Asymptotic Distributions of Quasi-Maximum Likelihood Estimators
for Spatial Econometric Models I: Spatial Autoregressive Processes

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

August 1999; March 2001

Abstract

Asymptotic properties of MLEs and QMLEs of spatial autoregressive processes are investigated. The stochastic rate of convergence of the MLE and QMLE for a spatial autoregressive process may be less than the $\sqrt{n}$-rate under some circumstances even though its limiting distribution is asymptotically normal. Implications of the possible low rate of convergence of the estimators on classical statistics such as the likelihood ratio, Wald, and efficient score statistics are analyzed.

Key Words:

Spatial autoregressive process, spatial autoregression, maximum likelihood estimator, quasi-maximum likelihood estimator, moment estimator, rate of convergence, asymptotic distribution, test statistics.

Correspondence Address:
Lung-fei Lee, Department of Economics, The Ohio State University, 410 Arps Hall, 1945 N. High St., Columbus, OH 43210-1172.
E-mail: llee@econ.ohio-state.edu

* This paper is a revised and expanded version of the first part of a paper previously circulated under the title *Asymptotic Distributions of Maximum Likelihood Estimators for Spatial Autoregressive Models*. The 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.
1. Introduction

Spatial econometrics consist of econometric techniques dealing with empirical economic problems caused by spatial autocorrelation in cross-sectional and/or panel data. Issues about spatial dependence are receiving more attention in both empirical and theoretical econometrics. Possible dependence across spatial units is a relevant issue in a range of economic fields including urban, real estate, regional, public, agricultural, and environmental economics as well as industrial organizational studies (see, e.g., the recent survey by Anselin and Bera, 1998, for references). There are a few books dealing with spatial econometrics issues (Cliff and Ord, 1973; Paelinck and Klaassen, 1979; Anselin, 1988; Cressie, 1993) and several collections of spatial econometrics papers published in special journals for regional and urban economics (Anselin, 1992; Anselin and Florax, 1995; Anselin and Rey, 1997).

To capture spatial dependence in cross-sectional units, the approach in spatial econometrics is to impose structures on the specification of a model. The spatial autoregressive models were first popularized by Whittle (1954), Mead (1967) and Ord (1975). While spatial autoregression extends autocorrelation in time series to spatial dimensions, the spatial aspect of a model has its own distinctive features that are not present in the time domain. Spatial autocorrelation by Whittle (1954) has the distinguishing feature of simultaneity in econometric equilibrium models.

The autoregressive model can be estimated by the method of maximum likelihood (ML). Ord (1975) discussed a computationally effective procedure to implement the ML method. As an alternative, Kelejian and Prucha (1999a) have suggested a generalized moments estimation method. In this paper, we investigate the possible rate of convergence and the asymptotic distribution of the maximum likelihood estimator (MLE) and the quasi-maximum likelihood estimator (QMLE) for the first-order spatial autoregressive model. The MLE is the one when disturbances are normally distributed. The QMLE is appropriate when the estimator is derived from a normal likelihood but the disturbances are not truly normally distributed. In the existing literature, the MLE of such a model is implicitly regarded to have the familiar $\sqrt{n}$-rate of convergence as the usual MLE for a nonlinear parametric statistical model (see, e.g., the reviews by Anselin, 1988 and Anselin and Bera, 1998). Our investigation reported 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 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 of the model’s parameters and may even be inconsistent. Possible implications for inferences are discussed.

This paper is organized as follows. In Sections 2, the spatial autoregressive model is presented and regular conditions for the model are specified. In Section 3, identification of the parameters in the model is investigated. Consistency of the QMLE and MLE are investigated. In Section 4, we investigate the rates of convergence and the asymptotic distributions of the estimators. All the proofs of the results are presented in Appendices. Section 5 concludes the paper. Appendix A collects some useful basic lemmas and propositions. Proofs of theorems are in Appendix B. Appendix C provides brief descriptions on some computationally simpler estimators.

2. The Spatial Autoregressive Model

The first-order spatial autoregressive model is specified as

$$Y_n = \lambda W_{n,n}Y_n + V_n,$$  \hspace{1cm} (2.1)

where $n$ is the total number of cross-sectional units, $Y_n$ is an $n$-dimensional vector of sample observations of the dependent variable, $W_{n,n}$ is an $n \times n$ matrix of constants known as spatial weights, $\lambda$ is a scalar unknown coefficient with $|\lambda| < 1$, and $V_n$ is an $n$-dimensional vector of i.i.d. disturbances with zero mean and a constant variance $\sigma^2$. This and related models were originated from Whittle (1954), Mead (1967), Cliff and Ord (1973), and Ord (1975). For the $i$th cross-sectional unit, this model specifies that

$$y_{ni} = \lambda w_{i,n}Y_n + v_i$$ \hspace{1cm} (2.1)'

where $y_{ni}$ and $v_i$ are, respectively, the $i$th element of $Y_n$ and $V_n$, and $w_{i,n}$ is the $i$th row of $W_{n,n}$. In geographic analysis, weights are functions of distance. In sociometrics, the weights reflect social networks of individuals (Doreian, 1980). In economic applications, the weights may be based on ‘economic’ distance such as income or racial composition (Case et al., 1993). As a convention, it is understood that $w_{n,ii} = 0$ for all $i$, where $w_{n,ij}$ is the $(i, j)$ entry of $W_{n,n}$.

