lambda)Y_n = S_n(\lambda)S_n^{-1}(\lambda_{n0})(X_n\beta_0 + \epsilon_n)$ and $S_n(\lambda)S_n^{-1}(\lambda_{n0}) = I_n - (\lambda - \lambda_{n0})G_n$,

$$\begin{aligned}\hat{\sigma}_n^2(\lambda) &= \frac{1}{n}(\lambda - \lambda_{n0})^2(G_nX_n\beta_0)'M_nG_nX_n\beta_0 + \frac{1}{n}\epsilon_n'(S_n(\lambda)S_n^{-1}(\lambda_{n0}))'M_nS_n(\lambda)S_n^{-1}(\lambda_{n0})\epsilon_n \\ &\quad -2\frac{\lambda - \lambda_{n0}}{n}\epsilon_n'(S_n(\lambda)S_n^{-1}(\lambda_{n0}))'M_nG_nX_n\beta_0. \end{aligned}\tag{28}$$

Correspondingly, we have $Q_n(\lambda) = \max_{\beta,\sigma^2} \mathrm{E}\ln L_n(\theta) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\sigma_n^{*2}(\lambda) + \ln|S_n(\lambda)|$, where $\sigma_n^{*2}(\lambda)$ is in (20). 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

$$(i)\ \frac{1}{n}[\ln L_n(\lambda_{n0}) - Q_n(\lambda_{n0})] \xrightarrow{p} 0, \tag{29}$$

$$(ii)\ \sup_{\lambda\in\Lambda}\frac{1}{n}|\ln L_n(\lambda) - Q_n(\lambda)| \xrightarrow{p} 0, \tag{30}$$

$$(iii)\ \liminf_{n\to\infty}\inf_{\lambda\in\Lambda,|\psi_n(\lambda-\lambda_{n0})|\geq\epsilon}[Q_n(\lambda_{n0}) - Q_n(\lambda)] > 0. \tag{31}$$

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 quadratic functions of 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 $\hat{\sigma}_n^2(\lambda)$ 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

$$d_{n1}(\lambda) = \ln\sigma_0^2 - \frac{2}{n}\ln|S_n(\lambda_{n0})| - \ln\sigma_n^2(\lambda) + \frac{2}{n}\ln|S_n(\lambda)|, \tag{32}$$

$$d_{n2}(\lambda) = \ln[(\lambda_{n0} - \lambda)^2\frac{1}{n\sigma_n^2(\lambda)}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + 1]. \tag{33}$$

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

To prove (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 (17) and $\Lambda_n^{(\mathfrak{M},\epsilon)}$ is defined as

$$\Lambda_n^{(\mathfrak{M},\epsilon)} = \{\lambda | \lambda \in \Lambda,\ \epsilon \leq |\psi_n(\lambda_{n0} - \lambda)| \leq \mathfrak{M}\}. \tag{34}$$

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 $\sigma_n^{*2}(\lambda) - \sigma_n^2(\lambda) = (\lambda_{n0} - \lambda)^2 (G_n X_n \beta_0)' M_n (G_n X_n \beta_0)$, 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), $(\lambda_{n0} - \lambda)^2 \frac{1}{n\sigma_n^2(\lambda)}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) = \frac{[\psi_n(\lambda_{n0}-\lambda)]^2}{\sigma_n^2(\lambda)} \frac{(G_n X_n \beta_0)' M_n (G_n X_n \beta_0)}{n\psi_n^2}$ is strictly positive and bounded away from zero in the limit, by using Lemma 4 and Assumption 7. 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), $d_{n2}(\lambda) = \ln\{[\psi_n(\lambda_{n0} - \lambda)]^2 \frac{1}{\sigma_n^2(\lambda) n \psi_n^2}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) + 1\}$. From Lemma 1 and under Assumption 7, $\frac{1}{n\psi_n^2}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0)$ 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 (27).

Using (7), we have $\hat{\beta}_n = \beta_0 + (\frac{1}{n}X_n' X_n)^{-1}[\frac{1}{n}X_n' \epsilon_n - \psi_n(\hat{\lambda}_n - \lambda_{n0})\frac{1}{n\psi_n}X_n' G_n \epsilon_n]$. As $\hat{\lambda}_n$ is ψ_n-consistent, $\hat{\beta}_n$ is consistent. Also, from (8), 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 $Y_n = Z_n\delta_{n0} + \epsilon_n$, we have $\hat{\sigma}_n^2 = \frac{1}{n}(Z_n \Upsilon_{\delta,n}^{-1} \cdot \Upsilon_{\delta,n}(\delta_{n0} - \hat{\delta}_n) + \epsilon_n)'(Z_n \Upsilon_{\delta,n}^{-1} \cdot \Upsilon_{\delta,n}(\delta_{n0} - \hat{\delta}_n) + \epsilon_n)$. Using Lemma 2 and the consistency rates of $\hat{\delta}_n$, the consistency of $\hat{\sigma}_n^2$ follows. ■

Proof for Theorem 2 When Assumption 7 fails, (29) and (30) still hold. However, the identification uniqueness condition (31), 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 (32), $d_{n1}(\lambda) = \ln\sigma_0^2 - \frac{2}{n}\ln|S_n(\lambda_{n0})| - \ln\sigma_n^2(\lambda) + \frac{2}{n}\ln|S_n(\lambda)|$. In the proof for Theorem 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 shall show that when $0 < \lim_{n\to\infty}\frac{m_n}{n} \leq 1/2$ or $\lim_{n\to\infty}\frac{m_n}{n} \leq \lim_{n\to\infty}\frac{1}{n}tr(W_n'^u W_n^u)$, $\lim_{n\to\infty} d_{n1}(\lambda) < 0$.

First, we can divide the situations into two cases.

