Econometrica, Vol. 72, No. 6 (November, 2004), 1899–1925

ASYMPTOTIC DISTRIBUTIONS OF QUASI-MAXIMUM LIKELIHOOD ESTIMATORS FOR SPATIAL AUTOREGRESSIVE MODELS

BY LUNG-FEI LEE[1]

This paper investigates asymptotic properties of the maximum likelihood estimator and the quasi-maximum likelihood estimator for the spatial autoregressive model. The rates of convergence of those estimators may depend on some general features of the spatial weights matrix of the model. It is important to make the distinction with different spatial scenarios. Under the scenario that each unit will be influenced by only a few neighboring units, the estimators may have $\sqrt{n}$-rate of convergence and be asymptotically normal. When each unit can be influenced by many neighbors, irregularity of the information matrix may occur and various components of the estimators may have different rates of convergence.

KEYWORDS: Spatial autoregression, maximum likelihood estimation, quasi-maximum likelihood estimator, rates of convergence, increasing-domain asymptotics, infill asymptotics.

1. INTRODUCTION

SPATIAL ECONOMETRICS CONSISTS of econometric techniques dealing with empirical economic problems caused by spatial autocorrelation in cross-sectional and/or panel data; see, e.g., the survey by Anselin and Bera (1998), and the books by Cliff and Ord (1973), Anselin (1988), and Cressie (1993). Possible dependence across spatial units is a relevant issue in urban, real estate, regional, public, agricultural, environmental economics, and industrial organization. To capture spatial dependence, the approaches in spatial econometrics are to impose structures on a model. One is in the domain of geostatistics where the spatial index is continuous (Conley (1999)). Another is where spatial sites form a countable lattice. In this paper, we are concerned about spatial models on lattices.

Among the lattice models, the class of spatial autoregressive (SAR) models by Cliff and Ord (1973) extends autocorrelation in time series to spatial dimensions. The spatial aspect of a SAR model has the distinguishing feature of simultaneity in econometric equilibrium models. Earlier development in testing and estimation of SAR models has been summarized in Anselin (1988), Cressie (1993), and Anselin and Bera (1998), among others. Recent empirical applications of the SAR model in the main stream economics journals include Case (1991), Case, Rosen, and Hines (1993), Besley and Case (1995), Brueckner (1998), Bell and Bockstael (2000), Bertrand, Luttmer, and Mullainathan (2000), and Topa (2001), among others. The SAR models can be estimated by the method of maximum likelihood (ML) (Ord (1975), Smirnov and Anselin (2001)) as well as methods of moments (Kelejian and Prucha (1999)). In this paper, we investigate asymptotic properties of the maximum likelihood estimator (MLE) and the quasi-maximum likelihood estimator (QMLE) for the SAR model under the normal

[1]I am grateful to the National Science Foundation, Economics Program for financial support under Grant SES-0111380. An earlier version of the paper was presented in seminars at HKUST, NWU, OSU, Princeton U., PSU, U. of Florida, U. of Illinois, and USC. I appreciate comments from participants of those seminars, referees, and the coeditor.

distributional specification. The QMLE is appropriate when the estimator is derived from a normal likelihood but the disturbances in the model are not truly normally distributed. In the existing literature, the MLE of such a model is implicitly regarded as having the familiar $\sqrt{n}$-rate of convergence as a usual MLE for a parametric statistical model with sample size n (see, e.g., the reviews by Anselin (1988) and Anselin and Bera (1998)). Manski (1993) has criticized the literature on the SAR model on the grounds that the equation of a SAR model does not specify how the spatial weights matrix should change as the sample size changes.[2]

Our investigation below provides a broader view of the asymptotic property of the MLE and the QMLE. It shows that the rates of convergence of the MLE and QMLE may depend on some general features of the spatial weights matrix of the model. The MLE and QMLE may indeed have a $\sqrt{n}$-rate of convergence and their limiting distributions are normal. But, under some circumstances, the estimators may have a low rate of convergence for some parameter components of the model and may even be inconsistent.

These results have some counterparts in spatial statistics. An asymptotic is called increasing-domain asymptotic when it is based on a growing observation region. It is called fixed-domain asymptotic (or infill asymptotic) when it is based on increasingly dense observations in a fixed and bounded region (Cressie (1993) and Stein (1999)). Mardia and Marshall (1984) and Cressie and Lahiri (1993) give consistency and asymptotic normality results for the MLE and related likelihood estimators under increasing-domain asymptotic for regression models with spatially correlated disturbances.[3] Ripley (1988) pointed out that for fixed-domain asymptotic, as interactions will increase with observations, there is no theoretical basis for the usual behavior of an MLE. No general results are available for the MLE under infill asymptotic (Cressie (1993, p. 101), Stein (1999)).

This paper is organized as follows. In Section 2, the spatial autoregressive model is presented and regularity conditions are specified. We make the important distinction between models with and without the presence of regressors. In Section 3, we show that when spatial varying regressors are really relevant, identification of parameters can be assured if there is no multicollinearity among the regressors and a spatially generated regressor. The MLE and QMLE can be $\sqrt{n}$-consistent and asymptotically normal under some regularity conditions on the spatial weights matrix. Section 4 considers the event of multicollinearity where the spatially generated regressor is collinear with the original regressors. Examples are given. Under such a circumstance, model parameters can be identified only through spatial correlation of outcomes. It is important to make the distinction with different spatial scenarios. Under the scenario that each unit will be influenced by only a few neighboring units, the MLE and QMLE may still have $\sqrt{n}$-rate of convergence and be asymptotic normal. Section 5 considers the spatial scenario that each unit can be influenced by many neighbors. In this situation, irregularity of the information matrix may occur and various components of the QMLEs may have different rates of convergence. This includes the MLE and QMLE for the (pure) SAR process. In Section 6, examples on the inconsistency of the QMLE are presented and this phenomena is related to the notion of infill asymptotic (Cressie (1993)). Section 7

[2]See footnote 7 in Manski (1993).

[3]Section 7.3.1 of Cressie (1993) provides a review of some related results under increasing domain asymptotic on the Markov random field.

provides the conclusions. Some useful lemmas and brief proofs are collected in the Appendix.[4]

2. SPATIAL AUTOREGRESSIVE MODELS AND QMLE

The SAR model is

$$Y_n = X_n\beta + \lambda W_n Y_n + V_n, \tag{2.1}$$

where n is the total number of spatial units, X_n is an $n \times k$ matrix of constant regressors, W_n is a specified constant spatial weights matrix, and V_n is an n-dimensional vector of i.i.d. disturbances with zero mean and finite variance σ^2. The weights may be based on physical distance, social networks, or "economic" distance (Case, Rosen, and Hines (1993)). This spatial model is an equilibrium model.[5] Let $\theta_0 = (\beta_0', \lambda_0, \sigma_0^2)'$ be the true parameter vector. Denote $S_n(\lambda) = I_n - \lambda W_n$ for any value of λ.[6] The equilibrium vector Y_n is

$$Y_n = S_n^{-1}(X_n\beta_0 + V_n), \tag{2.2}$$

where $S_n = S_n(\lambda_0)$ is nonsingular. When there are no regressors X_n in the model, it becomes a pure SAR process:

$$Y_n = \lambda W_n Y_n + V_n \tag{2.3}$$

and Y_n is simply derived from V_n. To emphasize the distinction of (2.1) and (2.3), the model with X_n in (2.1) is termed the mixed regressive, spatial autoregression model in Ord (1975) and Anselin (1988). Whether spatial varying regressors X_n in (2.1) are relevant or not plays a distinctive role in estimation. In the presence of spatial varying regressors X_n, in addition to the ML method, the method of instrumental variables (IV) can be used (Anselin (1988), Kelejian and Prucha (1998), and Lee (2002, 2003)). However, the IV estimation method will break down when all the spatial regressors are really irrelevant, and one cannot test the joint significance of the regressors in the IV framework (Kelejian and Prucha (1998)). These are so, because there are no valid IV's available when existing regressors are irrelevant. The ML method is still applicable. These features have interesting implications on model identification and asymptotic distribution of the MLE and QMLE.

Let $V_n(\delta) = Y_n - X_n\beta - \lambda W_n Y_n$, where $\delta = (\beta', \lambda)'$. Thus, $V_n = V_n(\delta_0)$. The log-likelihood function of (2.1) is

$$\ln L_n(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}V_n'(\delta)V_n(\delta), \tag{2.4}$$

[4]Detailed proofs can be found in the long version of this paper, which is available from the author's web site: *http://economics.sbs.ohio-state.edu/lee/*.

[5]Manski (1993) has introduced an endogenous social effect model where the expected values of spatial neighbors are used in place of W_nY_n in (2.1). The expected values satisfy social equilibrium equations and can be derived from them. Manski's model can be a competitive alternative to the SAR model. It is of interest to investigate model discrimination issues of these two models in future research.

[6]A list of frequently used notations in the text is summarized in the Appendix for reference.

where $\theta = (\beta', \lambda, \sigma^2)'$. The QMLE or MLE $\hat{\theta}_n$ is the extremum estimator derived from the maximization of (2.4). The estimation of the pure SAR process in (2.3) can be regarded as a constrained estimation of (2.1) by imposing $\beta = 0$. Computationally and analytically, it is convenient to work with the concentrated log-likelihood by concentrating out the β and σ^2. From the log-likelihood function (2.4), given λ, the QMLE of β is

$$\hat{\beta}_n(\lambda) = (X_n'X_n)^{-1}X_n'S_n(\lambda)Y_n, \tag{2.5}$$

and the QMLE of σ^2 is

$$\begin{aligned} \hat{\sigma}_n^2(\lambda) &= \frac{1}{n}[S_n(\lambda)Y_n - X_n\hat{\beta}_n(\lambda)]'[S_n(\lambda)Y_n - X_n\hat{\beta}_n(\lambda)] \\ &= \frac{1}{n}Y_n'S_n'(\lambda)M_nS_n(\lambda)Y_n, \end{aligned} \tag{2.6}$$

where $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$. The concentrated log-likelihood function of λ is

$$\ln L_n(\lambda) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\hat{\sigma}_n^2(\lambda) + \ln|S_n(\lambda)|. \tag{2.7}$$

The QMLE $\hat{\lambda}_n$ of λ maximizes the concentrated likelihood (2.7). The QMLEs of β and σ^2 are, respectively, $\hat{\beta}_n(\hat{\lambda}_n)$ and $\hat{\sigma}_n^2(\hat{\lambda}_n)$.

To provide a rigorous analysis of the QMLE, basic regularity conditions are assumed below. Additional regularity conditions will be subsequently listed.

ASSUMPTION 1: *The $\{v_i\}$, $i = 1, \ldots, n$, in $V_n = (v_1, \ldots, v_n)'$ are i.i.d. with mean zero and variance σ^2. Its moment $E(|v|^{4+\gamma})$ for some $\gamma > 0$ exits.*

ASSUMPTION 2: *The elements $w_{n,ij}$ of W_n are at most of order h_n^{-1}, denoted by $O(1/h_n)$, uniformly in all i, j,*[7] *where the rate sequence $\{h_n\}$ can be bounded or divergent. As a normalization, $w_{n,ii} = 0$ for all i.*

ASSUMPTION 3: *The ratio $h_n/n \to 0$ as n goes to infinity.*

ASSUMPTION 4: *The matrix S_n is nonsingular.*

ASSUMPTION 5: *The sequences of matrices $\{W_n\}$ and $\{S_n^{-1}\}$ are uniformly bounded in both row and column sums (Horn and Jonhson* (1985)).

ASSUMPTION 6: *The elements of X_n are uniformly bounded constants for all n. The $\lim_{n\to\infty} X_n'X_n/n$ exists and is nonsingular.*

ASSUMPTION 7: *$\{S_n^{-1}(\lambda)\}$ are uniformly bounded in either row or column sums, uniformly in λ in a compact parameter space Λ. The true λ_0 is in the interior of Λ.*

[7]That is, for some real constant c, there exists a finite integer N such that, for all $n \geq N$, $|h_n w_{n,ij}| < c$ for all i, j. See, e.g., White (1984, p. 14).