The parameter vector $\theta = (\lambda, \sigma^2)$ in (2.1) can be estimated by various methods. The popular estimation method is the method of maximum likelihood (ML) as suggested by Ord (1975). The log likelihood function of the model under normal disturbances is

$$\ln L_n(\theta) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln |I_n - \lambda W_{n,n}| - \frac{1}{2\sigma^2}Y_n'(I_n - \lambda W_{n,n})'(I_n - \lambda W_{n,n})Y_n.$$ \hspace{1cm} (2.2)
Alternatively, a method of moments has been suggested by Kelejian and Prucha (1999a). In this paper, we concentrate on the maximum likelihood estimator (MLE) and investigate its asymptotic distribution.\footnote{There are interesting computational issues that will not be addressed in this paper. Interested readers may consult Ord (1975) and find recent developments in Smirnov and Anselin (1999).} The method of moments will be touched upon briefly in an appendix. When $v_i$'s are not truly normally distributed, the MLE is a quasi-likelihood estimator (QMLE). To allow the later possibility, we investigate the asymptotic property of a QMLE. The QMLE will reduce to the true MLE when $v_i$'s are truly normally distributed.

To provide a rigorous analysis of the QMLE or MLE, we assume the following regularity conditions for the model.

**Assumption 1.** The disturbances $\{v_i\}$, $i = 1, \cdots, n$, are i.i.d. with mean zero and variance $\sigma^2$. Its moment $E(|v|^{4+2\delta})$ for some $\delta > 0$ exits.

**Assumption 2.** The elements $w_{n,ij}$ of the weights matrix $W_{n,n}$ are of order $O(\frac{1}{h_n})$ uniformly in all $i, j$, i.e., there exists a constant $c$ such that $|w_{n,ij}| \leq \frac{c}{h_n}$ for all $i, j$ and $n$, where the sequence of rates $\{h_n\}$ can be bounded or divergent.

**Assumption 3.** The ratio $\frac{h_n}{n} \to 0$ as $n$ goes to infinity.

Assumptions 2 and 3 describe how the elements of $W_{n,n}$ are related to sample size $n$.\footnote{Manski (1993) criticized the literature on spatial models for not having specified how the weights $w_{n,ij}$ should change with $n$ (footnote 7 in Manski). Our assumptions make this explicit.} Such features play important roles for asymptotic properties of an estimator. Assumption 3 is always satisfied if $\{h_n\}$ is a bounded sequence. Assumptions 2 and 3 are motivated by the design of row-normalized weights matrix. In empirical applications, it is a common practice to have the spatial weights matrix $W_{n,n}$ being row-normalized (or row-standardized, Haining, 1990, p.82) such that

$$w_{i,n} = [d_{i1}, d_{i2}, \ldots, d_{in}] / \sum_{j=1}^{n} d_{ij}$$

(2.3)

where $d_{ij}$ represents a function of the spatial distance of the $i$th and $j$th units in some (characteristic) space. The reason weights matrices are row-normalized is to ensure that all weights are between 0 and 1 and weighting operations can be interpreted as an average of neighboring values (e.g., Anselin and Rey, 1991; Kelejian and Robinson, 1993). The largest eigenvalue of a row-normalized weights matrix is always 1, which is often claimed to facilitate the interpretation of the autoregressive coefficient $\lambda$ (Ord, 1975; Anselin and Bera, 1998, p.243). For row-normalized weights matrix according to (2.3), as $d_{i,j}$ are nonnegative constants
and uniformly bounded and the sums \( \sum_{j=1}^{n} d_{ij}, \, i = 1, \ldots, n \), are uniformly bounded away from zero at the rate \( h_n \), the implied normalized weights matrix will have the property ascribed in Assumption 2. The assumption 3 leads us to focus on the cases where \( \sum_{j=1}^{n} d_{ij}, \, i = 1, \ldots, n \), do not diverge to infinity at a rate equal to or faster than the rate of the sample size \( n \). Assumptions 2 and 3 are, however, more general in that they shall cover spatial weights matrices where elements are not restricted to be nonnegative and weights matrices which might not be row-normalized. The following established asymptotic properties of the QMLE can thus be valid within a broad class of model framework.

The exact rate \( h_n \) in Assumption 2 will depend on a specified weights matrix. For a sample with \( n \) cross sectional units, one has to imagine how a single unit may influence other available cross sectional units. For example, in an empirical study of states expenditures in Case et al. (1993), several weights matrices have been tried. The simplest weights matrix specifies that \( d_{ij} = 1/s_i \), where \( s_i \) is the number of borders state \( i \) shares, if states \( i \) and \( j \) share a border. With this specification, it is reasonable to argue that \( \{h_n\} \) would be bounded even if more states could be available. An important case that \( h_n \) might diverge to infinity and satisfies Assumptions 2 and 3 can be that in Case (1991, 1992). Case (1992) studied the possible spatial effect of neighbors on a farmer’s adoption of new technologies. In her study, ‘neighbors’ refer to farmers who live in the same district (in rural Java, Indonesia). This characterization of the neighbors makes the \( W_{n,n} \) matrix block diagonal. The only non-zero elements appear as a block of households 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, \( W_{n,n} = I_r \otimes B_{m,m} \), where \( \otimes \) is the Kronecker product, is a symmetric block diagonal matrix with \( B_{m,m} = \frac{1}{(m-1)}(l_m l_m' - I_m) \), where \( l_m \) is a \( m \)-dimensional column vector of ones. For this example, \( h_n = (m-1) \) and \( \frac{h_n}{n} = (\frac{m-1}{m}) \cdot \frac{1}{r} = O(\frac{1}{r}) \). If the increase of sample size \( n \) is generated by the increase of both \( r \) and \( m \), then \( h_n \) goes to infinity and \( \frac{h_n}{n} \) goes to zero as \( n \) tends to infinity. In another study by Case (1991) on a consumer demand problem, a similar spatial weights matrix is utilized. Assumption 3, however, rules out the cases with \( h_n \) being equal or larger than \( n \) for the spatial autoregressive process as the MLE would likely be inconsistent for such cases.\(^3\) Examples will be provided for the latter.