Situation 1: Consider a 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 see that $\frac{1}{n}\sum_{i=m_n+1}^{n}[\ln(1-\lambda_{n0}d_{ni}) - \ln(1-\lambda_n d_{ni})]$ is 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. This means that, for any sequence which can converge to a limit value less than 1, it can 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 $d_{n1}(\lambda_n)$ will be bounded away from 0 (strictly negative).

From (32), $d_{n1}(\lambda_n) = \ln\sigma_0^2 - \frac{2}{n}\ln|S_n(\lambda_{n0})| - \ln(\sigma_n^2(\lambda_n)) + \frac{2}{n}\ln|S_n(\lambda_n)|$. Also, using $|S_n(\lambda_n)| = (1-\lambda_n)^{m_n}\prod_{i=m_n+1}^{n}(1-\lambda_n d_{ni})$, we have $\frac{1}{n}\ln|S_n(\lambda_{n0})| - \frac{1}{n}\ln|S_n(\lambda_n)| = \frac{m_n}{n}\ln\frac{1-\lambda_{n0}}{1-\lambda_n} + \frac{1}{n}\sum_{i=m_n+1}^{n}\ln(\frac{1-\lambda_{n0}d_{ni}}{1-\lambda_n d_{ni}})$. As $\frac{1-\lambda_{n0}}{1-\lambda_n} = \frac{1}{1+\psi_n(\lambda_{n0}-\lambda_n)}$ by using $1-\lambda_{n0} = \frac{1}{\psi_n}$, we have

$$d_{n1}(\lambda_n) = \ln\sigma_0^2 - \ln\sigma_n^2(\lambda_n) + 2\frac{m_n}{n}\ln(1+\psi_n(\lambda_{n0}-\lambda_n)) - \frac{2}{n}\sum_{i=m_n+1}^{n}\ln(\frac{1-\lambda_{n0}d_{ni}}{1-\lambda_n d_{ni}}). \tag{35}$$

For this case, we first show that $\frac{1}{n}\sum_{i=m_n+1}^{n}\ln(\frac{1-\lambda_{n0}d_{ni}}{1-\lambda_n d_{ni}})$ will go to zero, so it can be ignored. Intuitively, we can show that $\frac{1-\lambda_{n0}d_{ni}}{1-\lambda_n d_{ni}}$ converge to 1 uniformly in i. That is, for any $c > 0$, we want to show that for large enough n, $1-c \leq \frac{1-\lambda_{n0}d_{ni}}{1-\lambda_n d_{ni}} \leq 1+c$ for all i. The latter is equivalent to $-c \leq [(1-c)\lambda_n - \lambda_{n0}]d_{ni}$ and $[-\lambda_{n0} + (1+c)\lambda_n]d_{ni} \leq c$ for all i, which hold because both λ_{n0} and λ_n go to 1 and $|d_{ni}|$ 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 $d_{n1}(\lambda_n)$, which are combined into $d_{n1}^*(\lambda_n) = -\ln\{1 + 2[\psi_n(\lambda_{n0}-\lambda_n)]q_{n1} + [\psi_n(\lambda_{n0}-\lambda_n)]^2 q_{n2}\} + 2q_{n1}\ln[\psi_n(\lambda_{n0}-\lambda_n)+1]$, where $q_{n1} = \frac{m_n}{n}$ and $q_{n2} = \frac{1}{n}tr(W_n'^u W_n^u)$. 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 $2q_{n1}\ln(1+s_n^*) - \ln(1+2q_{1n}s_n^* + q_{2n}s_n^{*2}) \to 0$ as $n \to \infty$". Thus, 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 $s \in [0,\infty)$. Hence, if these two functions interact at a point different to zero or they have the same asymptotic value, 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 $\frac{dR(s)}{ds} = 2q_1 + 2q_2 s$ 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 are two cases under 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[19].

(2) $q_1 \leq q_2$:

The difference of $R(s)$ 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 $\lim_{n\to\infty} \frac{m_n}{n} \leq \lim_{n\to\infty} \frac{1}{n} tr(W_n'^u W_n^u)$. Therefore, the identification uniqueness condition (31) is proved. Combined with (29) and (30), the consistency follows. ■

Proof for Theorem 3 Similarly as Lee (2004), we have $\Upsilon_{\theta,n}^{-1} \frac{1}{n} \frac{\partial^2 \ln L_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'} \Upsilon_{\theta,n}^{-1} - E\Upsilon_{\theta,n}^{-1} \frac{1}{n} \frac{\partial^2 \ln L_n(\theta_{n0})}{\partial\theta\partial\theta'} \Upsilon_{\theta,n}^{-1} \xrightarrow{p} 0$. Using the CLT for the linear and quadratic form of ϵ_n in Kelejian and Prucha (2001), $\Upsilon_{\theta,n}^{-1} \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_{n0})}{\partial\theta} \xrightarrow{d} N(0, \Sigma + \Omega)$. Assumption 9 guarantees that Σ is nonsingular. Hence, $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0}) \xrightarrow{d} N(0, \Sigma^{-1} + \Sigma^{-1}\Omega\Sigma^{-1})$ follows from the Taylor expansion $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_n - \theta_{n0}) = -(\Upsilon_{\theta,n}^{-1} \frac{1}{n} \frac{\partial^2 \ln L_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'} \Upsilon_{\theta,n}^{-1})^{-1} \Upsilon_{\theta,n}^{-1} \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_{n0})}{\partial\theta}$. ■