Assumptions 1–3 are the assumptions that provide the essential features of the disturbances and the weights matrix for the model. Assumptions 2 and 3 link directly the expression of W_n to the sample size n. Assumption 2 is always satisfied if $\{h_n\}$ is a bounded sequence. In some empirical applications, it is a practice to have W_n be row-normalized (Anselin (1988)) such that its ith row $w_{i,n} = (d_{i1}, d_{i2}, \ldots, d_{in})/\sum_{j=1}^n d_{ij}$, where $d_{ij} \geq 0$, represents a function of the spatial distance of the ith and jth units in some (characteristic) space. The weighting operation can be interpreted as an average of neighboring values. For a row-normalized weights matrix, as $d_{i,j}$ are nonnegative constants and uniformly bounded, if the $\sum_{j=1}^n d_{ij}$, $i = 1, \ldots, n$, are uniformly bounded away from zero at the rate h_n in the sense that $\sum_{j=1}^n d_{ij} = O(h_n)$ uniformly in i and $\liminf_{n\to\infty} h_n^{-1} \sum_{j=1}^n d_{ij} > c$, where c is a positive constant independent of i and n, the implied normalized weights matrix will have the property ascribed in Assumption 2. Assumption 3 excludes the cases where the $\sum_{j=1}^n d_{ij}$, $i = 1, \ldots, n$, diverge to infinity at a rate equal to or faster than the rate of the sample size n, because the MLE would likely be inconsistent for those cases. Examples will be provided later. Bell and Bockstael (2000) argue that row-normalization for the weights matrix may not be meaningful for real estate problems with microlevel data. Assumptions 2 and 3 are general in that they cover spatial weights matrices where elements are not restricted to be nonnegative and those that might not be row-normalized. Empirical examples that satisfy the above assumptions include conventional spatial weights matrices where neighboring units are defined by only a few adjacent ones, and models of Case (1991) where all spatial units in a district are neighbors of each other. For models with a few neighboring units, $\{h_n\}$ would be bounded. An important case that h_n might diverge to infinity and satisfies Assumptions 2 and 3 is that of Case (1991). In Case's model, "neighbors" refer to farmers who live in the same district. Suppose that there are R districts and there are m farmers in each district (for simplicity). The sample size is $n = mR$. Case assumed that in a district, each neighbor of a farmer is given equal weight. In that case, $W_n = I_R \otimes B_m$, where $B_m = (l_m l_m' - I_m)/(m-1)$, $\otimes$ is the Kronecker product, and l_m is an m-dimensional column vector of ones. For this example, $h_n = (m-1)$ and $h_n/n = (m-1)/(mR) = O(1/R)$. If sample size n increases by increasing both R and m, then h_n goes to infinity and h_n/n goes to zero as n tends to infinity.[8]

Assumption 4 guarantees that the system (2.1) has an equilibrium and Y_n has mean $S_n^{-1} X_n \beta_0$ and variance $\sigma_0^2 S_n^{-1} S_n'^{-1}$, where σ_0^2 is the true variance of v_i. Assumption 5 is originated by Kelejian and Prucha (1998, 1999, 2001).[9] The uniform boundedness of $\{W_n\}$ and $\{S_n^{-1}\}$ is a condition to limit the spatial correlation to a manageable degree. It plays an important role in the asymptotic properties of estimators for spatial econometric models. For example, it guarantees that the variances of Y_n are bounded as n goes to infinity. Some discussions on uniform boundedness are in Appendix A.

When the mixed regressive model is used for analyzing cross-sectional units, it is meaningful to assume that the regressors are bounded as in Assumption 6.[10] Multicollinearity among the regressors of X_n are ruled out. Without regressors, it is a pure spatial autoregressive process and Assumption 6 is irrelevant.

[8]Whether $\{h_n\}$ is a bounded or divergent sequence has interesting implications on the least square approach. The least squares estimators of β and λ are inconsistent when $\{h_n\}$ is bounded, but they can be consistent when $\{h_n\}$ is divergent.

[9]Related conditions have also been adopted in Pinkse (1999) in a different context.

[10]If not, it can be replaced by stochastic regressors with certain finite moment conditions.

The uniform boundedness condition of S_n^{-1} at λ_0 in Assumption 5 implies that $S_n^{-1}(\lambda)$ are uniformly bounded in both row and column sums uniformly in a neighborhood of λ_0 (see Appendix A). Assumption 7 is needed to deal with the nonlinearity of $\ln|S_n(\lambda)|$ as a function of λ in (2.4). As in Appendix A, if $\| W_n \| \leq 1$ for all n, where $\| \cdot \|$ is a matrix norm, then $\{\| S_n^{-1}(\lambda) \|\}$ are uniformly bounded in any subset of $(-1, 1)$ bounded away from the boundary. In particular, if W_n is a row-normalized matrix, $S_n^{-1}(\lambda)$ is uniformly bounded in row sums norm uniformly in any closed subset of $(-1, 1)$. For this case, Λ in Assumption 7 can be taken as a single closed set contained in $(-1, 1)$ for all n.[11] For the case in which W_n is not normalized but its eigenvalues are real, since the Jacobian $|S_n(\lambda)|$ in (2.4) will be positive if $-1/|\mu_{n,\min}| < \lambda < 1/\mu_{n,\max}$, where $\mu_{n,\min}$ and $\mu_{n,\max}$ are the minimum and maximum eigenvalues of W_n (Anselin (1988)), Λ can be a closed interval contained in $(-1/|\mu_{n,\min}|, 1/\mu_{n,\max})$ for all n. It is clear from (2.5) and (2.6) that β_0 and σ_0^2 will be identifiable once λ_0 is identified, and the parameter space of β_0 and σ_0^2 do not need to be specified.

3. MIXED REGRESSIVE, SPATIAL AUTOREGRESSIVE MODELS: THE REGULAR CASE

The presence of X_n in (2.1) is a distinctive feature of the mixed regressive SAR model. From (2.1) and (2.2), the reduced form equation of Y_n can be represented as

$$Y_n = X_n\beta_0 + \lambda_0 G_n X_n \beta_0 + S_n^{-1} V_n \tag{3.1}$$

because $I_n + \lambda_0 G_n = S_n^{-1}$, where $G_n = W_n S_n^{-1}$.

ASSUMPTION 8: *The* $\lim_{n\to\infty}(X_n, G_n X_n \beta_0)'(X_n, G_n X_n \beta_0)/n$ *exists and is nonsingular.*

This assumption requires that the generated regressors $G_n X_n \beta_0$ in (3.1) and X_n are not asymptotically multicollinear. It is a sufficient condition for global identification of θ_0. Define $Q_n(\lambda) = \max_{\beta,\sigma^2} E(\ln L_n(\theta))$. The optimal solutions of this maximization problem are $\beta_n^*(\lambda) = (X_n' X_n)^{-1} X_n' S_n(\lambda) S_n^{-1} X_n \beta_0$ and

$$\begin{aligned}\sigma_n^{*2}(\lambda) &= \frac{1}{n} E\{[S_n(\lambda)Y_n - X_n\beta_n^*(\lambda)]'[S_n(\lambda)Y_n - X_n\beta_n^*(\lambda)]\} \\ &= \frac{1}{n}\{(\lambda_0 - \lambda)^2 (G_n X_n \beta_0)' M_n (G_n X_n \beta_0) + \sigma_0^2 \operatorname{tr}[S_n'^{-1} S_n'(\lambda) S_n(\lambda) S_n^{-1}]\}.\end{aligned} \tag{3.2}$$

Hence,

$$Q_n(\lambda) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln \sigma_n^{*2}(\lambda) + \ln|S_n(\lambda)|. \tag{3.3}$$

Identification of λ_0 can be based on the maximum values of $\{(Q_n(\lambda)/n)\}$. With identification and uniform convergence of $[\ln L_n(\lambda) - Q_n(\lambda)]/n$ to zero on Λ, consistency of the QMLE $\hat{\theta}_n$ follows.

[11]On the other hand, Assumption 7 rules out implicitly the consideration of models where the true λ_0 is close to 1 or -1.

THEOREM 3.1: *Under Assumptions* 1–8, θ_0 *is globally identifiable and* $\hat{\theta}_n$ *is a consistent estimator of* θ_0.

The asymptotic distribution of the QMLE $\hat{\theta}_n$ can be derived from the Taylor expansion of

$$\frac{\partial \ln L_n(\hat{\theta}_n)}{\partial \theta} = 0 \quad \text{at } \theta_0.$$

The first-order derivatives of the log-likelihood function at θ_0 are

$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \beta} = \frac{1}{\sigma_0^2 \sqrt{n}} X_n' V_n,$$

$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \sigma^2} = \frac{1}{2\sigma_0^4 \sqrt{n}} (V_n' V_n - n\sigma_0^2),$$

and

$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \lambda} = \frac{1}{\sigma_0^2 \sqrt{n}} (G_n X_n \beta_0)' V_n + \frac{1}{\sigma_0^2 \sqrt{n}} \big(V_n' G_n V_n - \sigma_0^2 \operatorname{tr}(G_n)\big). \tag{3.4}$$

These are linear and quadratic functions of V_n. The asymptotic distribution of (3.4) may be derived from central limit theorems for linear-quadratic functions. For the case $\{h_n\}$ being a bounded sequence, the central limit theorem for linear-quadratic forms in Kelejian and Prucha (2001) is applicable (see, also, Giraitis and Taqqu (1998)). For the case in which $\lim_{n\to\infty} h_n = \infty$, $(1/(\sigma_0^2\sqrt{n}))(G_n X_n \beta_0)' V_n$ will dominate the quadratic term of $(1/\sqrt{n})(\partial \ln L_n(\theta_0))/\partial\lambda$ under Assumption 8. This occurs because

$$\operatorname{var}\left(\frac{1}{\sqrt{n}} V_n' G_n V_n\right) = O\left(\frac{1}{h_n}\right),$$

and hence,

$$\frac{1}{\sqrt{n}} \big(V_n' G_n V_n - \sigma_0^2 \operatorname{tr}(G_n)\big) = o_P(1)$$

while

$$\frac{1}{\sqrt{n}} (G_n X_n \beta_0)' V_n = O_P(1).$$

Under this situation, Kolmogorov's central limit theorem can be applied.

The variance matrix of $(1/\sqrt{n})\, \partial \ln L_n(\theta_0)/\partial\theta$ is

$$E\left(\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta} \cdot \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta'}\right) = -E\left(\frac{1}{n} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta\, \partial \theta'}\right) + \Omega_{\theta,n},$$

where

$$
(3.5)\qquad -E\left(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\,\partial\theta'}\right)
= \begin{pmatrix}
\frac{1}{\sigma_0^2 n}X_n'X_n & \frac{1}{\sigma_0^2 n}X_n'(G_nX_n\beta_0) & 0 \\
\frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'X_n & \frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'(G_nX_n\beta_0)+\frac{1}{n}\operatorname{tr}(G_n^sG_n) & \frac{1}{\sigma_0^2 n}\operatorname{tr}(G_n) \\
0 & \frac{1}{\sigma_0^2 n}\operatorname{tr}(G_n) & \frac{1}{2\sigma_0^4}
\end{pmatrix}
$$

with $G_n^s = G_n + G_n'$, is the average Hessian matrix (information matrix when v's are normal), and

$$
(3.6)\qquad \Omega_{\theta,n} = \begin{pmatrix}
0 & * & * \\
\frac{\mu_3}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}x_{i,n} & \frac{2\mu_3}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}G_{in}X_n\beta_0 + \frac{(\mu_4-3\sigma_0^4)}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}^2 & * \\
\frac{\mu_3}{2\sigma_0^6 n}l_n'X_n & \frac{1}{2\sigma_0^6 n}[\mu_3 l_n'G_nX_n\beta_0 + (\mu_4-3\sigma_0^4)\operatorname{tr}(G_n)] & \frac{(\mu_4-3\sigma_0^4)}{4\sigma_0^8}
\end{pmatrix}
$$

is a symmetric matrix with $\mu_j = E(v_i^j)$, $j = 2, 3, 4$, being, respectively, the second, third, and fourth moments of v, where G_{in} is the ith row of G_n, $G_{n,ij}$ is the (i, j)th entry of G_n, and $x_{i,n}$ is the ith row of X_n. Assumption 8 is sufficient to guarantee that the average Hessian matrix is nonsingular for large enough n. If V_n is normally distributed, $\Omega_{\theta,n} = 0$.

THEOREM 3.2: *Under Assumptions* 1–8, $\sqrt{n}(\hat\theta_n - \theta_0) \xrightarrow{D} N(0, \Sigma_\theta^{-1} + \Sigma_\theta^{-1}\Omega_\theta\Sigma_\theta^{-1})$, *where* $\Omega_\theta = \lim_{n\to\infty}\Omega_{\theta,n}$ *and*

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

which are assumed to exist. If the v_i*'s are normally distributed, then* $\sqrt{n}(\hat\theta_n - \theta_0) \xrightarrow{D} N(0, \Sigma_\theta^{-1})$.[12]

The asymptotic results in Theorems 3.1 and 3.2 are valid regardedless of whether $\{h_n\}$ is a bounded or divergent sequence. For the case in which $\lim_{n\to\infty} h_n = \infty$, because $G_{n,ij} = O(1/h_n)$, the matrices (3.5) and (3.6) can be simplified to

$$
\Omega_\theta = \lim_{n\to\infty}\begin{pmatrix}
0 & 0 & * \\
0 & 0 & * \\
\frac{\mu_3}{2\sigma_0^6 n}l_n'X_n & \frac{\mu_3}{2\sigma_0^6 n}l_n'(G_nX_n\beta_0) & \frac{(\mu_4-3\sigma_0^4)}{4\sigma_0^8}
\end{pmatrix}
$$

[12]The estimation of the asymptotic variance of $\hat\theta_n$ is trivial. The Σ_θ can be estimated by (3.5) evaluated at $\hat\theta_n$. The Ω_θ can be estimated with (3.6). For the QMLE, the extra moments μ_3 and μ_4 in $\Omega_{\theta,n}$ can be estimated by the third and fourth order empirical moments based on estimated residuals of the v's.

and

$$\Sigma_\theta = \lim_{n\to\infty} \begin{pmatrix} \frac{1}{\sigma_0^2 n} X_n' X_n & \frac{1}{\sigma_0^2 n} X_n'(G_n X_n \beta_0) & 0 \\ \frac{1}{\sigma_0^2 n} (G_n X_n \beta_0)' X_n & \frac{1}{\sigma_0^2 n} (G_n X_n \beta_0)'(G_n X_n \beta_0) & 0 \\ 0 & 0 & \frac{1}{2\sigma_0^4} \end{pmatrix}.$$

The presence of X_n and the linear independence of $G_n X_n \beta_0$ and X_n are the crucial conditions for the asymptotic results in Theorem 3.2, in particular, the $\sqrt{n}$-rate of convergence of $\hat{\theta}_n$.