**Assumption 4.** At the true parameter \( \lambda_0 \), \( S_n = I_n - \lambda_0 W_{n,n} \), where \( I_n \) is the \( n \times n \) dimensional identity matrix, is nonsingular.

\(^3\) Note that the rate \( h_n \) can not be faster than \( n \) when \( d_{ij} \) are uniformly bounded in (2.3).
Assumption 4 guarantees that the system (2.1) has an equilibrium and

\[ Y_n = S_n^{-1}V_n, \]  

(2.4)

which has zero mean and variance \( \sigma_0^2 S_n^{-1} \sigma_0^{-1} \), where \( \sigma_0^2 \) denotes the true variance parameter of \( v_i \). A sufficient condition for Assumption 4 is \( \| \lambda_0 W_{n,n} \| < 1 \) for some matrix norm \( \| \cdot \| \) (Horn and Johnson 1985, p301). If \( W_{n,n} \) consists of nonnegative entries and is row-normalized, a sufficient condition for \( S_n(\lambda) \) to be invertible is that \( |\lambda| < 1 \) (Horn and Johnson, 1985). With elements of \( W_{n,n} \) being normalized as in (2.3), if \( d_{ij} = d_{ji} \) for all \( i, j \), Ord (1975) showed that the eigenvalues of \( W_{n,n} \) are all real. For the latter, assuming that the largest eigenvalue \( \omega_{\text{max}} \) of \( W_{n,n} \) is positive and the smallest eigenvalue \( \omega_{\text{min}} \) is negative, a sufficient condition for the invertibility of \( S_n \) is \( \frac{1}{\omega_{\text{min}}} < \lambda_0 < \frac{1}{\omega_{\text{max}}} \). With \( d_{ij} \) being nonnegative, it is known that \( \omega_{\text{max}} = 1 \) and the sufficient condition becomes \( \frac{1}{\omega_{\text{min}}} < \lambda_0 < 1 \).

**Definition:** Let \( \{A_n\} \) be a sequence of square \( n \times n \) matrices, where \( A_n = (a_{n,ij}) \).

1. The **column sums** of \( \{A_n\} \) are uniformly bounded (in absolute value) if there exists a finite constant \( c \) that does not depend on \( n \) such that \( \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{n,ij}| \leq c. \)

2. The **row sums** of \( \{A_n\} \) are uniformly bounded (in absolute value) if there exists a finite constant \( c \) that does not depend on \( n \) such that \( \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{n,ij}| \leq c. \)

The preceding notions of uniform boundedness can be defined in terms of some matrix norms. The maximum column sum matrix norm \( \| \cdot \|_1 \) of a square \( n \times n \) matrix \( A = (a_{ij}) \) is defined as \( \| A_n \|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}| \), and the maximum row sum matrix norm \( \| \cdot \|_{\infty} \) is \( \| A_n \|_{\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| \) (see Horn and Johnson (1985), pp.294-295). The uniformly boundedness of \( \{A_n\} \) in column (resp. row) sums is equivalent to the sequence \( \{ \| A_n \|_1 \} \) (resp. \( \{ \| A_n \|_{\infty} \} \) ) being bounded.

The uniform boundedness of row and column sums plays an important role in the asymptotic properties of estimators of spatial econometric models. It limits the correlation of \( y_{ni} \) across different units to a manageable degree (Kelejian and Prucha, 1998 and 1999a). We follow Kelejian and Prucha (1998) by imposing the following assumption in the model.\(^4\)

**Assumption 5.** The matrices \( \{W_{n,n}\} \) and \( \{S_{n}^{-1}\} \) are uniformly bounded in both row and column sum.

When \( W_{n,n} \) is row-normalized as in (2.3) with nonnegative \( d_{ij} \), \( \| W_{n,n} \|_{\infty} = 1 \) and, hence, its row sums must be uniformly bounded. Furthermore, if \( |\lambda_0| < 1 \), \( S_{n}^{-1} \) can be expanded into a convergent series \( \sum_{i=0}^{\infty} \lambda_i^i W_{n,n}^i \) (Horn and Johnson, 1985). From the series representation and \( W_{n,n} \) being row-normalized, \( S_{n}^{-1} \) must be

\(^4\) Related conditions have also been adopted in Pinkse (1999) in a different context.
uniformly bounded in row sums. For a symmetric matrix, uniform boundedness in row sums is equivalent to uniform boundedness in column sums. When $W_{n,n}$ is symmetric, $S_n$ will be symmetric. Thus, for a row-normalized symmetric $W_{n,n}$ with nonnegative elements, $W_{n,n}$ and $S_n$ with $|\lambda_0| < 1$ will satisfy Assumption 5. This is so for the spatial weights matrices in Case (1991, 1992). Effective restrictions by Assumption 5 are imposed on the column sums of $W_{n,n}$ and $S_n^{-1}$ if they are not symmetric matrices. The $\{W_{n,n}\}$ with (2.3) can be uniformly bounded in column sums under some conditions on $d_{ij}$ as stated in Lemma A.1 of Appendix A. The uniform boundedness of the column sums of $\{S_n^{-1}\}$ may impose restrictions on $\lambda_0$ as well as $W_{n,n}$. Lemma A.2 of Appendix A shows that, for any weights matrix, $\|\lambda_0 W_{n,n}\|_1 < 1$ and $\|\lambda_0 W_{n,n}\|_\infty < 1$ for all $n$, are sufficient conditions for $\{S_n^{-1}\}$ to be uniformly bounded in both row and column sums. Because a matrix norm $\|\cdot\|$ has the submultiplicative property that $\|A_n B_n\| \leq \|A_n\| \cdot \|B_n\|$, it is immediate that if $\{A_n\}$ and $\{B_n\}$ are matrices uniformly bounded in row sums, $\{A_n B_n\}$ is uniformly bounded in row sums; and if $\{A_n\}$ and $\{B_n\}$ are uniformly bounded in column sums, $\{A_n B_n\}$ is uniformly bounded in column sums.