Proof for Theorem 4 Using Assumptions 10 and CLT 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 $\hat{\sigma}^2_{n,2sls} = \frac{1}{n}(Y_n - Z_n\hat{\delta}_{n,2sls})'(Y_n - Z_n\hat{\delta}_{n,2sls})$ where $\hat{\delta}_{n,2sls} = (Z_n' P_{H_n} Z_n)^{-1} Z_n' P_{H_n} Y_n$. As $Y_n = Z_n\delta_{n0} + \epsilon_n$, we have, $\hat{\sigma}^2_{n,2sls} = \frac{1}{n}(Z_n(\delta_{n0} - \hat{\delta}_{n,2sls}) + \epsilon_n)'(Z_n(\delta_{n0} - \hat{\delta}_{n,2sls}) + \epsilon_n)$. Hence, $\sqrt{n}(\hat{\sigma}^2_{n,2sls} - \sigma_0^2) = \frac{1}{\sqrt{n}}(\epsilon_n'\epsilon_n - n\sigma_0^2) + o_p(1)$ using $\Upsilon_{\delta,n}$-consistency of $\hat{\delta}_{n,2sls}$, where the variance of $\epsilon_n'\epsilon_n$ is $n(\mu_4 - \sigma_0^4)$. The result follows. ■

Proof for Theorem 5 From (11), $\Upsilon_{\delta,n}\sqrt{n}(\hat{\delta}_{n,b2sls} - \delta_{n0}) = \left(\frac{1}{n}\Upsilon_{\delta,n}^{-1}\tilde{H}_n^{*\prime} Z_n \Upsilon_{\delta,n}^{-1}\right)^{-1} \frac{1}{\sqrt{n}}\Upsilon_{\delta,n}^{-1}\tilde{H}_n^{*\prime}\epsilon_n$. We can show that $\text{plim}_{n\to\infty} \frac{1}{n}\Upsilon_{\delta,n}^{-1}\tilde{H}_n^{*\prime}\epsilon_n = 0$ and $\text{plim}_{n\to\infty} \frac{1}{n}\Upsilon_{\delta,n}^{-1}\tilde{H}_n^{*\prime} Z_n \Upsilon_{\delta,n}^{-1} = \text{plim}_{n\to\infty} \frac{1}{n}\Upsilon_{\delta,n}^{-1} H_n^{*\prime} H_n^* \Upsilon_{\delta,n}^{-1}$, which provides consistency. As $\frac{1}{\sqrt{n}}\Upsilon_{\delta,n}^{-1}\tilde{H}_n^{*\prime}\epsilon_n \xrightarrow{d} N(0, \sigma_0^2 \lim_{n\to\infty} \frac{1}{n}\Upsilon_{\delta,n}^{-1} H_n^{*\prime} H_n^* \Upsilon_{\delta,n}^{-1})$, we have the asymptotic distribution of $\hat{\delta}_{n,b2sls}$. ■

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, from (12), $\hat{\theta}_{n2} = \hat{\theta}_{n1} - (\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'})^{-1}(\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta} + \frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'}(\hat{\theta}_{n1} - \theta_{n0}))$. Therefore, $\hat{\theta}_{n2} - \theta_{n0} = (I_{k+2} - (\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'})^{-1}\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'})(\hat{\theta}_{n1} - \theta_{n0}) - (\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'})^{-1}\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta}$. Similar to Lee (2004), $\Upsilon_{\theta,n}(\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'})^{-1}\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1} = (\frac{1}{n}\Upsilon_{\theta,n}^{-1}\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1})^{-1}\Upsilon_{\theta,n}^{-1}\frac{1}{n}\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1} \xrightarrow{p} I_{k+2}$. Hence, $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_{n2} - \theta_{n0}) = (-\frac{1}{n}\Upsilon_{\theta,n}^{-1}\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1})^{-1}\frac{1}{\sqrt{n}}\Upsilon_{\theta,n}^{-1}\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta} + o_p(1)$ by using $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_{n1} - \theta_{n0}) = O_p(1)$. From the proof of Theorem 3, we have established, $\frac{1}{\sqrt{n}}\Upsilon_{\theta,n}^{-1}\frac{\partial \ln L_n(\theta_{n0})}{\partial\theta} \xrightarrow{d} N(0, \Sigma + \Omega)$. Hence, as $-\frac{1}{n}\Upsilon_{\theta,n}^{-1}\frac{\partial^2 \ln L_n(\hat{\theta}_{n1})}{\partial\theta\partial\theta'}\Upsilon_{\theta,n}^{-1} \xrightarrow{p} \Sigma$, $\sqrt{n}\Upsilon_{\theta,n}(\hat{\theta}_{n2} - \theta_{n0}) \xrightarrow{d} N(0, \Sigma^{-1}(\Sigma + \Omega)\Sigma^{-1})$. ■