When v's are normally distributed, $\hat{\theta}_n$ is the MLE. When $\{h_n\}$ is bounded, the MLE's $\hat{\lambda}_n$ and $\hat{\sigma}_n^2$ will be asymptotically dependent because $\lim_{n\to\infty} \operatorname{tr}(G_n)/n$ is finite and may not be zero. Anselin and Bera (1998) discussed the implication of this dependence on statistical inference problems. We note, however, that for the case in which $\{h_n\}$ is a divergent sequence, $\lim_{n\to\infty} \operatorname{tr}(G_n)/n = 0$ and the MLE's $\hat{\lambda}_n$ and $\hat{\sigma}_n^2$ are asymptotically independent.

4. MIXED REGRESSIVE, SPATIAL AUTOREGRESSIVE MODELS: MULTICOLLINEARITY OF $G_n X_n \beta_0$ AND X_n

The set of the vectors $G_n X_n \beta_0$ and X_n can be linearly dependent under some circumstances. If $\beta_0 = 0$, $G_n X_n \beta_0 = 0$ and, hence, the set of $G_n X_n \beta_0$ and X_n is linearly dependent. This case corresponds to the pure spatial autoregressive process in (2.3). Another case is when W_n is row-normalized and the relevant regressor is only a constant term. Let $X_n = (l_n, X_{2n})$ and, conformably, $\beta_0 = (\beta_{01}, \beta_{02}')$, where $\beta_{02} = 0$. Consequently, as $X_n \beta_0 = l_n \beta_{01}$, $G_n X_n \beta_0 = (\beta_{01}/(1-\lambda_0)) l_n$ because $W_n l_n = l_n$ implies that $S_n l_n = (1-\lambda_0) l_n$ and $G_n l_n = l_n/(1-\lambda_0)$. The multicollinearity of $G_n X_n \beta_0$ and X_n is equivalent to the columns of $G_n X_n \beta_0$ lying in the space spanned by the columns of X_n, i.e., $M_n G_n X_n \beta_0 = 0$. It is also possible that even though $G_n X_n \beta_0$ and X_n are linear independent for finite n, they become asympototically multicollinear as n goes to infinity. This may happen for the spatial scenario in Case (1991) where the regressor vector x has a common mean across all districts with large group interactions. Quantitatively, this corresponds to $\lim_{n\to\infty} (1/n)(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) = 0$.[13] For subsequent analyses, Assumption 8 will be replaced by the following:

ASSUMPTION 8′: $\lim_{n\to\infty} (G_n X_n \beta_0)' M_n (G_n X_n \beta_0)/n = 0$.

Denote

$$\sigma_n^2(\lambda) = \frac{\sigma_0^2}{n} \operatorname{tr}[S_n'^{-1} S_n'(\lambda) S_n(\lambda) S_n^{-1}]. \tag{4.1}$$

Under the situation of Assumption 8′, $\lim_{n\to\infty} \sigma_n^{*2}(\lambda) = \lim_{n\to\infty} \sigma_n^2(\lambda)$ and $Q_n(\lambda)$ in (3.3) can be approximated by $Q_{a,n}(\lambda) = -(n/2)(\ln(2\pi)+1) - (n/2)\ln \sigma_n^2(\lambda) + \ln|S_n(\lambda)|$, which does not involve X_n. The identification condition of λ_0 can be stated in terms of the concentrated log-likelihood function of λ when $\{h_n\}$ is bounded.

[13]From the partition matrix formula, the $\lim_{n\to\infty} (1/n)(X_n, G_n X_n \beta_0)'(X_n, G_n X_n \beta_0)$ is nonsingular if and only if $\lim_{n\to\infty} (1/n) X_n' X_n$ and $\lim_{n\to\infty} (1/n)(G_n X_n \beta_0)' M_n (G_n X_n \beta_0)$ are nonsingular.

ASSUMPTION 9: *The $\{h_n\}$ is a bounded sequence and, for any $\lambda \neq \lambda_0$,*

$$\lim_{n\to\infty}\left(\frac{1}{n}\ln|\sigma_0^2 S_n^{-1}S_n'^{-1}| - \frac{1}{n}\ln|\sigma_n^2(\lambda)S_n^{-1}(\lambda)S_n'^{-1}(\lambda)|\right) \neq 0. \tag{4.2}$$

For the SAR model, as $Y_n = S_n^{-1}X_n\beta_0 + S_n^{-1}V_n$, the variance matrix of Y_n is $\sigma_0^2 S_n^{-1}S_n'^{-1}$. Assumption 9 is a global identification condition related to the uniqueness of the variance matrix of Y_n.

THEOREM 4.1: *For the situation of Assumption* 8′, *the QMLE $\hat{\theta}_n$ is a consistent estimator of θ_0 under Assumptions* 1–7 *and* 9.

For the situation of Assumption 8′, Σ_θ can be nonsingular if

$$\lim_{1\to\infty}\frac{1}{n}\operatorname{tr}(C_n^s C_n^s) \neq 0, \tag{4.3}$$

where $C_n = G_n - (\operatorname{tr}(G_n)/n)I_n$ and $C_n^s = C_n' + C_n$. This property is implied by Assumption 9. We note that $\operatorname{tr}(C_n^s C_n^s) = 2[\operatorname{tr}(G_nG_n') + \operatorname{tr}(G_n^2) - (2/n)\operatorname{tr}^2(G_n)]$, which is the square of the Euclidean norm of C_n^s, so in general $(1/n)\operatorname{tr}(C_n^s C_n^s) > 0$. The global identification condition in Assumption 9 guarantees that the limit in (4.3) does not vanish. As it shall be noted later, Assumption 9 and (4.3) can be valid only under the scenario that $\{h_n\}$ is a bounded sequence because $\operatorname{tr}(C_n^s C_n^s) = O(n/h_n)$. The asymptotic distribution of the QMLE $\hat{\theta}_n$ is $\sqrt{n}$-consistent and asymptotically normal when $\{h_n\}$ is a bounded sequence.

THEOREM 4.2: *Under Assumptions* 1–7, 8′, *and* 9, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \Sigma_\theta^{-1} + \Sigma_\theta^{-1}\Omega_\theta\Sigma_\theta^{-1})$. *Furthermore, if the v_i's are normally distributed, then* $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \Sigma_\theta^{-1})$.

When $\{h_n\}$ is a bounded sequence, all the QMLE's of λ_0, β_0 and σ_0^2 have the usual $\sqrt{n}$-rate of convergence from Theorem 4.2. This includes the QMLE for the pure SAR process in (2.3). For the pure SAR process, its concentrated log-likelihood function of λ is similar to that in (2.7) with X_n being zero in (2.6). These conclusions will subsequently be changed when $\{h_n\}$ is a divergence sequence.

5. MIXED REGRESSIVE MODELS WITH SINGULAR INFORMATION MATRICES

When $\lim_{n\to\infty} h_n = \infty$,

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

can be nonsingular only if $G_nX_n\beta_0$ and X_n are not asymptotically multicollinear as in Assumption 8. For the situation under Assumption 8′, when $\lim_{n\to\infty} h_n = \infty$, Σ_θ will become singular because

$$(1/n)\operatorname{tr}[(C_n' + C_n)(C_n' + C_n)'] = O(1/h_n) = o(1).$$

For the pure SAR process with $\theta = (\lambda, \sigma^2)$, as $\lim_{n\to\infty} h_n = \infty$,

$$\Sigma_\theta = \begin{pmatrix} 0 & 0 \\ 0 & 1/(2\sigma_0^4) \end{pmatrix}.$$

There are other cases in which the irregularity occurs. If W_n is row-normalized and $X_n = l_n$, $W_n X_n = l_n$ and $G_n X_n = l_n/(1-\lambda_0)$. In this case, when $\lim_{n\to\infty} h_n = \infty$, Σ_θ is singular because $(1/n)\operatorname{tr}(G_n)$ and $(1/n)[\operatorname{tr}(G_n' G_n) + \operatorname{tr}(G_n^2)]$ are $O(1/h_n)$, which goes to zero, and the submatrix

$$\frac{1}{n}(X_n,\ G_n X_n \beta_0)'(X_n,\ G_n X_n \beta_0) = \begin{pmatrix} 1 & \frac{\beta_0}{(1-\lambda_0)} \\ \frac{\beta_0}{(1-\lambda_0)} & (\frac{\beta_0}{(1-\lambda_0)})^2 \end{pmatrix}$$

is singular. When all spatially varying regressors X_{2n} in $X_n = (l_n, X_{2n})$ are irrelevant but are included in estimation, the coefficient β_{02} of X_{2n} in $\beta_0 = (\beta_{01}, \beta_{02}')'$ is zero. Consequently, $X_n\beta_0 = l_n\beta_{01}$ and $G_n X_n \beta_0 = (\beta_{01}/(1-\lambda_0))l_n$, when W_n is row-normalized. It follows that

$$\begin{pmatrix} \frac{1}{n}X_n'X_n & \frac{1}{n}X_n'(G_nX_n\beta_0) \\ \frac{1}{n}(G_nX_n\beta_0)'X_n & \frac{1}{n}(G_nX_n\beta_0)'(G_nX_n\beta_0) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}X_n'X_n & \frac{\beta_{01}}{1-\lambda_0}\cdot\frac{1}{n}X_n'l_n \\ \frac{\beta_{01}}{1-\lambda_0}\cdot\frac{1}{n}l_n'X_n & (\frac{\beta_{01}}{1-\lambda_0})^2\cdot\frac{1}{n}l_n'l_n \end{pmatrix}$$

is singular because the last column is proportional to the first one. The irregularity also occurs under Case's spatial scenario when x has a common mean across all districts (see footnote 15).

The singularity of the information matrix has implications on the rate of convergence of the estimators. When $\lim_{n\to\infty} h_n = \infty$, $(1/n)\ln L_n(\theta)$ is rather flat in λ and the convergence of $(1/n)(\ln L_n(\lambda) - Q_n(\lambda))$ to zero is too fast to be useful. However, with a properly adjusted rate,

$$\begin{aligned}&(h_n/n)\big[\big(\ln L_n(\lambda) - \ln L_n(\lambda_0)\big) - \big(Q_n(\lambda) - Q_n(\lambda_0)\big)\big]\\ &\quad \xrightarrow{p} 0 \quad \text{uniformly in } \lambda \text{ in } \Lambda,\end{aligned}$$

which shall be the useful one. We consider the situation in which

$$\lim_{n\to\infty} (h_n/n)(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) = c, \quad \text{where} \quad 0 \le c < \infty.$$

In this situation, it is natural that elements of $M_nG_nX_n\beta_0$ are of uniform order $O(1/\sqrt{h_n})$. In the event that $c = 0$, Assumption 9 shall be modified with a proper normalization.

ASSUMPTION 10: *The* $\{h_n\}$ *is a divergent sequence, elements of* $M_n(G_nX_n\beta_0)$ *have the uniform order* $O(1/\sqrt{h_n})$, *and* $\lim_{n\to\infty}(h_n/n)(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) = c$ *with* $0 \le c < \infty$. *Under this situation, either* (a) $c > 0$, *or* (b) *if* $c = 0$,

$$\lim_{n\to\infty}\left(\frac{h_n}{n}\ln|\sigma_0^2 S_n^{-1}S_n'^{-1}| - \frac{h_n}{n}\ln|\sigma_n^2(\lambda)S_n^{-1}(\lambda)S_n'^{-1}(\lambda)|\right) \neq 0,$$

whenever $\lambda \neq \lambda_0$.

Assumption 10(b) modifies Assumption 9 with the factor h_n to account for the proper rate of convergence.

THEOREM 5.1: *For the situation of Assumption* 10, *the QMLE* $\hat{\lambda}_n$ *derived from the maximization of* $\ln L_n(\lambda)$ *in* (2.7) *is a consistent estimator, under Assumptions* 1–7.