In our subsequent analysis, products of matrices such as $W_{n,n}' W_{n,n}$ and $S_n^{-1} S_n^{-1}$, etc., appear. Assumption 5 guarantees that those products of matrices will be uniformly bounded in row and column sums.

For any $\lambda$, denote $S_n(\lambda) = I_n - \lambda W_{n,n}$. Then, $S_n$ corresponds to $S_n(\lambda_0)$. Assumption 5 imposes the uniform boundedness conditions for $S_n^{-1}(\lambda)$ only at the point $\lambda = \lambda_0$. But, as shown in Lemma A.3 of Appendix A, this is sufficient to guarantee that $S_n^{-1}(\lambda)$ are uniformly bounded in row and column sums, uniformly in a neighborhood of $\lambda_0$. Indeed, if $\|W_{n,n}\|_1 \leq 1$ and $\|W_{n,n}\|_\infty \leq 1$ for all $n$, Lemma A.4 shows that $S_n^{-1}(\lambda)$ will be uniformly bounded in row and column sums, uniformly in any closed subset of $(-1,1)$. These uniform properties in $\lambda$ are useful in handling the nonlinearity of $S_n^{-1}(\lambda)$ in the analysis of the (quasi) likelihood function of the model.

3. Identification and Consistency

The log likelihood function of this model is

$$\ln L_n(\theta) = \frac{n}{2} \ln (2\pi) - \frac{n}{2} \ln \sigma^2 + \ln |S_n(\lambda)| - \frac{1}{2\sigma^2} Y_n' S_n^{-1}(\lambda) S_n(\lambda) Y_n. \quad (3.1)$$

As in common practice, $E(\ln L_n(\theta))$ refers to the expectation of $\ln L_n(\theta)$ with respect to the true distribution of $Y_n$. When $L_n(\theta)$ in (3.1) is a quasi-likelihood, $E^*(\ln L_n(\theta))$ shall denote the expectation of $\ln L_n(\theta)$ with respect to the normal distribution with mean 0 and variance $\sigma_0^2 S_n^{-1} S_n^{-1}$ to distinguish it from the true expectation. However, the value of $E(\ln L_n(\theta))$ is invariant with respect to any distribution of $V_n$ with
\( E(V_nV_n') = \sigma_0^2 I_n \). Thus, \( E(\ln L_n(\theta)) = E^*(\ln L_n(\theta)) \) when the variance of \( V_n \) is correctly specified and there is no need to distinguish \( E^*(\ln L_n(\theta)) \) from \( E(\ln L_n(\theta)) \) in our subsequent analysis.

Under Assumptions 1-5, the following theorem shows that \( \frac{1}{n} \ln L_n(\theta) - E\left(\frac{1}{n} \ln L_n(\theta)\right) \) converges to zero in probability uniformly in a compact parameter space \( \Theta \) of \( \theta \).

**Theorem 1.** Under Assumptions 1-5, \( \sup_{\theta \in \Theta} \left| \frac{1}{n} \ln L_n(\theta) - E\left(\frac{1}{n} \ln L_n(\theta)\right) \right| \xrightarrow{p} 0 \).

The consistency of the QMLE \( \hat{\theta}_n \) would follow from the uniform convergence if the identified uniqueness assumption (White, 1994, Theorem 3.4) were satisfied. For each \( n \), Jensen’s inequality guarantees that \( E(\ln L_n(\theta)) \) is uniquely maximized at the true parameter vector \( \theta_0 = (\lambda_0, \sigma_0^2) \). For identification, a sufficient condition called the identifiable uniqueness condition in White (1994, p.28) is needed to guarantee the uniqueness of the maximizer in the limiting process as \( n \) goes to infinity. For any \( \epsilon > 0 \), let \( N_\epsilon(\theta_0) \) be an open ball of \( \theta_0 \) with radius \( \epsilon \) and \( \bar{N}_\epsilon(\theta_0) \) denote the complement of \( N_\epsilon(\theta_0) \) contained in \( \Theta \). The identifiable uniqueness assumption is

\[
\limsup_{n \to \infty} \left[ \max_{\theta \in N_\epsilon(\theta_0)} E\left(\frac{1}{n} \ln L_n(\theta)\right) - E\left(\frac{1}{n} \ln L_n(\theta_0)\right) \right] < 0. \tag{3.2}
\]

For some cases, the identifiable uniqueness condition might not be easily verifiable. For the cases with \( \lim_{n \to \infty} h_n = \infty \), \( E\left(\frac{1}{n} \ln L_n(\theta)\right) - E\left(\frac{1}{n} \ln L_n(\theta_0)\right) \) can be very flat with respect to the component \( \lambda \) for large \( n \). This can be seen from the partial derivatives or the function itself. Consider specifically the model with a row-normalized nonnegative weights matrix in (2.3), where \( \{S^{-1}(\lambda)\} \) will be uniformly bounded in row sums uniformly in \( \lambda \) in any closed subset of \((-1,1)\) from Lemma A.4. As

\[
E\left(\frac{1}{n} \ln L_n(\theta)\right) - E\left(\frac{1}{n} \ln L_n(\theta_0)\right) = \frac{1}{2}(1 + \ln \sigma_0^2 - \ln \sigma^2) + \frac{1}{n} (\ln |S_n(\lambda)| - \ln |S_n(\lambda_0)|) - \frac{\sigma_0^2}{2\sigma^2 n} \text{tr}(S_n^{-1}S_n'\lambda S_n(\lambda)S_n^{-1}) \tag{3.3}
\]

and \( \frac{\partial \ln |S_n(\lambda)|}{\partial \lambda} = -\text{tr}(W_{n,n}S_n^{-1}(\lambda)) \) (see, e.g., Dhrymes 1978, Corollaries 30), one has