Table 1: 2SLSEs and MLEs with normal disturbances: $n = 49$

Method	λ	β_1	β_2	β_3	
	Mean (SD)[RMSE]{CI}	Mean (SD)[RMSE]{CI}	Mean (SD)[RMSE]{CI}	Mean (SD)[RMSE]{CI}	CS
λ_{n0}=0.6					
2SLS	0.6504 (0.1565)[0.1644]{0.8720}	-0.9911 (0.1546)[0.1548]{0.9340}	0.0109 (0.1528)[0.1532]{0.9370}	0.9944 (0.1511)[0.1512]{0.9310}	0.8090
B2SLS	0.5399 (0.3735)[0.3783]{0.8609}	-1.0000 (0.1572)[0.1572]{0.9355}	0.0106 (0.1579)[0.1583]{0.9375}	1.0002 (0.1567)[0.1567]{0.9284}	0.8024
MLE	0.5705 (0.1117)[0.1156]{0.9390}	-1.0050 (0.1523)[0.1524]{0.9330}	0.0118 (0.1530)[0.1534]{0.9360}	1.0063 (0.1492)[0.1494]{0.9340}	0.9030
LMLE	0.5885 (0.1198)[0.1203]{0.9130}	-1.0030 (0.1528)[0.1529]{0.9345}	0.0116 (0.1528)[0.1532]{0.9355}	1.0050 (0.1491)[0.1492]{0.9325}	0.8317
λ_{n0}=0.9					
2SLS	0.9258 (0.0739)[0.0782]{0.7740}	-0.9913 (0.1551)[0.1553]{0.9370}	0.0110 (0.1530)[0.1534]{0.9350}	0.9946 (0.1512)[0.1513]{0.9310}	0.7170
B2SLS	0.8940 (0.4697)[0.4697]{0.8439}	-1.0092 (0.2180)[0.2183]{0.9355}	0.0134 (0.1739)[0.1744]{0.9367}	1.0088 (0.2176)[0.2177]{0.9276}	0.7817
MLE	0.8775 (0.0549)[0.0593]{0.9300}	-1.0105 (0.1536)[0.1539]{0.9340}	0.0117 (0.1541)[0.1545]{0.9320}	1.0115 (0.1513)[0.1517]{0.9260}	0.9270
LMLE	0.8890 (0.0569)[0.0579]{0.8484}	-1.0094 (0.1541)[0.1544]{0.9310}	0.0128 (0.1528)[0.1534]{0.9378}	1.0114 (0.1534)[0.1538]{0.9208}	0.7715
λ_{n0}=0.99					
2SLS	0.9947 (0.0279)[0.0283]{0.8600}	-1.0003 (0.1538)[0.1538]{0.9390}	0.0120 (0.1541)[0.1545]{0.9360}	1.0040 (0.1501)[0.1502]{0.9350}	0.7640
B2SLS	0.9653 (0.4890)[0.4896]{0.9108}	-0.9856 (0.5680)[0.5682]{0.9259}	0.0229 (0.2426)[0.2437]{0.9314}	1.0013 (0.5243)[0.5243]{0.9218}	0.8230
MLE	0.9833 (0.0201)[0.0211]{0.9110}	-1.0084 (0.1526)[0.1529]{0.8680}	0.0121 (0.1541)[0.1546]{0.8560}	1.0112 (0.1507)[0.1511]{0.8560}	0.9290
LMLE	0.9827 (0.0194)[0.0207]{0.8752}	-1.0103 (0.1543)[0.1546]{0.9273}	0.0130 (0.1537)[0.1542]{0.9273}	1.0190 (0.1539)[0.1551]{0.9273}	0.7874
λ_{n0}=0.999					
2SLS	0.9996 (0.0065)[0.0065]{0.8770}	-1.0042 (0.1536)[0.1536]{0.9350}	0.0124 (0.1550)[0.1555]{0.9300}	1.0077 (0.1498)[0.1500]{0.9380}	0.7960
B2SLS	0.9993 (0.0361)[0.0361]{0.9493}	-1.0061 (0.1543)[0.1544]{0.9252}	0.0156 (0.1604)[0.1612]{0.9279}	1.0123 (0.1580)[0.1585]{0.9252}	0.8318
MLE	0.9982 (0.0049)[0.0049]{0.9150}	-1.0048 (0.1520)[0.1520]{0.6950}	0.0120 (0.1530)[0.1534]{0.6780}	1.0075 (0.1489)[0.1491]{0.6720}	0.9570
LMLE	0.9979 (0.0038)[0.0039]{0.8745}	-1.0115 (0.1532)[0.1536]{0.9279}	0.0152 (0.1557)[0.1564]{0.9252}	1.0173 (0.1523)[0.1533]{0.9319}	0.7944

Note:

1. $\beta_{01} = -1$, $\beta_{02} = 0$ and $\beta_{03} = 1$, rep=1000.
2. We report all the estimates where the estimated $\hat{\lambda}_n$ could be greater than 1 in absolute value. For the B2SLS and LMLE which use initial estimates from 2SLS, we use the initial estimates that are smaller than 1; so the number of repetitions for them is the number of 2SLSEs with $\hat{\lambda}_n$ less than one.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, 749 for 2SLS, B2SLS, MLE and LMLE