Asymptotic distribution of the QMLE $\hat{\lambda}_n$ can be derived from the concentrated log-likelihood function. Once the asymptotic distribution of $\hat{\lambda}_n$ is available, those of the QMLE's $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ from (2.5) and (2.6) can be derived. The limiting distribution of $\partial \ln L_n(\lambda_0)/\partial\lambda$ depends on the quadratic form of V_n. The original central limit theorem in Kelejian and Prucha (2001) is not directly applicable to the case with $\{h_n\}$ being a divergent sequence. But their theorem and its proof can be generalized to cover the divergent case (see Appendix A). Assumption 3 needs to be slightly strengthened.

ASSUMPTION 3′: $h_n^{1+\eta}/n \to 0$ *for some* $\eta > 0$ *as n goes to infinity.*

The central limit theorem for a linear-quadratic form implies that $(\sqrt{h_n/n}) \times \partial \ln L_n(\lambda_0)/\partial\lambda$ is asymptotically normal. The asymptotic distribution of $\hat{\lambda}_n$ follows from

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = -\left(\frac{h_n}{n}\frac{\partial^2 \ln L_n(\tilde{\lambda}_n)}{\partial\lambda^2}\right)^{-1}\sqrt{\frac{h_n}{n}}\frac{\partial \ln L_n(\lambda_0)}{\partial\lambda}.$$

Assumption 10(b) implies the local identification condition that $\lim_{n\to\infty}(h_n/n) \times \operatorname{tr}(C_n^s C_n^s) \neq 0$. Let $\operatorname{vec}_D(A)$ be the vector formed by the diagonal elements of a square matrix A.

THEOREM 5.2: *Under Assumptions* 1, 2, 3′, 4–7, *and* 10,

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \xrightarrow{D} N(0, \sigma_\lambda^2),$$

where

$$\begin{aligned}\sigma_\lambda^2 = \lim_{n\to\infty}&\left\{\frac{h_n}{n}\left[\frac{1}{\sigma_0^2}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \operatorname{tr}(C_nC_n^s)\right]\right\}^{-2}\\ &\times \frac{h_n}{n}\left[\frac{1}{\sigma_0^2}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \operatorname{tr}(C_nC_n^s)\right.\\ &\left.+ 2\frac{\mu_3}{\sigma_0^4}(G_nX_n\beta_0)'M_n\operatorname{vec}_D(C_n'M_n)\right].\end{aligned}$$

In the special case with $c = 0$ *in Assumption* 10, $\sigma_\lambda^2 = \lim_{n\to\infty}\{(h_n/n)\operatorname{tr}(C_nC_n^s)\}^{-1}$.

The possible slower rate of convergence of $\hat{\lambda}_n$ in Theorem 5.2 implies that, for statistical inference, one shall take into account the factor h_n in addition to the sample size n. Some practical formulas for classical inference statistics can be valid. In general, the t statistic for testing λ as a specific constant, say λ_c, is asymptotically valid when the proper asymptotic standard deviation of $\hat{\lambda}_n$ is used. Suppose that the disturbances are normally distributed. Let

$$\hat{\omega}_{\lambda,n}^2 = -\left(\frac{\partial^2 \ln L_n(\hat{\lambda}_n)}{\partial\lambda^2}\right)^{-1}.$$

This $(n/h_n)\hat{\omega}^2_{\lambda,n}$ is a consistent estimate of σ^2_λ. The conventional test statistic for testing $H_0:\lambda_0=\lambda_c$ is $(\hat{\lambda}_n-\lambda_c)/\hat{\omega}_{\lambda,n}$. This statistic is asymptotically standard normal, because

$$\frac{\hat{\lambda}_n-\lambda_0}{\hat{\omega}_{\lambda,n}}=\left(-\frac{h_n}{n}\frac{\partial^2\ln L_n(\hat{\lambda}_n)}{\partial\lambda^2}\right)^{-1/2}\sqrt{\frac{h_n}{n}}\frac{\partial\ln L_n(\lambda_0)}{\partial\lambda}+o_P(1)\xrightarrow{D}N(0,1)$$

under the null hypothesis. In addition to the Wald-type statistic, the conventional likelihood ratio and efficient score test statistics are also valid for testing $\lambda_0=\lambda_c$ under normal disturbances. This is so because, under the null hypothesis,

$$\begin{aligned}&2[\ln L_n(\hat{\lambda}_n)-\ln L_n(\lambda_c)]\\&=-\frac{\partial^2\ln L_n(\bar{\lambda}_n)}{\partial\lambda^2}(\hat{\lambda}_n-\lambda_0)^2\\&=\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n-\lambda_0)\Sigma^{-1}_{\lambda\lambda}\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n-\lambda_0)+o_P(1)\xrightarrow{D}\chi^2(1).\end{aligned}$$

The efficient score statistic

$$\frac{\partial\ln L_n(\lambda_c)}{\partial\lambda}\left(-\frac{\partial^2\ln L_n(\lambda_c)}{\partial\lambda^2}\right)^{-1}\frac{\partial\ln L_n(\lambda_c)}{\partial\lambda}$$

is asymptotically chi-square distributed because $-(h_n/n)\,\partial^2\ln L_n(\lambda_c)/\partial\lambda^2$ is a consistent estimate of the limiting variance of $(\sqrt{h_n/n}\,)\,\partial\ln L_n(\lambda_c)/\partial\lambda$ under the null hypothesis. From our results, we note that, even when the $\{v_i\}$ are not normally distributed, these classical statistics based on the concentrated likelihood can be asymptotically valid as long as $\lim_{n\to\infty}h_n=\infty$ and $\mu_3=0$.

With $\hat{\lambda}_n$, the QMLE's of β_0 and σ_0^2 are $\hat{\beta}_n=(X_n'X_n)^{-1}X_n'S_n(\hat{\lambda}_n)Y_n$ and $\hat{\sigma}_n^2=\frac{1}{n}Y_n'S_n'(\hat{\lambda}_n)M_nS_n(\hat{\lambda}_n)Y_n$.

THEOREM 5.3: *Under Assumptions* 1, 2, 3′, 4–7, *and* 10,

$$\begin{aligned}(5.1)\qquad&\sqrt{\frac{n}{h_n}}(\hat{\beta}_n-\beta_0)\\&=\sqrt{\frac{n}{h_n}}(X_n'X_n)^{-1}X_n'V_n\\&\quad-\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n-\lambda_0)\cdot(X_n'X_n)^{-1}X_n'G_nX_n\beta_0+O_p\left(\frac{1}{\sqrt{n}}\right)\\&\xrightarrow{D}N\Big(0,\sigma^2_\lambda\lim_{n\to\infty}(X_n'X_n)^{-1}X_n'(G_nX_n\beta_0)(G_nX_n\beta_0)'X_n(X_n'X_n)^{-1}\Big)\end{aligned}$$

and

$$\sqrt{n}(\hat{\sigma}_n^2-\sigma_0^2)=\frac{1}{\sqrt{n}}\sum_{i=1}^n(v_i^2-\sigma_0^2)+o_P(1)\xrightarrow{D}N(0,\mu_4-\sigma_0^4).$$

However, when $\beta_0 = 0$,

$$\sqrt{n}\hat{\beta}_n \xrightarrow{D} N\left(0, \sigma_0^2 \lim_{n\to\infty}\left(\frac{X_n'X_n}{n}\right)^{-1}\right).$$

The asymptotic distribution of $\hat{\lambda}_n$ has the $\sqrt{n/h_n}$-rate of convergence in Theorem 5.2. As h_n is divergent, this rate of convergence is lower than $\sqrt{n}$. For the Case spatial scenario, this corresponds to $\sqrt{R}$, where R is the number of districts in the sample. The asymptotic distribution of the QMLE $\hat{\beta}_n$ and its low rate of convergence in Theorem 5.3 are determined by the asymptotic distribution of $\hat{\lambda}_n$ that forms the leading term in the asymptotic expansion (5.1). When $\beta_0 = 0$, this leading term vanishes and $\hat{\beta}_n$ converges to β_0 with the usual $\sqrt{n}$-rate. The asymptotic distribution of $\hat{\sigma}_n^2$ has the usual $\sqrt{n}$-rate of convergence.

The rate of convergence of $\hat{\beta}_n$ can be improved in the event that $(1/n)X_n'G_nX_n\beta_0$ may vanish asymptotically. However, the exact rate of convergence will depend on how fast $(1/n)X_n'G_nX_n\beta_0$ will vanish in the limit. When $G_nX_n\beta_0$ and X_n are multicollinear for finite n, the implications of Theorem 5.3 on the various components of $\hat{\beta}_n$ can be spelled out more explicitly. Suppose there exists a column vector c_n such that $G_nX_n\beta_0 = X_nc_n$; then the asymptotic distribution of $\hat{\beta}_n$ in (5.1) can be rewritten as $\sqrt{(n/h_n)}(\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, \sigma_\lambda^2 \lim_{n\to\infty} c_nc_n')$. If some components of c_n are zero, the corresponding limiting variances will be zero. These components of $\hat{\beta}_n$ will have degenerate distributions and may converge at a rate faster than $\sqrt{n/h_n}$, while the estimates of the remaining components will converge at the $\sqrt{n/h_n}$-rate. From (5.1),

$$\hat{\beta}_n - \beta_0 = (X_n'X_n)^{-1}X_n'V_n - (\hat{\lambda}_n - \lambda_0)c_n + O_P\big(\sqrt{h_n}/n\big).$$

If $c_{1n} \neq 0$ but $c_{2n} = 0$, where $c_n = (c_{1n}', c_{2n}')'$, $\hat{\beta}_{n1}$ may be affected by the limiting distribution of $\hat{\lambda}_n$ but $\hat{\beta}_{n2}$ will not. This is because the dominated term for $\hat{\beta}_{n1}$ is $(\hat{\lambda}_n - \lambda_0)c_{1n}$. For $\hat{\beta}_{n2}$, as the corresponding component $(\hat{\lambda}_n - \lambda_0)c_{2n}$ vanishes, $\hat{\beta}_{n2}$ has the usual $\sqrt{n}$-rate of convergence regardless if $\{h_n\}$ is divergent or not.

THEOREM 5.4: *Under Assumptions* 1, 2, 3′, 4–7, *and* 10(b), *and* $G_nX_n\beta_0 = X_{1n}c_{1n}$ *for some* c_{1n}, *where* $X_n = (X_{1n}, X_{2n})$,

$$\sqrt{\frac{n}{h_n}}(\hat{\beta}_{n1} - \beta_{01}) \xrightarrow{D} N(0, \sigma_\lambda^2 c_1c_1'), \quad \textit{where} \quad c_1 = \lim_{n\to\infty} c_{1n},$$

but

$$\sqrt{n}(\hat{\beta}_{n2} - \beta_{02}) \xrightarrow{D} N\left(0, \sigma_0^2\left[\lim_{n\to\infty}\frac{1}{n}X_{2n}'\big(I_n - X_{1n}(X_{1n}'X_{1n})^{-1}X_{1n}'\big)X_{2n}\right]^{-1}\right).$$

In summary, consider the SAR model where all the included spatial varying regressors are irrelevant, i.e., $X_n = (l_n, X_{2n})$ and $\beta = (\beta_1, \beta_2')'$ with $\beta_{02} = 0$. Because $\beta_{02} = 0$ is an unknown event, one estimates both β_1 and β_2. Because $G_nX_n\beta_0 = \beta_{01}G_nl_n$, $G_nX_n\beta_0$ and l_n can be distinguished regressors if G_nl_n is not linearly dependent on l_n. In that case, Theorems 3.1 and 3.2 are applicable and the QMLE $\hat{\theta}_n$ can be $\sqrt{n}$-consistent. In the event that G_nl_n and l_n are multicollinear but $\{h_n\}$ is a bounded sequence, Theorems 4.1 and 4.2 are applicable and $\hat{\theta}_n$ is still $\sqrt{n}$-consistent. The irregular case occurs when $\lim_{n\to\infty} h_n = \infty$ and $G_n l_n$ and l_n are multicollinear. If β_{01} were zero, it would correspond to $\beta_0 = 0$ covered by the last part of Theorem 5.3. For the model with $\beta_{01} \neq 0$ but $\beta_{02} = 0$ and the weights matrix row-normalized, as $G_n X_n \beta_0 = (\beta_{01}/(1-\lambda_0)) l_n$, $c_{1n} = \beta_{01}/(1-\lambda_0) \neq 0$ and $c_{2n} = 0$. For this case, Theorem 5.4 implies that, when $\lim_{n\to\infty} h_n = \infty$, $\hat{\beta}_{n1}$ has the same low rate of convergence as that of $\hat{\lambda}_n$, but $\hat{\beta}_{n2}$ will converge to zero in probability at the usual $\sqrt{n}$-rate.