\[
\frac{\partial}{\partial \lambda} [E\left(\frac{1}{n} \ln L_n(\theta)\right) - E\left(\frac{1}{n} \ln L_n(\theta_0)\right)] = -\frac{1}{n} \text{tr}(W_{n,n}S_n^{-1}(\lambda)) + \frac{\sigma_0^2}{\sigma^2 n} \text{tr}(S_n^{-1}W_{n,n}S_n(\lambda)S_n^{-1}) = O\left(\frac{1}{h_n}\right)
\]

uniformly in \( \lambda \) in any bounded subset of \((-1,1)\), by Lemmas A.4 and A.9. Thus, if \( \lim_{n \to \infty} h_n = \infty \), this derivative at any \( \lambda \in (-1,1) \) goes to zero. Alternatively, denote \( \sigma_n^2(\lambda) = \frac{\sigma_0^2}{\sigma^2} \text{tr}(S_n^{-1}S_n'\lambda S_n(\lambda)S_n^{-1}) \). The function (3.3) can be rewritten as

\[
E\left(\frac{1}{n} \ln L_n(\theta)\right) - E\left(\frac{1}{n} \ln L_n(\theta_0)\right) = \frac{1}{2}(\ln \sigma_0^2 - \ln \sigma^2) + \frac{\sigma_0^2}{n} (\ln |S_n(\lambda)| - \ln |S_n(\lambda_0)|) - \frac{\sigma_0^2(\lambda) - \sigma_0^2}{2\sigma^2} \tag{3.3'}
\]
Denote $G_n = W_{n,n} S_n^{-1}$. Because $\sigma^2_n(\lambda)$ is a quadratic function of $\lambda$,
\begin{equation}
\sigma^2_n(\lambda) = \sigma^2_0[1 - 2(\lambda - \lambda_0)\frac{tr(G_n)}{n} + (\lambda - \lambda_0)^2 \frac{tr(G'_n G_n)}{n}],
\end{equation}
As $tr(G_n) = O(\frac{n}{h_n})$ and $tr(G'_n G_n) = O(\frac{n}{h_n})$ by Lemma A.9, $\sigma^2_n(\lambda) - \sigma^2_0 = O(\frac{1}{h_n})$. By the mean value theorem,
\begin{equation}
\frac{1}{n}[\ln |S_n(\lambda)| - \ln |S_n(\lambda_0)|] = \frac{1}{n} \frac{\partial \ln |S_n(\lambda)|}{\partial \lambda}(\lambda - \lambda_0) = - \frac{1}{n} tr(W_{n,n} S_n^{-1}(\lambda)) (\lambda - \lambda_0) = O(\frac{1}{h_n}),
\end{equation}
because $tr(W_{n,n} S_n^{-1}(\lambda)) = O(\frac{n}{h_n})$, uniformly in $\lambda \in (-1, 1)$ by Lemmas A.4 and A.9. Thus, if $h_n$ goes to infinity,
\begin{equation}
\lim_{n \to \infty} [E(\frac{1}{n} \ln L_n(\theta)) - E(\frac{1}{n} \ln L_n(\theta_0))] = \frac{1}{2}(\ln \sigma^2_0 - \ln \sigma^2) + \frac{(\sigma^2 - \sigma^2_0)}{2\sigma^2}.
\end{equation}
This observation indicates that the identifiable uniqueness condition (3.2) might only be relevant for the identification of $\sigma^2$ for given $\lambda$ or for the cases with a bounded sequence $h_n$.

For further investigation, we consider the concentrated likelihood function of $\lambda$. Given any value of $\lambda$, the QMLE of $\sigma^2$ can be solved from the first-order condition $\frac{\partial \ln L_n(\theta)}{\partial \sigma^2} = 0$, which is
\begin{equation}
\hat{\sigma}^2_n(\lambda) = \frac{1}{n} Y'_n S'_n(\lambda) S_n(\lambda) Y_n.
\end{equation}
The concentrated log likelihood function of $\lambda$ is
\begin{equation}
\ln L_n(\lambda) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \hat{\sigma}^2_n(\lambda) + \ln |S_n(\lambda)| - \frac{1}{2\hat{\sigma}^2_n(\lambda)} Y'_n S'_n(\lambda) S_n(\lambda) Y_n
\end{equation}
\begin{equation}
= -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2} \ln \hat{\sigma}^2_n(\lambda) + \ln |S_n(\lambda)|.
\end{equation}
The QMLE $\hat{\lambda}_n$ of $\lambda$ is the maximizer of $\ln L_n(\lambda)$ and it satisfies the first order condition that $\frac{\partial \ln L_n(\hat{\lambda}_n)}{\partial \lambda} = 0$.

The corresponding QMLE of $\sigma^2$ is
\begin{equation}
\hat{\sigma}^2_n(\hat{\lambda}_n) = \frac{1}{n} Y'_n S'_n(\hat{\lambda}_n) S_n(\hat{\lambda}_n) Y_n.
\end{equation}
Instead of the expected log likelihood function, a better analysis may investigate some related functions of the concentrated log likelihood function in (3.6). Given $\lambda$, consider the maximization of $E(\ln L_n(\lambda, \sigma^2))$ with respect to $\sigma^2$. The maximizer is $\sigma^2_n(\lambda) = \frac{\hat{\sigma}^2_n}{n} tr(S_n^{-1} S'_n(\lambda) S_n(\lambda) S_n^{-1})$. Define the function
\begin{equation}
Q_n(\lambda) = \max_{\sigma^2} E(\ln L_n(\lambda, \sigma^2)) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2} \ln \sigma^2_n(\lambda) + \ln |S_n(\lambda)|.
\end{equation}
\(Q_n(\lambda)\) is related to the concentrated log likelihood function \(\ln L_n(\lambda)\) as it will be shown that

\[
\frac{h_n}{n} (\ln L_n(\lambda) - Q_n(\lambda)) = -\frac{h_n}{2} (\ln \tilde{\sigma}_n^2(\lambda) - \ln \sigma_n^2(\lambda))
\]