Table 2: 2SLSEs and MLEs with normal disturbances: $n = 245$

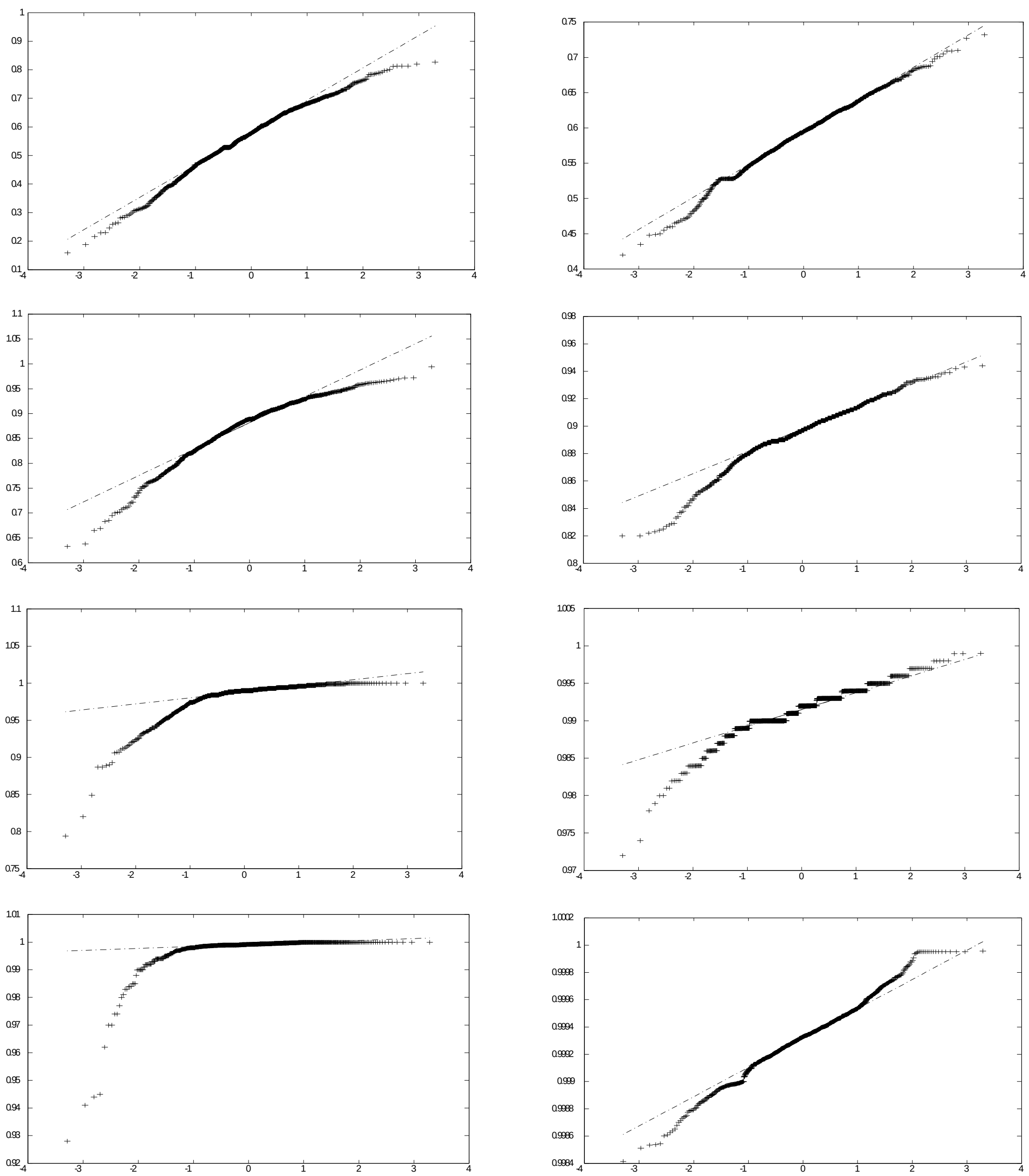
Method	λ	β_1	β_2	β_3	
	Mean (SD)[RMSE]{CI}	Mean (SD)[RMSE]{CI}	Mean (SD)[RMSE]{CI}	Mean (SD)[RMSE]{CI}	CS
λ_{n0}=0.6					
2SLS	0.6120 (0.0623)[0.0635]{0.9460}	-0.9982 (0.0653)[0.0653]{0.9490}	-0.0024 (0.0665)[0.0666]{0.9350}	0.9977 (0.0650)[0.0651]{0.9470}	0.9260
B2SLS	0.5950 (0.0650)[0.0652]{0.8990}	-1.0007 (0.0652)[0.0652]{0.9450}	-0.0025 (0.0668)[0.0668]{0.9340}	1.0003 (0.0649)[0.0649]{0.9450}	0.9030
MLE	0.5924 (0.0473)[0.0479]{0.9410}	-1.0019 (0.0647)[0.0648]{0.9480}	-0.0023 (0.0667)[0.0668]{0.9370}	1.0015 (0.0646)[0.0646]{0.9460}	0.9520
LMLE	0.5979 (0.0465)[0.0465]{0.9400}	-1.0010 (0.0647)[0.0647]{0.9470}	-0.0023 (0.0666)[0.0667]{0.9380}	1.0006 (0.0645)[0.0645]{0.9490}	0.9260
λ_{n0}=0.9					
2SLS	0.9053(0.0283)[0.0288]{0.9210}	-0.9978 (0.0659)[0.0659]{0.9500}	-0.0024 (0.0664)[0.0664]{0.9370}	0.9971 (0.0654)[0.0655]{0.9520}	0.9140
B2SLS	0.9078 (0.4730)[0.4731]{0.8920}	-0.9985 (0.1595)[0.1595]{0.9410}	-0.0054 (0.1133)[0.1134]{0.9370}	1.0024 (0.0652)[0.0653]{0.9480}	0.8950
MLE	0.8958 (0.0198)[0.0202]{0.9360}	-1.0028 (0.0649)[0.0650]{0.9460}	-0.0023 (0.0668)[0.0669]{0.9360}	1.0023 (0.0646)[0.0647]{0.9470}	0.9580
LMLE	0.8981 (0.0200)[0.0201]{0.9290}	-1.0017 (0.0650)[0.0651]{0.9460}	-0.0023 (0.0667)[0.0667]{0.9380}	1.0012 (0.0647)[0.0647]{0.9480}	0.9100
λ_{n0}=0.99					
2SLS	0.9909 (0.0061)[0.0061]{0.8310}	-0.9990 (0.0654)[0.0655]{0.9470}	-0.0023 (0.0664)[0.0665]{0.9410}	0.9985 (0.0648)[0.0648]{0.9520}	0.8320
B2SLS	0.9875 (0.0507)[0.0508]{0.8661}	-1.0044 (0.0736)[0.0737]{0.9456}	-0.0032 (0.0871)[0.0871]{0.9425}	1.0022 (0.0676)[0.0676]{0.9446}	0.8724
MLE	0.9914 (0.0030)[0.0033]{0.9980}	-0.9980 (0.0640)[0.0641]{0.7800}	-0.0020 (0.0664)[0.0665]{0.8290}	0.9974 (0.0640)[0.0641]{0.7700}	0.9630
LMLE	0.9899 (0.0033)[0.0033]{0.8755}	-1.0025 (0.0645)[0.0645]{0.9477}	-0.0024 (0.0665)[0.0665]{0.9425}	1.0012 (0.0640)[0.0640]{0.9456}	0.8630
λ_{n0}=0.999					
2SLS	0.9991 (0.0009)[0.0009]{0.8180}	-0.9996 (0.0653)[0.0653]{0.9500}	-0.0023 (0.0666)[0.0666]{0.9430}	0.9991 (0.0645)[0.0645]{0.9560}	0.8200
B2SLS	0.9989 (0.0015)[0.0015]{0.8687}	-1.0027 (0.0643)[0.0643]{0.9509}	-0.0014 (0.0679)[0.0679]{0.9370}	1.0022 (0.0643)[0.0643]{0.9434}	0.8741
MLE	0.9993 (0.0003)[0.0004]{1.0000}	-0.9961 (0.0639)[0.0640]{0.8350}	-0.0019 (0.0663)[0.0663]{0.7350}	0.9954 (0.0640)[0.0642]{0.8610}	0.9590
LMLE	0.9990 (0.0004)[0.0004]{0.8538}	-1.0025 (0.0641)[0.0642]{0.9498}	-0.0024 (0.0673)[0.0673]{0.9402}	1.0009 (0.0639)[0.0639]{0.9424}	0.8474