When the constraint $\beta_{02} = 0$ is correctly imposed, the model for estimation becomes a spatial autoregressive model with an unknown intercept: $Y_n = \beta_1 l_n + \lambda W_n Y_n + V_n$. The unknown parameters are β_1, λ, and σ^2. Given a λ, the QMLE's of β_1 and σ^2 are, respectively, $\hat{\beta}_{n1}(\lambda) = (1/n) l_n' S_n(\lambda) Y_n$ and $\hat{\sigma}_n^2(\lambda) = (1/n) Y_n' S'(\lambda) M_{1n} S_n(\lambda) Y_n$, where $M_{1n} = I_n - l_n l_n'/n$. The concentrated log-likelihood function of λ is in (2.7) with M_n replaced by M_{1n}. Because M_{1n} is a special case for M_n (with $X_n = l_n$), Theorems 5.1–5.4 hold also for the restricted parameter estimates $\hat{\lambda}_n$, $\hat{\beta}_{n1}$, and $\hat{\sigma}_n^2$. For the pure SAR process (2.3), the estimation corresponds to imposing $\beta_0 = (\beta_{01}, \beta_{02}) = 0$. The concentrated log-likelihood of λ corresponds to the one in (2.7) with $M_n = I_n$. Theorems 5.1 and 5.2 hold also for the SAR process.

6. INCONSISTENCY WHEN $\lim_{n\to\infty}(h_n/n) > 0$

The preceding results are derived with $\lim_{n\to\infty}(h_n/n) = 0$ under Assumption 3. That is, either $\{h_n\}$ is a bounded sequence or $\{h_n\}$ diverges to infinity at a rate slower than n. In this section, we provide an example in which the QMLE $\hat{\theta}_n$ may not be consistent if h_n has the rate n.

Consider $W_n = (1/(n-1))(l_n l_n' - I_n)$ in Case (1991) when sample data are collected only from a single district. In this case, $h_n = (n-1)$ is $O(n)$. For simplicity, consider $\sigma_0^2 = 1$ as known. With this W_n,

$$S_n^{-1}(\lambda) = \left(1 + \frac{\lambda}{n-1}\right)^{-1}\left(I_n + \frac{\lambda}{1-\lambda}\frac{l_n l_n'}{n-1}\right) \quad \text{and}$$

$$W_n S_n^{-1}(\lambda) = \frac{1}{n-1+\lambda}\left(\frac{l_n l_n'}{1-\lambda} - I_n\right).$$

As X_n includes an intercept term,

$$G_n X_n \beta_0 = \frac{n}{n-1+\lambda_0}\left(l_n' X_n \beta_0 \frac{l_n}{(1-\lambda_0)n} - \frac{X_n \beta_0}{n}\right)$$

is multicollinear with X_n.

The log-likelihood function is

$$\ln L_n(\delta) = -\frac{n}{2}\ln(2\pi) + \ln|S_n(\lambda)| - \frac{1}{2}V_n'(\delta)V_n(\delta),$$

where $\delta = (\beta', \lambda)'$. Given λ, the QMLE of β_0 is $\hat{\beta}_n(\lambda) = (X_n' X_n)^{-1} X_n' S_n(\lambda) Y_n$ and the concentrated log-likelihood function of λ is

$$\text{(6.1)} \qquad \ln L_n(\lambda) = -\frac{n}{2}\ln(2\pi) + \ln|S_n(\lambda)| - \frac{1}{2}Y_n' S_n'(\lambda) M_n S_n(\lambda) Y_n.$$

Because $M_nG_nX_n\beta_0 = 0$,

$$\frac{\partial \ln L_n(\lambda)}{\partial \lambda} = -\operatorname{tr}(W_nS_n^{-1}(\lambda)) + V_n'M_nG_nV_n + V_n'G_n'M_nG_nV_n(\lambda_0 - \lambda).$$

Because $\operatorname{tr}(G_n) = (n/(n-1+\lambda_0))\lambda_0/(1-\lambda_0)$ and $M_nG_n = -M_n/(n-1+\lambda_0)$,

$$\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} = -\frac{n}{n-1+\lambda_0}\frac{\lambda_0}{1-\lambda_0} - \frac{V_n'M_nV_n}{n-1+\lambda_0}.$$

The second-order derivative of (6.1) is

$$\begin{aligned}\frac{\partial^2 \ln L_n(\lambda)}{\partial \lambda^2} &= -\operatorname{tr}\bigl[(W_nS_n^{-1}(\lambda))^2\bigr] - V_n'G_n'M_nG_nV_n \\ &= -\frac{n^2}{(n-1+\lambda)^2}\left[\frac{1-2(1-\lambda)/n}{(1-\lambda)^2} + \frac{1}{n}\right] - \frac{V_n'M_nV_n}{(n-1+\lambda_0)^2}.\end{aligned}$$

By the mean value theorem,

$$\hat{\lambda}_n = \lambda_0 - \left(\frac{\partial^2 \ln L_n(\bar{\lambda}_n)}{\partial \lambda^2}\right)^{-1}\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda},$$

where $\bar{\lambda}_n$ lies between $\hat{\lambda}_n$ and λ_0. Suppose $\hat{\lambda}_n$ were consistent, we shall show that there would be a contradiction. If $\hat{\lambda}_n$ were consistent, it would imply that $\bar{\lambda}_n \xrightarrow{p} \lambda_0$ and, hence, $\partial^2 \ln L_n(\bar{\lambda}_n)/\partial\lambda^2 \xrightarrow{p} -1/(1-\lambda_0)^2$. As $(1/n)V_n'M_nV_n = (1/n)V_n'V_n + o_P(1) \xrightarrow{p} 1$, $\partial \ln L_n(\lambda_0)/\partial\lambda \xrightarrow{p} 1 - \lambda_0/(1-\lambda_0)$. Consequently, $\hat{\lambda}_n \xrightarrow{p} \lambda_0 + (1-\lambda_0)(1-2\lambda_0) \neq \lambda_0$ in general, a contradiction.

For the pure SAR process, it corresponds to $\beta_0 = 0$ imposed in estimation. As

$$\begin{aligned}&\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} \\ &\quad = -\frac{\lambda_0}{1-\lambda_0}\left(1 - \frac{1-\lambda_0}{n}\right)^{-1} + \frac{1}{n-1+\lambda_0}V_n'\left(\frac{l_nl_n'}{1-\lambda_0} - I_n\right)V_n\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{n}V_n'\left(\frac{l_nl_n'}{1-\lambda_0} - I_n\right)V_n - \frac{\lambda_0}{1-\lambda_0} \\ &\quad = \frac{1}{1-\lambda_0}\left[\left(\frac{\sum_{i=1}^n v_i}{\sqrt{n}}\right)^2 - 1\right] + \left(1 - \frac{\sum_{i=1}^n v_i^2}{n}\right) \xrightarrow{p} \frac{\xi - 1}{1-\lambda_0},\end{aligned}$$

where ξ is a $\chi^2(1)$ variable, $\partial \ln L_n(\lambda_0)/\partial\lambda \xrightarrow{D} (\xi-1)/(1-\lambda_0)$. The second-order derivative is

$$\begin{aligned}\frac{\partial^2 \ln L_n(\lambda)}{\partial \lambda^2} &= -\left(\frac{n}{n-1+\lambda}\right)^2\left(\frac{1-2(1-\lambda)/n}{(1-\lambda)^2} + \frac{1}{n}\right) \\ &\qquad - \frac{1}{(n-1+\lambda)^2}V_n'\left(\frac{n-2(1-\lambda)}{(1-\lambda)^2}l_nl_n' + I_n\right)V_n.\end{aligned}$$

By the mean value theorem,

$$\hat{\lambda}_n = \lambda_0 - \left(\frac{\partial^2 \ln L_n(\bar{\lambda}_n)}{\partial \lambda^2}\right)^{-1} \frac{\partial \ln L_n(\lambda_0)}{\partial \lambda},$$

where $\bar{\lambda}_n$ lies between $\hat{\lambda}_n$ and λ_0. If $\hat{\lambda}_n$ were a consistent estimator, $\bar{\lambda}_n \xrightarrow{p} \lambda_0$ and

$$\frac{\partial^2 \ln L_n(\bar{\lambda}_n)}{\partial \lambda^2} \xrightarrow{D} -\frac{\xi+1}{(1-\lambda_0)^2}.$$

Thus, if $\hat{\lambda}_n$ were a consistent estimator, it would imply $\hat{\lambda}_n - \lambda_0 \xrightarrow{D} (1-\lambda_0)(\xi-1)/(\xi+1)$. This would be a contradiction as $(1-\lambda_0)(\xi-1)/(\xi+1)$ would not have a degenerate distribution (at zero). So $\hat{\lambda}_n$ could not be a consistent estimator of λ_0.

7. MONTE CARLO RESULTS

To investigate finite sample properties of the QMLE by a Monte Carlo study, we focus on the spatial scenario in Case (1991) with an R number of districts, m members in each district, and with each neighbor of a member in a district given equal weight, i.e., $W_n = I_R \otimes B_m$, where $B_m = (1/(m-1))(l_m l_m' - I_m)$ as in Section 2. We consider models with and without regressors.[14]

The first model (SAR) in the study is a spatial process $Y_n = \lambda W_n Y_n + V_n$, where $V_n \sim N(0, \sigma^2 I_n)$. The sample data are generated with $\lambda = .5$ and $\sigma^2 = 1$. The second model (MRSAR-1) extends the first model to $Y_n = \lambda W_n Y_n + X_n \beta + V_n$ by including a regressor, where $X_n \sim N(0, I_n)$ and $\beta = 1$. The regressors are i.i.d. across districts as well as members in a district. The third model (MRSAR-2) specifies a regressor where its values for members in a single district can be correlated. Let z_r, $r = 1, \ldots, R$, be generated by $N(0,1)$. The regressor x_{ir} of the ith member in the district r is generated as $x_{ir} = (z_r + z_{ir})/\sqrt{2}$, where z_{ir} are i.i.d. $N(0,1)$ for all i and r and are independent of z_r. This specification implies that the average value of x_{ir} of the district r will converge in probability to z_r as m goes to infinity in MRSAR-2. On the other hand, the average value for each district in MRSAR-1 will go to zero, which is their mean by design.[15]

We have experimented with different values of R from 30 to 120 and m from 3 to 100. For each case, there are 400 repetitions.[16] The optimization is performed with the Brent

[14]Monte Carlo studies for the MLE under spatial scenarios for which each unit has a few neighbors can be found in Anselin (1988).

[15]If the mean μ_r of x_{ir} conditional on a district is the same across districts, i.e., $\mu_r = \mu$ for all r, then, when either $\mu = 0$ or X_n includes an intercept term,

$$M_n G_n X_n = M_n\{(m/(1-\lambda_0)(m-1+\lambda_0))((\bar{x}_{.1}-\mu)' l_m' \cdots (\bar{x}_{.R}-\mu)' l_m')' - X_n/(m-1+\lambda_0)\}$$

and its elements are $O(1/\sqrt{m})$, where $\bar{x}_{.r}$ is the mean of x in the rth district. This case corresponds to the situation in Assumption 10. If μ_r's are different across different districts,

$$M_n G_n X_n = M_n\{(m/(1-\lambda_0)(m-1+\lambda_0))(\bar{x}_{.1}' l_m' \cdots \bar{x}_{.R}' l_m')' - X_n/(m-1+\lambda_0)\}$$

and its elements will, in general, have $O(1)$.

[16]The regressor matrix is randomly generated in each Monte Carlo trial.