(3.9)

can converge to zero in probability uniformly in \(\lambda\).\footnote{We note that \(Q_n(\lambda)\) is not the expectation of \(\ln L_n(\lambda)\) for any finite \(n\).}

**Theorem 2.** Suppose the assumed regularity conditions 1-5 are satisfied, then

1. there exists a neighborhood \(\Lambda\) of \(\lambda_0\) such that

\[
\frac{h_n}{n} (\ln L_n(\lambda) - Q_n(\lambda)) \xrightarrow{P} 0
\]

(3.10)

uniformly in \(\lambda\) in \(\Lambda\);

2. for the case \(\lim_{n \to \infty} h_n = \infty\), the convergence in (3.10) can be uniform in \(\lambda\) in any bounded set; and

3. when \(\{h_n\}\) is a bounded sequence, if \(\limsup_{n \to \infty} \frac{\text{tr}^2(G_n)}{n \text{tr}(G_n G_n)} < 1\), then the convergence in (3.10) can be uniform in \(\lambda\) in any bounded set.

We note that the condition \(\limsup_{n \to \infty} \frac{\text{tr}^2(G_n)}{n \text{tr}(G_n G_n)} < 1\) for a bounded sequence \(\{h_n\}\) in part (3) of the above theorem is a very mild condition on the spatial weights matrix \(W_{n,n}\). Let \(g_{n,ij}\) denote the elements of \(G_n\).

It follows that \(\text{tr}(G_n) = \sum_{l=1}^n g_{n,ll}\) and \(\text{tr}(G_n^r G_n) = \sum_{l=1}^n \sum_{k=1}^n g_{n,ik}^2\) (the square of the Euclidean matrix norm of \(G_n\)). Thus,

\[
\frac{1}{n} \text{tr}(G_n G_n) - \frac{\text{tr}^2(G_n)}{n^2} = \left[\frac{1}{n} \sum_{l=1}^n \sum_{k=1}^n g_{n,ik}^2 - \frac{1}{n} \sum_{l=1}^n g_{n,ll}^2\right] + \left[\frac{1}{n} \sum_{l=1}^n g_{n,ll}^2 - \left(\frac{1}{n} \sum_{l=1}^n g_{n,ll}\right)^2\right].
\]

As an empirical variance of \(g_{n,ll}\), \(l = 1, \cdots, n\), \(\frac{1}{n} \sum_{l=1}^n g_{n,ll}^2 - \left(\frac{1}{n} \sum_{l=1}^n g_{n,ll}\right)^2 \geq 0\). The condition will be satisfied as long as either the limiting empirical distribution of \(g_{n,ll}\) for \(l = 1, \cdots, n\) does not degenerate, i.e., \(g_{n,ll}\) does not converge to a constant for all \(l = 1, \cdots, n\) as \(n\) goes to infinity, or the average of the squares of the off diagonal elements of \(G_n\) does not tend to zero.

Consistency of the estimator \(\hat{\lambda}_n\) will follow if \(\lambda_0\) can be uniquely identified from the limiting process of \(\frac{h_n}{n} (Q_n(\lambda) - Q_n(\lambda_0))\). The identifiable uniqueness condition can be established in a neighborhood of \(\lambda_0\) under some mild regular conditions. Let \(C_n = G_n - \frac{\text{tr}(G_n)}{n} I_n\).

**Assumption 6.** The limit of the sequence \(\frac{h_n}{n} \text{tr}((C_n' + C_n)(C_n' + C_n))\) exists and is positive.

We note that \(\text{tr}((C_n' + C_n)(C_n' + C_n)) = 2[\text{tr}(C_n C_n') + \text{tr}(C_n^2)] = 2[\text{tr}(G_n G_n') + \text{tr}(G_n^2) - \frac{2}{n} \text{tr}^2(G_n)].\) By Lemma A.9, \(\frac{h_n}{n} \text{tr}((C_n' + C_n)(C_n' + C_n)) = O(1)\). It is, in general, positive. This is so as follows. For any matrix \(A\), \(\text{tr}(AA')\) is the square of the Euclidean norm of \(A\). Therefore, \(\text{tr}(AA') = 0\) if and only if \(A = 0\).
For our case, $C_n + C_n = G_n + G_n - 2\frac{tr(G_n)}{n}I_n$ is, in general, not a zero matrix. If it were zero, it would imply that all the diagonal elements of $G_n$ would equal each other and $G_n$ would be a skew-symmetric matrix, i.e., $G_n' = -G_n$. Thus, in general, $\frac{h_n}{n}[tr(C_n' + C_n)(C_n' + C_n)] > 0$ for each $n$. Assumption 6 rules out irregular cases that the limit of the relevant sequence vanishes. For the example from Case (1992) with $W_{n,m} = I_r \otimes B_{m,m}$ where $r$ is the number of districts and $m$ is the number of households in a district,

$$
\frac{h_n}{n}[tr(C_n' + C_n)(C_n' + C_n)]' = 4\frac{(m - 1)}{m}[(\frac{m}{m - 1 + \lambda_0})^2 \frac{(1 - 2(1 - \lambda_0)/m}{(1 - \lambda_0)^2} + \frac{1}{m}) - \frac{1}{m}(\frac{\lambda_0}{1 - \lambda_0})^2(1 - \frac{(1 - \lambda_0)}{m})^{-2}] \to \frac{4}{(1 - \lambda_0)^2},
$$

which is finite and positive (as $m$ goes to infinity).