Note:

1. $\beta_{01} = -1$, $\beta_{02} = 0$ and $\beta_{03} = 1$, rep=1000.
2. We report all the estimates where the estimated $\hat{\lambda}_n$ could be greater than 1 in absolute value. For the B2SLS and LMLE which use initial estimates from 2SLS, we use the 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 $\hat{\lambda}_n$: MLE

$n = 49$ $\qquad$ $n = 245$

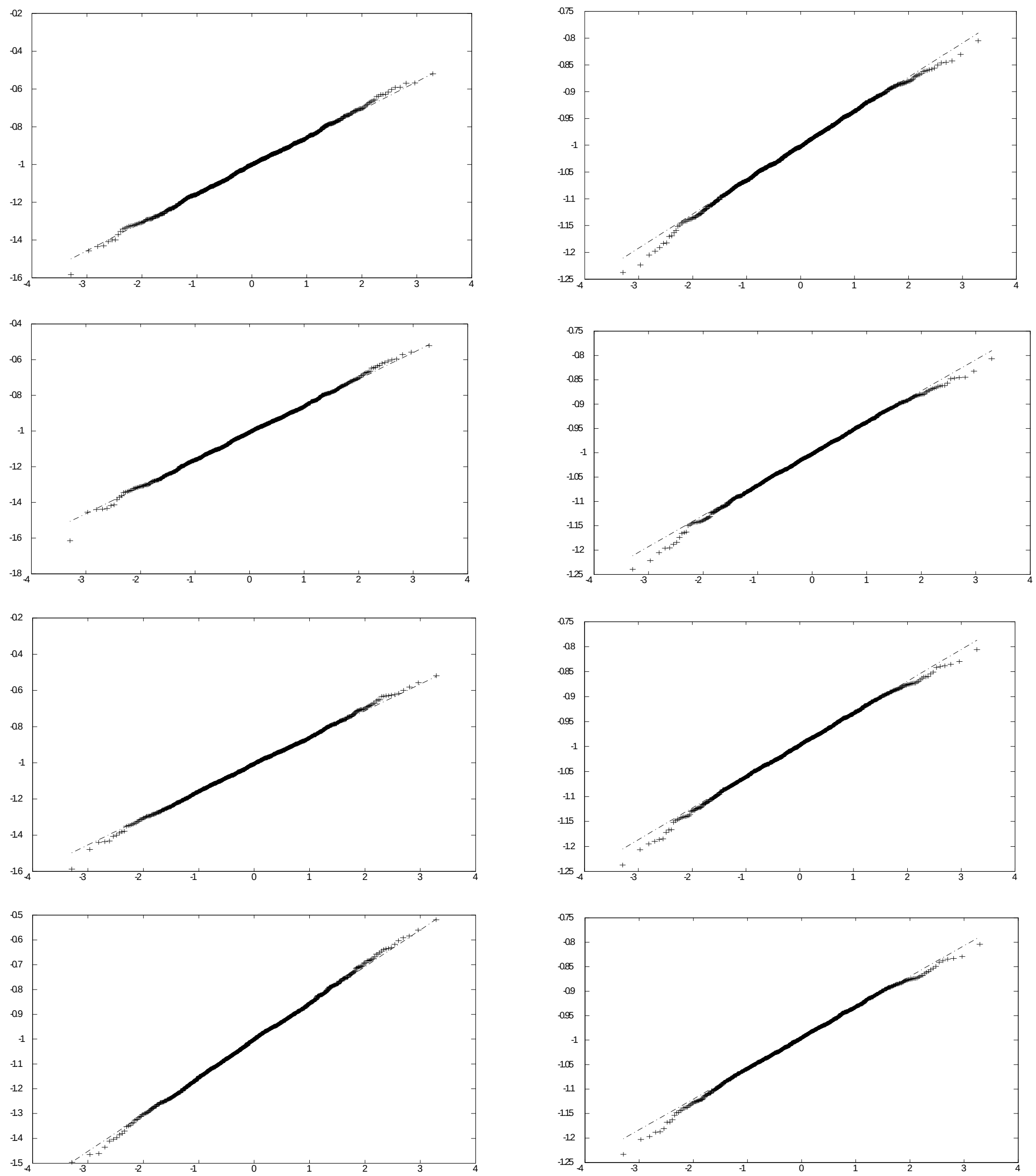


Note:

1. From 1st row to fourth row are $\lambda_{n0} = 0.6, 0.9, 0.99, 0.999$.

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

$n = 49$ $\qquad$ $n = 245$



Note:

1. From 1st row to fourth row are $\lambda_{n0} = 0.6, 0.9, 0.99, 0.999$.

Table 5: 2SLSEs and MLEs with exponential disturbances: $n = 245$

Method	λ	β_1	β_2	β_3	
	Mean (SD)[RMSE]{CI}	Mean (SD)[RMSE]{CI}	Mean (SD)[RMSE]{CI}	Mean (SD)[RMSE]{CI}	CS
λ_{n0}=0.6					
2SLS	0.6085 (0.0689)[0.0694]{0.9410}	-0.9970 (0.0663)[0.0664]{0.9470}	0.0025 (0.0646)[0.0647]{0.9430}	0.9968 (0.0654)[0.0655]{0.9490}	0.8770
B2SLS	0.5915 (0.0724)[0.0729]{0.8830}	-0.9994 (0.0663)[0.0663]{0.9470}	0.0025 (0.0647)[0.0648]{0.9440}	0.9993 (0.0655)[0.0655]{0.9510}	0.8570
MLE	0.5909 (0.0490)[0.0498]{0.9420}	-1.0005 (0.0658)[0.0658]{0.9480}	0.0025 (0.0647)[0.0648]{0.9430}	1.0007 (0.0653)[0.0653]{0.9520}	0.9220
LMLE	0.5962 (0.0489)[0.0491]{0.9370}	-0.9996 (0.0658)[0.0658]{0.9490}	0.0025 (0.0647)[0.0647]{0.9430}	0.9998 (0.0653)[0.0653]{0.9500}	0.8860
λ_{n0}=0.9					
2SLS	0.9041(0.0301)[0.0304]{0.9040}	-0.9965 (0.0665)[0.0666]{0.9490}	0.0025 (0.0646)[0.0647]{0.9430}	0.9963 (0.0660)[0.0661]{0.9510}	0.8600
B2SLS	0.8928 (0.0330)[0.0337]{0.8690}	-1.0012 (0.0667)[0.0667]{0.9500}	0.0026 (0.0649)[0.0649]{0.9440}	1.0013 (0.0658)[0.0658]{0.9500}	0.8599
MLE	0.8954 (0.0202)[0.0207]{0.9310}	-1.0013 (0.0661)[0.0661]{0.9490}	0.0024 (0.0648)[0.0649]{0.9440}	1.0015 (0.0657)[0.0657]{0.9530}	0.9510
LMLE	0.8976 (0.0209)[0.0210]{0.9149}	-1.0003 (0.0660)[0.0660]{0.9479}	0.0024 (0.0647)[0.0648]{0.9429}	1.0003 (0.0656)[0.0656]{0.9510}	0.8799
λ_{n0}=0.99					
2SLS	0.9910 (0.0064)[0.0065]{0.8150}	-0.9970 (0.0661)[0.0662]{0.9470}	0.0022 (0.0648)[0.0648]{0.9440}	0.9971 (0.0658)[0.0658]{0.9500}	0.7930
B2SLS	0.9889 (0.0121)[0.0122]{0.8411}	-1.0009 (0.0665)[0.0665]{0.9481}	0.0014 (0.0659)[0.0659]{0.9449}	1.0005 (0.0662)[0.0665]{0.9460}	0.8475
MLE	0.9914 (0.0031)[0.0033]{0.9960}	-0.9962 (0.0655)[0.0657]{0.7750}	0.0023 (0.0646)[0.0646]{0.8370}	0.9964 (0.0653)[0.0654]{0.7790}	0.9560
LMLE	0.9898 (0.0034)[0.0034]{0.8591}	-1.0001 (0.0656)[0.0656]{0.9460}	0.0022 (0.0636)[0.0636]{0.9481}	0.9994 (0.0653)[0.0656]{0.9492}	0.8347
λ_{n0}=0.999					
2SLS	0.9991 (0.0008)[0.0009]{0.7990}	-0.9974 (0.0661)[0.0661]{0.9480}	0.0021 (0.0650)[0.0650]{0.9430}	0.9976 (0.0656)[0.0657]{0.9490}	0.7700
B2SLS	0.9989 (0.0011)[0.0011]{0.8529}	-1.0000 (0.0663)[0.0663]{0.9495}	0.0016 (0.0659)[0.0659]{0.9418}	0.9998 (0.0656)[0.0656]{0.9429}	0.8474
MLE	0.9993 (0.0003)[0.0004]{1.0000}	-0.9941 (0.0654)[0.0657]{0.8260}	0.0023 (0.0646)[0.0646]{0.7380}	0.9944 (0.0651)[0.0653]{0.8440}	0.9540
LMLE	0.9990 (0.0004)[0.0004]{0.8595}	-0.9998 (0.0660)[0.0660]{0.9462}	0.0022 (0.0637)[0.0637]{0.9462}	0.9993 (0.0654)[0.0654]{0.9506}	0.8299

Note:

1. $\beta_{01} = -1$, $\beta_{02} = 0$ and $\beta_{03} = 1$, rep=1000.
2. We report all the estimates where the estimated $\hat{\lambda}_n$ could be greater than 1 in absolute value. For the B2SLS and LMLE which use initial estimates from 2SLS, we use the 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 $\lambda_{n0} = 0.9$, the numbers of estimates $\hat{\lambda}_n$ that are smaller than 1 are, respectively, 999, 999, 1000, 999 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, 944, 938, 1000, 944 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, 911, 907, 1000, 911 for 2SLS, B2SLS, MLE and LMLE.