TABLE I

ML ESTIMATION OF SPATIAL AUTOREGRESSIVE MODELS

R	SAR	m = 3	5	10	20	50	100
30	λ	.3896	.4290	.4552	.4741	.4834	.4849
		(.0734)	(.0778)	(.0764)	(.0681)	(.0692)	(.0722)
	σ	.9533	.9852	.9975	1.0008	.9992	.9998
		(.0769)	(.0582)	(.0407)	(.0282)	(.0183)	(.0131)
60	λ	.3930	.4365	.4679	.4830	.4906	.4917
		(.0520)	(.0519)	(.0504)	(.0480)	(.0471)	(.0493)
	σ	.9586	.9879	.9985	.9986	.9997	1.0005
		(.0504)	(.0409)	(.0282)	(.0202)	(.0132)	(.0094)
120	λ	.3978	.4430	.4725	.4858	.4927	.4939
		(.0373)	(.0372)	(.0351)	(.0351)	(.0332)	(.0350)
	σ	.9613	.9886	.9964	.9989	1.0004	1.0002
		(.0362)	(.0280)	(.0203)	(.0148)	(.0095)	(.0067)
R	MRSAR-1						
30	λ	.3992	.4367	.4624	.4775	.4827	.4881
		(.0676)	(.0600)	(.0595)	(.0577)	(.0562)	(.0507)
	β	.9512	.9831	.9946	.9981	.9970	.9997
		(.1041)	(.0820)	(.0568)	(.0410)	(.0264)	(.0191)
	σ	.9403	.9792	.9950	.9998	1.0001	.9997
		(.0718)	(.0572)	(.0396)	(.0284)	(.0174)	(.0123)
60	λ	.3990	.4403	.4672	.4846	.4876	.4937
		(.0469)	(.0427)	(.0423)	(.0385)	(.0373)	(.0365)
	β	.9526	.9848	.9960	.9972	.9997	.9995
		(.0769)	(.0562)	(.0411)	(.0300)	(.0191)	(.0128)
	σ	.9520	.9852	.9978	.9994	.9996	.9997
		(.0513)	(.0391)	(.0283)	(.0198)	(.0123)	(.0090)
120	λ	.4000	.4421	.4718	.4854	.4907	.4949
		(.0320)	(.0303)	(.0290)	(.0264)	(.0265)	(.0265)
	β	.9573	.9861	.9950	.9989	.9995	.9997
		(.0527)	(.0412)	(.0300)	(.0221)	(.0127)	(.0089)
	σ	.9580	.9881	.9973	.9994	.9996	.9999
		(.0373)	(.0277)	(.0198)	(.0141)	(.0090)	(.0063)

Remarks: (1) SAR: $Y_n = \lambda W_n Y_n + V_n, V_n \sim N(0, \sigma^2 I_n)$; (2) MRSAR-1: $Y_n = \lambda W_n Y_n + X_n \beta + V_n$, where $V_n \sim N(0, \sigma^2 I_n)$ and $X_n \sim N(0, I_n)$; (3) the R is the number of districts and m is the number of members in a district.

method in one-dimensional search with first derivatives (Press et al. (1992, Ch. 10)). The empirical mean and standard deviation (in bracket) for each parameter estimator are reported in Tables I and II. The effects of m on $\hat{\lambda}_n$ are of interest. There are biases in $\hat{\lambda}_n$ in all three models. The biases of $\hat{\lambda}_n$ decrease as m becomes larger. The biases of $\hat{\sigma}_n$ and $\hat{\beta}_n$ are rather small. The empirical standard errors of $\hat{\beta}_n$ and $\hat{\sigma}_n$ decrease as either R or m increases. For a fixed R, the empirical standard errors of $\hat{\lambda}_n$ do not change much as m becomes large for both the SAR process and the MRSAR-1 model.

TABLE II

ML ESTIMATION OF SPATIAL AUTOREGRESSIVE MODELS

		m					
R	MRSAR-2	3	5	10	20	50	100
30	λ	.3912	.4358	.4661	.4829	.4915	.4964
		(.0635)	(.0550)	(.0428)	(.0366)	(.0248)	(.0184)
	β	.9684	.9880	1.0006	1.0039	.9983	1.0023
		(.1151)	(.1026)	(.0727)	(.0526)	(.0362)	(.0248)
	σ	.9524	.9808	.9955	.9990	1.0009	1.0002
		(.0728)	(.0612)	(.0420)	(.0275)	(.0184)	(.0132)
60	λ	.3985	.4415	.4689	.4846	.4930	.4974
		(.0462)	(.0364)	(.0332)	(.0259)	(.0158)	(.0117)
	β	.9614	.9863	1.0023	1.0011	.9999	1.0008
		(.0852)	(.0696)	(.0527)	(.0388)	(.0258)	(.0166)
	σ	.9537	.9865	.9974	.9987	1.0005	.9995
		(.0513)	(.0431)	(.0288)	(.0193)	(.0136)	(.0090)
120	λ	.3986	.4424	.4717	.4860	.4940	.4973
		(.0324)	(.0253)	(.0227)	(.0178)	(.0113)	(.0091)
	β	.9625	.9871	.9994	.9995	.9991	1.0007
		(.0597)	(.0474)	(.0380)	(.0281)	(.0175)	(.0123)
	σ	.9580	.9878	.9973	1.0001	1.0001	.9997
		(.0381)	(.0297)	(.0204)	(.0143)	(.0091)	(.0062)

Remarks: MRSAR-2: $Y_n = \lambda W_n Y_n + X_n\beta + V_n$, where the elements x_{ir} of X_n are $x_{ir} = (z_r + z_{ir})/\sqrt{2}$. The z_{ir}'s and z_r's are i.i.d. $N(0,1)$.

They decrease as m increases for the MRSAR-2 model. This behavior of $\hat{\lambda}_n$ confirms the implication of our theoretical analysis as $\sqrt{n/h_n} = \sqrt{R}$ here.[17]

8. CONCLUSION

The examples of inconsistent QMLE have samples from a single district. By increasing n, it increases spatial units in the (same) district. That corresponds to the notion of 'infill asymptotics' (Cressie (1993, p. 101)). This example shows that the QMLE under infill asymptotics alone may not be consistent. If there are many separate districts from which samples are obtained, the QMLE's can be consistent if the number of districts R increases to infinity. The latter scenario corresponds to the notion of "increasing-domain asymptotics" (Cressie (1993, p. 100)). Consistency of the QMLE can be achieved with increasing-domain asymptotics. From our results, the QMLE under the increasing-domain asymptotics alone can have the usual $\sqrt{n}$-rate of conver-

[17]For the MRSAR-1 model, the standard error of $\hat{\beta}_n$ decreases with increasing m. This is a special result of Theorem 5.3. As x_{ri} are i.i.d. with zero mean, $X_n'G_nX_n = O_P(1/m)$ for the MRSAR-1 model. In this case,

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = (X_n'X_n/n)^{-1}(1/\sqrt{n})X_n'V_n + o_P(1)$$

from (5.1).

gence. But, when both infill and increasing-domain asymptotics are operating, the rates of convergence of the QMLE's for various parameters can be different and some may have slower rates than the usual $\sqrt{n}$ one.

Dept. of Economics, The Ohio State University, Columbus, OH 43210-1172, U.S.A.; lflee@econ.ohio-state.edu; http://economics.sbs.ohio-state.edu/lee/.

Manuscript received July, 2003; final revision received February, 2004.

APPENDIX A

Notations

The following list summarizes some frequently used notations in the text:
$S_n(\lambda) = I_n - \lambda W_n$ for any possible λ.
$S_n = I_n - \lambda_0 W_n$.
$G_n = W_n S_n^{-1}$.
$C_n = G_n - (\mathrm{tr}(G_n)/n) I_n$.
$\ln L_n(\theta)$ is the log-likelihood of $\theta = (\beta', \lambda, \sigma^2)'$.
$\ln L_n(\lambda)$ is the concentrated log-likelihood function of λ.
$Q_n(\lambda) = \max_{\beta,\sigma^2} E(\ln L_n(\theta))$.
$\sigma_n^2(\lambda) = (\sigma_0^2/n)\,\mathrm{tr}[S_n'^{-1} S_n'(\lambda) S_n(\lambda) S_n^{-1}]$.
$M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$.

Some Basic Properties

The following statements summarize some basic properties on spatial weight matrices and some laws of large numbers and central limit theorems on linear and quadratic forms. The elements, the v_i's, of $V_n = (v_1, \ldots, v_n)'$ are assumed to be i.i.d. with zero mean and a finite variance σ_0^2. For quadratic forms involving V_n, the fourth moment μ_4 of the v's is assumed to exist.

- Suppose that the spatial weights matrix W_n is a row-normalized matrix with its (i, j)th element being $w_{n,i} = d_{ij}/\sum_{l=1}^n d_{il}$ and $d_{ij} \geq 0$ for all i, j. If $d_{ij} = d_{ji}$ for all i and j and $\sum_{j=1}^n d_{ij}$ are $O(h_n)$ and are bounded away from zero at the rate h_n uniformly in i, then $\{W_n\}$ are uniformly bounded in column sums.
- Suppose that $\{\|W_n\|\}$ and $\{\|S_n^{-1}\|\}$, where $\|\cdot\|$ is a matrix norm, are bounded. Then $\{\|S_n(\lambda)^{-1}\|\}$ are uniformly bounded in a neighborhood of λ_0.
- Suppose that $\|W_n\| \leq 1$ for all n, where $\|\cdot\|$ is a matrix norm; then $\{\|S_n(\lambda)^{-1}\|\}$ are uniformly bounded in any closed subset of $(-1, 1)$.
- Suppose that elements of the $n \times k$ matrices X_n are uniformly bounded for all n, and $\lim_{n\to\infty} X_n'X_n/n$ exists and is nonsingular; then the projectors $X_n(X_n'X_n)^{-1}X_n'$ and $I_n - X_n(X_n'X_n)^{-1}X_n'$ are uniformly bounded in both row and column sums.
- Suppose that A_n is a square matrix with its column sums being uniformly bounded and elements of the $n \times k$ matrix Z_n are uniformly bounded. Then, $(1/\sqrt{n})Z_n'A_nV_n = O_P(1)$. Furthermore, if the limit of $Z_n'A_nA_n'Z_n/n$ exists and is positive definite, then $(1/\sqrt{n})Z_n'A_nV_n \xrightarrow{D} N(0, \sigma_0^2 \lim_{n\to\infty} Z_n'A_nA_n'Z_n/n)$.
- Let $A_n = [a_{ij}]$ be an n-dimensional square matrix. Then, $E(V_n'A_nV_n) = \sigma_0^2\,\mathrm{tr}(A_n)$ and $\mathrm{var}(V_n'A_nV_n) = (\mu_4 - 3\sigma_0^4)\sum_{i=1}^n a_{ii}^2 + \sigma_0^4[\mathrm{tr}(A_nA_n') + \mathrm{tr}(A_n^2)]$.
- Suppose the elements $a_{n,ij}$ of the $n \times n$ matrices A_n are $O(1/h_n)$ uniformly for all i, j. If $n \times n$ matrices $\{B_n\}$ are uniformly bounded in column sums (respectively, row sums), then the elements of A_nB_n (respectively, B_nA_n) have the uniform order $O(1/h_n)$. For these cases, $\mathrm{tr}(A_nB_n) = \mathrm{tr}(B_nA_n) = O(n/h_n)$.
- Suppose that $\{A_n\}$ are uniformly bounded either in row or column sums and their elements $a_{n,ij}$ have $O(1/h_n)$ uniformly in i and j. Then $E(V_n'A_nV_n) = O(n/h_n)$ and $\mathrm{var}(V_n'A_nV_n) = O(n/h_n)$. If $\lim_{n\to\infty}(h_n/n) = 0$, then $(h_n/n)[V_n'A_nV_n - E(V_n'A_nV_n)] = o_P(1)$.

- Suppose that $\{A_n\}$ is a sequence of symmetric matrices with row and column sums uniformly bounded in absolute value and $\{b_n\}$ is a sequence of constant vectors with its elements uniformly bounded. The moment $E(|v|^{4+2\delta})$ for some $\delta > 0$ of v exists. Let $\sigma_{Q_n}^2$ be the variance of Q_n, where $Q_n = b_n' + V_n' A_n V_n - \sigma^2 \operatorname{tr}(A_n)$. Assume that the variance $\sigma_{Q_n}^2$ is $O(n/h_n)$ with $\{(h_n/n)\sigma_{Q_n}^2\}$ bounded away from zero, the elements of A_n are of uniform order $O(1/h_n)$ and the elements of b_n are of uniform order $O(1/\sqrt{h_n})$. If $\lim_{n\to\infty}(h_n^{1+2/\delta}/n) = 0$, then $Q_n/\sigma_{Q_n} \xrightarrow{D} N(0,1)$.
- Suppose that A_n is a constant $n \times n$ matrix uniformly bounded in both row and column sums. Let c_n be a column vector of constants. If $(h_n/n)c_n' c_n = o(1)$, then $(\sqrt{h_n/n})c_n' A_n \mathcal{E}_n = o_P(1)$. On the other hand, $(\sqrt{h_n/n})c_n' A_n \mathcal{E}_n = O_P(1)$ if $(h_n/n)c_n' c_n = O(1)$.

APPENDIX B

PROOF OF THEOREM 3.1 AND THEOREM 4.1: The consistency of $\hat{\theta}_n$ will follow from the uniform convergence of $(1/n)(\ln L_n(\lambda) - Q_n(\lambda))$ to zero on Λ and the uniqueness identification condition that, for any $\epsilon > 0$, $\limsup_{n\to\infty} \max_{\lambda \in \bar{N}_\epsilon(\lambda_0)} (1/n)[Q_n(\lambda) - Q_n(\lambda_0)] < 0$, where $\bar{N}_\epsilon(\lambda_0)$ is the complement of an open neighborhood of λ_0 in Λ of diameter ϵ (White (1994, Theorem 3.4)).