**Theorem 3.** Under the assumed conditions, there exists a set $\Lambda$ with $\lambda_0$ in its interior such that, for any neighborhood $N_\epsilon(\lambda_0)$ of $\lambda_0$ with radius $\epsilon$,

$$
\lim_{n \to \infty} \sup_{\lambda \in \Lambda \setminus N_\epsilon(\lambda_0)} \frac{h_n}{n}Q_n(\lambda) - \frac{h_n}{n}Q_n(\lambda_0) < 0.
$$

Furthermore, the QMLE $\lambda_n$ derived from the maximization of $\ln L_n(\lambda)$ with $\lambda$ in $\Lambda$ is a consistent estimator. As $\lambda_n$ is a consistent estimate of $\lambda$, the consistency of $\hat{\sigma}_n^2(\lambda_n)$ in (3.7) can be shown as follows. As $\hat{\sigma}_n^2(\lambda_n) = \frac{1}{n}V_n'n_n(\lambda_n)S_n(\lambda_n)Y_n$, by expansion

$$
\hat{\sigma}_n^2(\lambda_n) = \frac{1}{n}V_n'n_n + 2(\lambda_0 - \lambda_n)\frac{1}{n}Y_n'n_n'W_n'n_nY_n + (\lambda_0 - \lambda_n)^2\frac{1}{n}Y_n'n_n'W_n'n_nW_{n,n}Y_n = \frac{1}{n}V_n'n_n + o_P(1),
$$

because $(\lambda_n - \lambda_0) = o_P(1)$, $\frac{1}{n}Y_n'n_n'W_n'n_nY_n = O_P(\frac{1}{n})$ and $\frac{1}{n}Y_n'n_n'W_n'n_nW_{n,n}Y_n = O_P(\frac{1}{n})$ by Lemma A.14. The consistency of $\hat{\sigma}_n^2(\lambda_n)$ follows as $\frac{1}{n}V_n'n_n + o_P(1)$ by the law of large number for i.i.d. variables.

Assumption 6 is essentially a local identification condition. For global identification, one has to impose identification conditions outside of $\Lambda$. For global identification, the true parameter $\lambda_0$ needs to be the unique global maximizer in the limiting process of $\frac{h_n}{n}(Q_n(\lambda) - Q_n(\lambda_0))$ on its parameter space. For each $n$, $Q_n(\lambda)$ may have a unique maximum at $\lambda_0$ if the variance of $Y_n$ is unique at $(\lambda_0, \sigma_0^2)$, i.e., whenever $(\lambda, \sigma^2) \neq (\lambda_0, \sigma_0^2)$, $\sigma^2S_n^{-1}(\lambda)S_n^{-1}(\lambda) \neq \sigma_0^2S_n^{-1}S_n^{-1}$. By Jensen’s inequality, $E(\ln L(\lambda, \sigma^2)) \leq E(\ln L(\lambda_0, \sigma_0^2))$ for all $(\lambda, \sigma^2)$, and, hence, $Q_n(\lambda) \leq Q_n(\lambda_0)$. When the variance of $Y_n$ is unique at $(\lambda_0, \sigma_0^2)$, as the ratio of the normal densities with different variances cannot be a constant, Jensen’s inequality applied to the logarithmic of the ratio of normal densities will become a strict inequality and one has $E(\ln L(\lambda, \sigma^2)) < E(\ln L(\lambda_0, \sigma_0^2))$ for all $(\lambda, \sigma^2) \neq (\lambda_0, \sigma_0^2)$. Under such a circumstance, $Q_n(\lambda) - Q_n(\lambda_0) < 0$ whenever $\lambda \neq \lambda_0$. However, this is not sufficient to guarantee the identification uniqueness of $\lambda_0$ in the limiting process of $\frac{h_n}{n}(Q_n(\lambda) - Q_n(\lambda_0))$ as
goes to infinity (see, e.g., Amemiya (1985), p.116). Assumption 6 is not sufficient for global identification. For example, Assumption 6 does not rule out cases with \( W_{n,n} \) being skew-symmetric. It is apparent that if \( W_{n,n} \) were skew-symmetric, the variance matrix of \( Y_n \) would be 
\[
\sigma_0^2 S_n^{-1} S_n^{-1} = \sigma_0^2 (I_n + \lambda_0^2 W_n W_n). 
\]
As only \( \lambda_0^2 \) would appear in the variance matrix, the values of \( \lambda = \lambda_0 \) and \( \lambda = -\lambda_0 \) would generate the same variance matrix of \( Y_n \) and the true parameter of \( \lambda \) could not be globally identified.

\( \lambda_0 \) would be globally identifiable if for any \( \lambda \neq \lambda_0 \),
\[
\lim_{n \to \infty} \left( \frac{h_n}{n} \ln |\sigma_0^2 S_n^{-1} S_n^{-1}| - \frac{h_n}{n} \ln |\sigma_0^2 (\lambda) S_n^{-1} (\lambda) S_n^{-1} (\lambda)| \right) \neq 0,
\]
because it implies that
\[
\lim_{n \to \infty} \frac{h_n}{n} \left[ \ln |S_n(\lambda)| - \ln |S_n(\lambda_0)| - \frac{n}{2} (\ln \sigma_n^2 (\lambda) - \ln \sigma_n^2 (\lambda_0)) \right] \neq 0 \quad \text{for any } \lambda \neq \lambda_0.
\]
A direct verification of the global identification of \( \lambda_0 \) would depend on the specific sequence of weights matrices \( \{W_{n,n}\} \) in a model. Simple and intuitive global identification conditions are available for some important weights matrices when \( \lim_{n \to \infty} h_n = \infty \). These cases include the spatial weights matrix in Case (1991, 1992) which is symmetric, and the familiar case of Ord (1975) where \( W_{n,n} \) is row-normalized and the eigenvalues of \( \{W_{n,n}\} \) are real, in the following theorem.