Table 6: 2SLSEs and MLEs under misspecifications: $n = 245$

Method	λ	β_1	β_2	β_3
	Mean (SD)[RMSE]	Mean (SD)[RMSE]	Mean (SD)[RMSE]	Mean (SD)[RMSE]
(1) λ_0=0.6, γ_0=0.2, ρ_0=0.2				
2SLS	0.9860 (0.0580)[0.3903]	-0.9612 (0.1094)[0.1161]	-0.0058 (0.1107)[0.1109]	0.9594 (0.1151)[0.1221]
B2SLS	0.9838 (0.7386)[0.8324]	-0.9661 (0.3321)[0.3338]	-0.0057 (0.5522)[0.5523]	0.9823 (0.4301)[0.4305]
MLE	0.9844 (0.0107)[0.3845]	-0.9439 (0.1048)[0.1188]	-0.0060 (0.1083)[0.1085]	0.9438 (0.1126)[0.1258]
LMLE	0.9704 (0.0265)[0.3713]	-0.9606 (0.1084)[0.1153]	-0.0082 (0.1071)[0.1074]	0.9510 (0.1114)[0.1217]
(2) λ_0=0.2, γ_0=0.4, ρ_0=0.4				
2SLS	0.9520 (0.1297)[0.7631]	-0.9958 (0.1415)[0.1415]	-0.0069 (0.1431)[0.1433]	0.9966 (0.1465)[0.1467]
B2SLS	0.9585 (3.5761)[3.6556]	-1.0948 (1.0068)[1.0113]	0.0198 (1.2044)[1.2046]	0.9811 (0.5693)[0.5696]
MLE	0.9530 (0.0184)[0.7532]	-0.9755 (0.1365)[0.1387]	-0.0075 (0.1410)[0.1412]	0.9777 (0.1451)[0.1468]
LMLE	0.9268 (0.0638)[0.7296]	-0.9965 (0.1396)[0.1396]	-0.0085 (0.1405)[0.1407]	0.9894 (0.1420)[0.1424]
(3) λ_0=0, γ_0=0.5, ρ_0=0.5				
2SLS	0.9330 (0.1658)[0.9476]	-1.0251 (0.1658)[0.1677]	-0.0082 (0.1687)[0.1689]	1.0276 (0.1713)[0.1735]
B2SLS	0.4063 (9.3034)[9.3122]	-1.1068 (0.6353)[2.6374]	-0.1644 (2.1524)[2.1587]	1.0967 (0.7844)[0.7904]
MLE	0.9331 (0.0222)[0.9334]	-1.0042 (0.1611)[0.1612]	-0.0087 (0.1667)[0.1670]	1.0079 (0.1705)[0.1707]
LMLE	0.9052 (0.0842)[0.9091]	-1.0260 (0.1632)[0.1652]	-0.0090 (0.1668)[0.1671]	1.0245 (0.1656)[0.1674]
(4) λ_0=0.628, γ_0=0.661, ρ_0=-0.296				
2SLS	0.9895 (0.0824)[0.3707]	-0.9561 (0.2248)[0.2291]	-0.0097 (0.2261)[0.2263]	0.9609 (0.2246)[0.2280]
B2SLS	0.9329 (0.4732)[0.5629]	-1.0043 (0.3319)[0.3319]	-0.0226 (0.3751)[0.3758]	0.9628 (0.3248)[0.3269]
MLE	0.9532 (0.0170)[0.3256]	-0.9443 (0.2215)[0.2284]	-0.0091 (0.2283)[0.2285]	0.9513 (0.2265)[0.2317]
LMLE	0.9515 (0.0346)[0.3254]	-0.9794 (0.2180)[0.2190]	-0.0163 (0.2254)[0.2260]	0.9683 (0.2225)[0.2248]
(5) λ_0=0, γ_0=0.3, ρ_0=0.3				
2SLS	0.1202 (0.2058)[0.2383]	-0.9973 (0.1180)[0.1181]	-0.0038 (0.1219)[0.1019]	0.9972 (0.1258)[0.1259]
B2SLS	-0.0276 (0.2827)[0.2840]	-0.9930 (0.1252)[0.1254]	-0.0026 (0.1277)[0.1277]	0.9914 (0.1338)[0.1341]
MLE	0.2999 (0.0815)[0.3107]	-1.0031 (0.1156)[0.1157]	-0.0047 (0.1201)[0.1202]	1.0034 (0.1231)[0.1231]
LMLE	0.2887 (0.0849)[0.3009]	-1.0023 (0.1156)[0.1156]	-0.0047 (0.1201)[0.1202]	1.0026 (0.1233)[0.1233]
(6) λ_0=0, γ_0=0.6, ρ_0=0				
2SLS	0.1064 (0.2799)[0.2994]	-0.9886 (0.1916)[0.1919]	-0.0080 (0.1939)[0.1941]	0.9953 (0.1946)[0.1947]
B2SLS	-0.0544 (0.3584)[0.3625]	-0.9845 (0.1956)[0.1963]	-0.0054 (0.1954)[0.1954]	0.9878 (0.2001)[0.2004]
MLE	-0.0013 (0.0981)[0.0981]	-1.0001 (0.1898)[0.1898]	-0.0057 (0.1944)[0.1945]	1.0051 (0.1926)[0.1927]
LMLE	0.0237 (0.1161)[0.1185]	-0.9983 (0.1902)[0.1903]	-0.0062 (0.1942)[0.1943]	1.0039 (0.1928)[0.1928]
(7) λ_0=0, γ_0=0, ρ_0=0.6				
2SLS	0.1108 (0.1858)[0.2163]	-1.0007 (0.0969)[0.0969]	-0.0014 (0.1019)[0.1019]	0.9966 (0.1046)[0.1047]
B2SLS	-0.0219 (0.2550)[0.2559]	-0.9968 (0.1034)[0.1035]	-0.0007 (0.1075)[0.1075]	0.9919 (0.1117)[0.1120]
MLE	0.3231 (0.0809)[0.3330]	-1.0056 (0.0940)[0.0942]	-0.0028 (0.0995)[0.0996]	1.0017 (0.1018)[0.1018]
LMLE	0.3042 (0.0840)[0.3156]	-1.0047 (0.0940)[0.0942]	-0.0027 (0.0995)[0.0995]	1.0009 (0.1019)[0.1019]

Note:

1. $\beta_{01} = -1$, $\beta_{02} = 0$ and $\beta_{03} = 1$, rep=1000.
2. We report all the estimates where the estimated $\hat{\lambda}_n$ could be greater than 1 in absolute value. For the B2SLS and LMLE which use initial estimates from 2SLS, we use the initial estimates regardless of whether it is larger or smaller than 1.
3. For SDPD DGP (4), the coefficients come from the empirical estimates in Keller and Shiue (2007).
4. For SDPD DGP (1), the numbers of estimates $\hat{\lambda}_n$ that are smaller than 1 are, respectively, 620, 523, 620 for 2SLS, B2SLS and LMLE. For SDPD DGP (2), those numbers are 660, 536, 660. For SDPD DGP (3), they are 684, 537, 684. For SDPD DGP (4), they are 571, 457, 571. For SDPD DGP (5), they are 1000, 988, 1000. For SDPD DGP (6), they are 1000, 984, 1000. For SDPD DGP (7), they are 1000, 994, 1000.