Note that $(1/n)(\ln L_n(\lambda) - Q_n(\lambda)) = -(1/2)(\ln \hat{\sigma}_n^2(\lambda) - \ln \sigma_n^{*2}(\lambda))$. The $\sigma_n^{*2}(\lambda)$ and $\hat{\sigma}_n^2(\lambda)$ can be written as $\sigma_n^{*2}(\lambda) = (\lambda_0 - \lambda)^2 (G_n X_n \beta_0)' M_n (G_n X_n \beta_0)/n + \sigma_n^2(\lambda)$, where

$$\sigma_n^2(\lambda) = \frac{\sigma_0^2}{n} \operatorname{tr}\big(S_n'^{-1} S_n'(\lambda) S_n(\lambda) S_n^{-1}\big) \quad \text{and}$$

$$\begin{aligned}\hat{\sigma}_n^2(\lambda) &= \frac{1}{n} Y_n' S_n'(\lambda) M_n S_n(\lambda) Y_n \\ &= (\lambda_0 - \lambda)^2 \frac{1}{n} (G_n X_n \beta_0)' M_n (G_n X_n \beta_0) + 2(\lambda_0 - \lambda) H_{1n}(\lambda) + H_{2n}(\lambda),\end{aligned}$$

where $H_{1n}(\lambda) = (1/n)(G_n X_n \beta_0)' M_n S_n(\lambda) S_n^{-1} V_n$ and

$$H_{2n}(\lambda) = \frac{1}{n} V_n' S_n'^{-1} S_n'(\lambda) M_n S_n(\lambda) S_n^{-1} V_n.$$

It can be shown that $H_{1n}(\lambda) = o_P(1)$ and $H_{2n}(\lambda) - \sigma_n^2(\lambda) = o_P(1)$ uniformly on Λ. Therefore, $\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda) = o_P(1)$ uniformly on Λ. Consequently, $\sup_{\lambda \in \Lambda} |(1/n)(\ln L_n(\lambda) - Q_n(\lambda))| = o_P(1)$. The identification uniqueness condition can be established by a counter argument. First,

$$\frac{1}{n}[Q_n(\lambda) - Q_n(\lambda_0)] = \frac{1}{n}\big(Q_{p,n}(\lambda) - Q_{p,n}(\lambda_0)\big) - \frac{1}{2}[\ln \sigma_n^{*2}(\lambda) - \ln \sigma_n^2(\lambda)],$$

where

$$Q_{p,n}(\lambda) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2} \ln \sigma_n^2(\lambda) + \ln |S_n(\lambda)|.$$

The $Q_n(\lambda)/n$ is uniformly equicontinuous on Λ. By Jensen's inequality, $(1/n)(Q_{p,n}(\lambda) - Q_{p,n}(\lambda_0)) \le 0$ for all λ. Furthermore, $\sigma_n^{*2}(\lambda) \ge \sigma_n^2(\lambda)$. If the identification uniqueness condition were not satisfied, without loss of generality, there would exist a sequence $\lambda_n \in \Lambda$ that would converge to a point $\lambda_+ \ne \lambda_0$ such that $\lim_{n\to\infty}(1/n)[Q_n(\lambda_n) - Q_n(\lambda_0)] = 0$. This would be possible only if $\lim_{n\to\infty}(\sigma_n^{*2}(\lambda_+) - \sigma_n^2(\lambda_+)) = 0$ and $\lim_{n\to\infty}(1/n)[Q_{p,n}(\lambda_+) - Q_{p,n}(\lambda_0)] = 0$. The latter would generate a contradiction to either $\lim_{n\to\infty}(1/n)(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) \ne 0$ or Assumption 9. *Q.E.D.*

PROOF OF THEOREM 3.2: Except λ, β and $1/\sigma^2$ appear either linearly or in quadratic form in $\partial^2 \ln L_n(\theta)/(\partial\theta\,\partial\theta')$. The second derivative with λ is

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda^2} = -\operatorname{tr}\big([W_n S_n^{-1}(\lambda)]^2\big) - \frac{Y_n' W_n' W_n Y_n}{\sigma^2}.$$

Denote $G_n(\lambda) = W_n S_n(\lambda)$. By the mean value theorem, $\operatorname{tr}(G_n^2(\tilde\lambda_n)) = \operatorname{tr}(G_n^2) + 2\operatorname{tr}(G_n^3(\bar\lambda_n)) \times (\tilde\lambda_n - \lambda_0)$. Assumption 5 implies that $G_n(\bar\lambda_n)$ is uniformly bounded in row and column sums uniformly in a neighborhood of λ_0. Hence,

$$\begin{aligned}
&\frac{1}{n}\left[\frac{\partial^2 \ln L_n(\tilde\theta_n)}{\partial\lambda^2} - \frac{\partial^2 \ln L_n(\theta_0)}{\partial\lambda^2}\right]\\
&= -2\frac{\operatorname{tr}(G_n^3(\bar\lambda_n))}{n}(\tilde\lambda_n - \lambda_0) + \left[\frac{1}{\sigma_0^2} - \frac{1}{\tilde\sigma_n^2}\right]\frac{Y_n' W_n' W_n Y_n}{n}\\
&= o_p(1),
\end{aligned}$$

because $\operatorname{tr}(G_n^3(\bar\lambda_n)) = O(n/h_n)$ and $Y_n' W_n' W_n Y_n = O_P(n/h_n)$. As other terms of the second order derivatives can be easily analyzed,

$$\frac{1}{n}\left[\frac{\partial^2 \ln L_n(\tilde\theta_n)}{\partial\theta\,\partial\theta'} - \frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\,\partial\theta'}\right] \xrightarrow{p} 0.$$

The convergence of

$$\frac{1}{n}\left[\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\,\partial\theta'} - E\left(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\,\partial\theta'}\right)\right]$$

to zero in probability is straightforward by showing that linear functions and quadratic functions of V_n, deviated from their means, e.g., $X_n' G_n V_n/n$ and $(1/n)(V_n' G_n V_n - \sigma_0^2 \operatorname{tr}(G_n))$, are all $o_P(1)$.

The components of $(1/\sqrt{n})\,\partial \ln L_n(\theta_0)/\partial\theta$ are linear or quadratic functions of V_n. With the existence of high-order moments of v in Assumption 1, the central limit theorem for linear-quadratic forms of Kelejian and Prucha (2001) can be applied and

$$\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta_0)}{\partial\theta} \xrightarrow{D} N(0, \Sigma_\theta + \Omega_\theta).$$

Assumption 8 guarantees that Σ_θ is nonsingular. The asymptotic distribution of $\hat\theta_n$ follows from the expansion

$$\sqrt{n}(\hat\theta_n - \theta_0) = -\left(\frac{1}{n}\frac{\partial^2 \ln L_n(\tilde\theta_n)}{\partial\theta\,\partial\theta'}\right)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta},$$

where $\tilde\theta_n$ converges to θ_0 in probability. *Q.E.D.*

PROOF OF THEOREM 4.2: The nonsingularity of Σ_θ is now guaranteed by Assumption 9. The remaining arguments are the same as in the proof of Theorem 3.2. *Q.E.D.*

PROOF OF THEOREM 5.1: By the mean value theorem,

$$\begin{aligned}
&\frac{h_n}{n}\big[\ln L_n(\lambda) - \ln L_n(\lambda_0) - \big(Q_n(\lambda) - Q_n(\lambda_0)\big)\big]\\
&= -\frac{h_n}{2}\frac{\partial[\ln \hat\sigma_n^2(\bar\lambda_n) - \ln \sigma_n^{*2}(\bar\lambda_n)]}{\partial\lambda}(\lambda - \lambda_0)\\
&= \frac{1}{\hat\sigma_n^2(\bar\lambda_n)}\frac{h_n}{n}\left\{B_n(\bar\lambda_n) - \frac{\hat\sigma_n^2(\bar\lambda_n) - \sigma_n^{*2}(\bar\lambda_n)}{\sigma_n^{*2}(\bar\lambda_n)}A_n(\bar\lambda_n)\right\}(\lambda - \lambda_0),
\end{aligned}$$

where

$$A_n(\lambda) = (\lambda_0 - \lambda)(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) + \sigma_0^2 \operatorname{tr}(G_n' S_n(\lambda) S_n^{-1})$$

and $B_n(\lambda) = Y_n' W_n' M_n S_n(\lambda) Y_n - A_n(\lambda)$. We have $(h_n/n)(V_n' M_n G_n V_n - \sigma_0^2 \operatorname{tr}(G_n)) = o_P(1)$ and $(h_n/n)(V_n' G_n' M_n G_n V_n - \sigma_0^2 \operatorname{tr}(G_n' G_n)) = o_P(1)$ by the law of large numbers for quadratic forms; $(h_n/n)(G_n X_n \beta_0)' M_n V_n = o_P(1)$ and $(h_n/n)(G_n X_n \beta_0)' M_n G_n V_n = o_P(1)$ under Assumption 10. Therefore,

$$\begin{aligned}\frac{h_n}{n} B_n(\lambda) &= \frac{h_n}{n}\{(G_n X_n \beta_0)' M_n V_n + 2(\lambda_0 - \lambda)(G_n X_n \beta_0)' M_n G_n V_n + V_n' G_n' M_n V_n \\ &\quad + (\lambda_0 - \lambda) V_n' G_n' M_n G_n V_n - \sigma_0^2 \operatorname{tr}(G_n') - \sigma_0^2 (\lambda_0 - \lambda) \operatorname{tr}(G_n' G_n)\} \\ &= o_P(1),\end{aligned}$$

uniformly on Λ. $(h_n/n) A_n(\lambda)$ has $O(1)$ uniformly on Λ. With expressions in the proof of Theorem 3.1, $\hat\sigma_n^2(\lambda) - \sigma_n^{*2}(\lambda) = o_P(1)$ uniformly on Λ. $\sigma_n^{*2}(\bar\lambda_n)$ and $\hat\sigma_n^2(\bar\lambda_n)$ are bounded away from zero in probability. Hence, $(h_n/n)[\ln L_n(\lambda) - \ln L_n(\lambda_0) - (Q_n(\lambda) - Q_n(\lambda_0))]$ converges in probability uniformly on Λ.

$$\begin{aligned}&\frac{h_n}{n}(Q_n(\lambda) - Q_n(\lambda_0)) \\ &\quad = -\frac{h_n}{n}(\ln \sigma^{*2}(\lambda) - \ln \sigma_0^2) + \frac{h_n}{n}(\ln|S_n(\lambda)| - \ln|S_n(\lambda_0)|)\end{aligned}$$

is uniformly equicontinuous on Λ. Firstly, $(h_n/n)(\ln|S_n(\lambda_2)| - \ln|S_n(\lambda_1)|) = (h_n/n) \times \operatorname{tr}(W_n S_n^{-1}(\bar\lambda_n))(\lambda_2 - \lambda_1)$ by the mean value theorem, and it is uniformly equicontinuous on Λ because $(h_n/n)\operatorname{tr}(W_n S_n^{-1}(\bar\lambda_n)) = O(1)$. Also, $h_n(\ln \sigma_n^{*2}(\lambda) - \ln \sigma_0^2) = h_n(\sigma_n^{*2}(\lambda) - \sigma_0^2)/\bar\sigma_n^{*2}(\lambda)$ is uniformly continuous because $\bar\sigma_n^{*2}(\lambda)$ is uniformly bounded away from zero and

$$\begin{aligned}h_n(\sigma_n^{*2}(\lambda) - \sigma_0^2) &= (\lambda - \lambda_0)^2 \frac{h_n}{n}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) \\ &\quad + \sigma_0^2\left[2\frac{h_n}{n}\operatorname{tr}(G_n) + (\lambda_0 - \lambda)\frac{h_n}{n}\operatorname{tr}(G_n' G_n)\right](\lambda_0 - \lambda)\end{aligned}$$

is uniformly equicontinuous. The latter follows because $(h_n/n)(G_n X_n \beta_0)' M_n (G_n X_n \beta_0)$, $(h_n/n) \times \operatorname{tr}(G_n)$, and $(h_n/n)\operatorname{tr}(G_n' G_n)$ are of $O(1)$. For identification, let

$$D_n(\lambda) = -\frac{h_n}{2}(\ln \sigma_n^2(\lambda) - \ln \sigma_0^2) + \frac{h_n}{n}(\ln|S_n(\lambda)| - \ln|S_n(\lambda_0)|).$$

Then, $(h_n/n)(Q_n(\lambda) - Q_n(\lambda_0)) = D_n(\lambda) - (h_n/2)(\ln \sigma_n^{*2}(\lambda) - \ln \sigma_n^2(\lambda))$. Assumption 10(a) implies that $\lim_{n\to\infty} h_n(\ln \sigma_n^{*2}(\lambda) - \ln \sigma_n^2(\lambda)) > 0$ for any $\lambda \neq \lambda_0$. Also, $D_n(\lambda) < 0$ whenever $\lambda \neq \lambda_0$ under Assumption 10(b). Overall, $\lim_{n\to\infty}(h_n/n)(Q_n(\lambda) - Q_n(\lambda_0)) < 0$ whenever $\lambda \neq \lambda_0$. Together, these imply that λ_0 is uniquely identifiable. The consistency of $\hat\lambda_n$ follows. *Q.E.D.*

PROOF OF THEOREM 5.2: The first- and second-order derivatives of the concentrated log-likelihood are

$$\frac{\partial \ln L_n(\lambda)}{\partial \lambda} = \frac{1}{\hat\sigma_n^2(\lambda)} Y_n' W_n' M_n S_n(\lambda) Y_n - \operatorname{tr}(W_n S_n^{-1}(\lambda)) \quad \text{and}$$

$$\frac{\partial^2 \ln L_n(\lambda)}{\partial \lambda^2} = \frac{2}{n\hat\sigma_n^4(\lambda)}(Y_n' W_n' M_n S_n(\lambda) Y_n)^2 - \frac{1}{\hat\sigma_n^2(\lambda)} Y_n' W_n' M_n W_n Y_n - \operatorname{tr}([W_n S_n^{-1}(\lambda)]^2),$$

where $\hat{\sigma}_n^2(\lambda) = (1/n)Y_n'S_n'(\lambda)M_nS_n(\lambda)Y_n$. For the pure SAR process, $\beta_0 = 0$ and the corresponding derivatives are similar with M_n replaced by the identity I_n.

Under Assumption 10,

$$\frac{h_n}{n}Y_n'W_n'M_nW_nY_n = \frac{h_n}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \frac{h_n}{n}V_n'G_n'M_nG_nV_n + o_P(1)$$

and

$$\begin{aligned}&\frac{h_n}{n}Y_n'W_n'M_nS_n(\lambda)Y_n\\ &\quad= \frac{h_n}{n}V_n'G_n'M_nV_n + (\lambda_0-\lambda)\left[\frac{h_n}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \frac{h_n}{n}V_n'G_n'M_nG_nV_n\right]\\ &\qquad+ o_P(1).\end{aligned}$$

When $\lim_{n\to\infty} h_n = \infty$, $(1/n)Y_n'W_n'M_nS_n(\lambda)Y_n = o_P(1)$ and $\hat{\sigma}_n^2(\lambda) = \sigma_0^2 + o_P(1)$ uniformly on Λ. Hence,

$$\begin{aligned}\frac{h_n}{n}\frac{\partial^2 \ln L_n(\lambda)}{\partial\lambda^2} &= -\frac{1}{\sigma_0^2}\left[\frac{h_n}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \frac{h_n}{n}V_n'G_n'M_nG_nV_n\right]\\ &\quad - \frac{h_n}{n}\operatorname{tr}\big([W_nS_n^{-1}(\lambda)]^2\big) + o_P(1),\end{aligned}$$

uniformly on Λ. Under Assumption 7, $(h_n/n)\operatorname{tr}(G_n^3(\lambda)) = O(1)$ uniformly on Λ. Therefore, by the Taylor expansion,

$$\begin{aligned}\frac{h_n}{n}\left(\frac{\partial^2 \ln L_n(\tilde{\lambda}_n)}{\partial\lambda^2} - \frac{\partial^2 \ln L_n(\lambda_0)}{\partial\lambda^2}\right) &= -\frac{h_n}{n}\left\{\operatorname{tr}\big([W_nS_n^{-1}(\tilde{\lambda}_n)]^2\big) - \operatorname{tr}(G_n^2)\right\} + o_P(1)\\ &= -2\frac{h_n}{n}\operatorname{tr}(G_n^3(\bar{\lambda}_n))(\tilde{\lambda}_n - \lambda_0) + o_P(1)\\ &= o_P(1)\end{aligned}$$

for any $\tilde{\lambda}_n$ which converges in probability to λ_0.

Define

$$P_n(\lambda_0) = -\frac{1}{\sigma_0^2}[(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + V_n'G_n'M_nG_nV_n] - \operatorname{tr}(G_n^2).$$

Then

$$\frac{h_n}{n}\frac{\partial^2 \ln L_n(\lambda_0)}{\partial\lambda^2} = \frac{h_n}{n}P_n(\lambda_0) + o_P(1) \quad \text{and}$$

$$E(P_n(\lambda_0)) = -\frac{(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)}{\sigma_0^2} - [\operatorname{tr}(G_nG_n') + \operatorname{tr}(G_n^2)] + O(1).$$

Because $(h_n/n)[P_n(\lambda_0) - E(P_n(\lambda_0))] = -(1/\sigma_0^2)\Delta_n + o(1)$, where

$$\Delta_n = \frac{h_n}{n}[V_n'G_n'M_nG_nV_n - \sigma_0^2\operatorname{tr}(G_n'M_nG_n)] = o_P(1),$$

$$\frac{h_n}{n}\big[P_n(\lambda_0) - E(P_n(\lambda_0))\big] = o_P(1).$$

One has

$$\sqrt{\frac{h_n}{n}}\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} = \frac{1}{\hat{\sigma}_n^2(\lambda_0)}\sqrt{\frac{h_n}{n}}[(G_nX_n\beta_0)'M_nV_n + q_n],$$

where $q_n = V_n'C_n'M_nV_n$. The mean and variance of q_n are $E(q_n) = \sigma_0^2 \operatorname{tr}(M_nC_n) = O(1)$ and

$$\sigma_{q_n}^2 = (\mu_4 - 3\sigma_0^4)\sum_{i=1}^{n} C_{n,ii}^2 + \sigma_0^4[\operatorname{tr}(C_n'C_n) + \operatorname{tr}(C_n^2)] + O(1).$$

The variance of $((G_nX_n\beta_0)'M_nV_n + q_n)$ is

$$\sigma_{lq_n}^2 = \sigma_0^2(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \sigma_{q_n}^2 + 2(G_nX_n\beta_0)'M_n\text{vec}_D(C_n'M_n)\mu_3.$$

Since $(q_n - E(q_n))/\sigma_{lq_n} \xrightarrow{D} N(0,1)$ by the central limit theorem for linear-quadratic functions (Appendix A), it follows that

$$\sqrt{\frac{h_n}{n}}\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} = \frac{\sqrt{\frac{h_n}{n}}\sigma_{lq_n}}{\hat{\sigma}_n^2(\lambda_0)}\cdot\frac{[(G_nX_n\beta_0)'M_nV_n + (q_n - E(q_n))]}{\sigma_{lq_n}} + o_P(1)$$

$$\xrightarrow{P} N\left(0, \lim_{n\to\infty}\frac{h_n}{n}\frac{\sigma_{lq_n}^2}{\sigma_0^4}\right). \qquad \textit{Q.E.D.}$$

PROOF OF THEOREM 5.3: Note that

$$\hat{\beta}_n(\hat{\lambda}_n) - \beta_0 = (X_n'X_n)^{-1}X_n'V_n - (\hat{\lambda}_n - \lambda_0)(X_n'X_n)^{-1}X_n'G_nX_n\beta_0 + O_p\big(\sqrt{h_n}/n\big).$$

Therefore,

$$\sqrt{n/h_n}(\hat{\beta}_n(\hat{\lambda}_n) - \beta_0) = -\sqrt{n/h_n}(\hat{\lambda}_n - \lambda_0)\cdot(X_n'X_n)^{-1}X_n'G_nX_n\beta_0 + O_p\big(1/\sqrt{h_n}\big),$$

and its limited distribution is a linear function of that of $\hat{\lambda}_n$. If β_0 is zero,

$$\begin{aligned}\sqrt{n}(\hat{\beta}_n(\hat{\lambda}_n) - \beta_0) &= (X_n'X_n/n)^{-1}X_n'V_n/\sqrt{n} + O_p\big(\sqrt{h_n/n}\big)\\ &\xrightarrow{D} N\Big(0, \sigma_0^2 \lim_{n\to\infty}(X_n'X_n/n)^{-1}\Big).\end{aligned}$$

For $\hat{\sigma}_n^2$,

$$\begin{aligned}&\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)\\ &= \frac{1}{\sqrt{n}}(V_n'V_n - n\sigma_0^2) - \frac{1}{\sqrt{n}}V_n'X_n(X_n'X_n)^{-1}X_n'V_n\\ &\quad - 2\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0)\cdot\frac{\sqrt{h_n}}{n}Y_n'W_n'M_nS_nY_n + \sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0)^2\cdot\frac{\sqrt{h_n}}{n}Y_n'W_n'M_nW_nY_n.\end{aligned}$$

Under Assumption 10, $(\sqrt{h_n}/n)Y_n'W_n'M_nS_nY_n = O(1/\sqrt{h_n})$ and $(\sqrt{h_n}/n)Y_n'W_n'M_nW_nY_n = O(1/\sqrt{h_n})$. Hence, as $\lim_{n\to\infty} h_n = \infty$, $\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2) = (1/\sqrt{n})(V_n'V_n - n\sigma_0^2) + o_P(1) \xrightarrow{D} N(0, \mu_4 - \sigma^4)$. *Q.E.D.*

PROOF OF THEOREM 5.4: Let $X_n = (X_{1n}, X_{2n})$, $M_{1n} = I_n - X_{1n}(X'_{1n}X_{1n})^{-1}X'_{1n}$, and $M_{2n} = I_n - X_{2n}(X'_{2n}X_{2n})^{-1}X'_{2n}$. Using a matrix partition for $(X'_nX_n)^{-1}$,

$$
\begin{aligned}
&\sqrt{\frac{n}{h_n}}(\hat{\beta}_{n1} - \beta_{01}) \\
&= \frac{1}{\sqrt{h_n}}\left(\frac{1}{n}X'_{1n}M_{2n}X_{1n}\right)^{-1}\frac{1}{\sqrt{n}}X'_{1n}M_{2n}V_n - c_{1n}\cdot\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= -c_{1n}\cdot\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) + O_P\left(\frac{1}{\sqrt{h_n}}\right),
\end{aligned}
$$

and $\sqrt{n}(\hat{\beta}_{n2} - \beta_{20}) = (X'_{2n}M_{1n}X_{2n}/n)^{-1}\cdot(1/\sqrt{n})X'_{2n}M_{1n}V_n + O_P(\sqrt{h_n/n})$. The asymptotic distributions of $\hat{\beta}_{n1}$ and $\hat{\beta}_{n2}$ follow. *Q.E.D.*

REFERENCES

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

ANSELIN, L., AND A. K. BERA (1998): "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," in *Handbook of Applied Economics Statistics*, ed. by A. Ullah and D. E. A. Giles. New York: Marcel Dekker.

BELL, K. P., AND N. E. BOCKSTAEL (2000): "Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data," *The Review of Economics and Statistics*, 82, 72–82.

BESLEY, T., AND A. CASE (1995): "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition," *The American Economic Review*, 85, 25–45.

BERTRAND, M., E. F. P. LUTTMER, AND S. MULLAINATHAN (2000): "Network Effects and Welfare Cultures," *Quarterly Journal of Economics*, 115, 1019–1055.

BRUECKNER, J. K. (1998): "Testing for Strategic Interaction Among Local Governments: The Case of Growth Controls," *Journal of Urban Economics*, 44, 438–467.

CASE, A. C. (1991): "Spatial Patterns in Household Demand," *Econometrica*, 59, 953–965.

CASE, A. C., H. S. ROSEN, AND J. R. HINES (1993): "Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States," *Journal of Public Economics*, 52, 285–307.

CLIFF, A. D., AND J. K. ORD (1973): *Spatial Autocorrelation*. London: Pion Ltd.

CONLEY, T. G. (1999): "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92, 1–45.

CRESSIE, N. (1993): *Statistics for Spatial Data*. New York: John Wiley & Sons.

CRESSIE, N., AND S. N. LAHIRI (1993): "The Asymptotic Distribution of REML Estimators," *Journal of Multivariate Analysis*, 45, 217–233.

GIRAITIS, L., AND M. S. TAQQU (1998): "Central Limit Theorems for Quadratic Forms with Time-Domain Conditions," *The Annals of Probability*, 26, 377–398.

HORN, R., AND C. JOHNSON (1985): *Matrix Analysis*. New York: Cambridge University Press.

KELEJIAN, H. H., AND I. R. PRUCHA (1998): "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17, 99–121.

——— (1999): "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economics Review*, 40, 509–533.

——— (2001): "On the Asymptotic Distribution of the Moran I Test Statistic with Applications," *Journal of Econometrics*, 104, 219–257.

LEE, L. F. (2002): "Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models," *Econometric Theory*, 18, 252–277.

——— (2003): "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances," *Econometric Reviews*, 22, 307–335.
MANSKI, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60, 531–542.
MARDIA, K. V., AND R. J. MARSHALL (1984): "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression," *Biometrika*, 73, 135–146.
ORD, J. K. (1975): "Estimation Methods for Models of Spatial Interaction," *Journal of American Statistical Association*, 70, 120–126.
PINKSE, J. (1999): "Asymptotic Properties of Moran and Related Tests and Testing for Spatial Correlation in Probit Models," Unpublished Manuscript, University of British Columbia.
PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING, AND B. P. FLANNERY (1992): *Numerical Recipes* (Second Ed.). New York: Cambridge University Press.
RIPLEY, B. D. (1988): *Statistical Inference for Spatial Processes*. New York: Cambridge University Press.
SMIRNOV, O., AND L. ANSELIN (2001): "Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: A Characteristic Polynomial Approach," *Computational Statistics and Data Analysis*, 35, 301–319.
STEIN, M. L. (1999): *Interpolation of Spatial Data*. New York: Springer-Verlag.
TOPA, G. (2001): "Social Interactions, Local Spillovers and Unemployment," *Review of Economic Studies*, 68, 261–295.
WHITE, H. (1984): *Asymptotic Theory for Econometricians*. Orlando: Academic Press.
——— (1994): *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.