**Theorem 4.** For the case with \( \lim_{n \to \infty} h_n = \infty \), if the weights matrix \( W_{n,n} \) is symmetric or \( W_{n,n} = \Lambda_n^{-1} D_n \) where \( \Lambda_n \) is a diagonal matrix and \( D_n \) is a symmetric matrix, then, for large enough \( n \),
\[
\frac{h_n}{n} (Q_n(\lambda) - Q_n(\lambda_0))
\]
is concave on any bounded set of \( \lambda \) and \( \lambda_0 \) is its unique global maximizer. Hence, \( \lambda_0 \) is uniquely identifiable on any bounded parameter space. The estimator \( \hat{\lambda}_n \) derived from the maximization of \( \ln L_n(\lambda) \) over any bounded parameter space of \( \lambda \) is consistent.

For cases where global identification for the QMLE is hard to check, an alternative strategy is to design a consistent estimation method under some regularity conditions and, hopefully, easily verifiable identification conditions (e.g., a method of moments) to obtain an initial consistent estimate \( \tilde{\lambda}_n \) of \( \lambda \). As \( \tilde{\lambda}_n \) converges to \( \lambda_0 \) in probability, \( \tilde{\lambda}_n \) will lie in any neighborhood \( \Lambda \) of \( \lambda_0 \) with probability close to 1 for large enough \( n \) and the maximization of \( \ln L_n(\lambda) \) can start with the initial consistent estimate. Appendix C provides an example of moment estimator. It is shown that one step Newton-Raphson iteration starting with an initial consistent estimator \( \hat{\lambda}_n \) will provide a second round estimator of \( \lambda \), which has the same asymptotic distribution of the QMLE.

4. Asymptotic Distributions of the MLEs
As \( Y'_{n}W_{n,n}Y_{n} = Y'_{n}W'_{n,n}Y_{n} \), it follows from (3.5) that \( \frac{\partial \sigma^2(\lambda)}{\partial \lambda} = -\frac{2}{n} Y'_{n}W'_{n,n}S(\lambda)Y_{n} \). Therefore,

\[
\frac{\partial \ln L_n(\lambda)}{\partial \lambda} = \frac{1}{\sigma^2_n(\lambda)} Y'_{n}W'_{n,n}S(\lambda)Y_{n} - tr(W_{n,n}S^{-1}(\lambda)).
\] (4.1)

At the true parameter \( \lambda_0 \), \( \hat{\sigma}^2_n(\lambda_0) = \frac{1}{n}V'_{n}V_{n} \) and

\[
\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} = \frac{1}{\hat{\sigma}^2_n(\lambda_0)} V'_{n}G'_{n}V_{n} - tr(G_{n}).
\] (4.2)

Because \( \frac{\partial S^{-1}(\lambda)}{\partial \lambda} = S^{-1}(\lambda)W_{n,n}S^{-1}(\lambda) \) (Dhrymes 1978, p.120, Corollary 38),

\[
\frac{\partial^2 \ln L_n(\lambda)}{\partial \lambda^2} = \frac{2}{n\hat{\sigma}^4_n(\lambda)} (V'_{n}G'_{n}V_{n})^2 - \frac{1}{\hat{\sigma}^2_n(\lambda)} V'_{n}G'_{n}G_{n}V_{n} - tr(G_{n}^2).
\] (4.3)

At \( \lambda_0 \),

\[
\frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2} = \frac{2}{n\hat{\sigma}^4_n(\lambda_0)} (V'_{n}G'_{n}V_{n})^2 - \frac{1}{\hat{\sigma}^2_n(\lambda_0)} V'_{n}G'_{n}G_{n}V_{n} - tr(G_{n}^2).
\] (4.4)

As \( \hat{\lambda}_n \) satisfies the first-order condition that \( \frac{\partial \ln L_n(\hat{\lambda}_n)}{\partial \lambda} = 0 \), by the mean-value theorem

\[
\hat{\lambda}_n - \lambda_0 = -\left( \frac{\partial^2 \ln L_n(\hat{\lambda}_n)}{\partial \lambda^2} \right)^{-1} \frac{\partial \ln L_n(\hat{\lambda}_n)}{\partial \lambda},
\] (4.5)

where \( \hat{\lambda}_n \) lies between \( \hat{\lambda}_n \) and \( \lambda_0 \). To derive the asymptotic distribution of the QMLE \( \hat{\lambda}_n \), we investigate the rate of convergence of the first and second derivatives of the log likelihood function in (4.2) and (4.4). For the second derivative, only its probability limit is needed. For the first derivative, we need to investigate its limiting distribution also.

Define the function

\[
P_n(\lambda_0) = \frac{2}{n\sigma^4_0} (V'_{n}G'_{n}V_{n})^2 - \frac{1}{\sigma^2_0} V'_{n}G'_{n}G_{n}V_{n} - tr(G_{n}^2).
\] (4.6)

This function differs from \( \frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2} \) in that \( \hat{\sigma}^2_n(\lambda_0) \) of the latter is replaced by the true parameter \( \sigma^2_0 \).

**Proposition 1.** Under the regularity conditions of Assumptions 1-5,

\[
E(P_n(\lambda_0)) = \frac{2}{n} tr^2(G_{n}) - tr(G_{n}G'_{n}) - tr(G_{n}^2) + O\left(\frac{1}{h_n}\right)
\]

and

\[
\frac{h_n}{n} P_n(\lambda_0) \xrightarrow{p} \lim_{n \to \infty} \frac{h_n}{n} \left( \frac{2}{n} tr^2(G_{n}) - tr(G_{n}G'_{n}) - tr(G_{n}^2) \right).
\] (4.7)

In particular, if \( \lim_{n \to \infty} h_n = \infty \), then

\[
\frac{h_n}{n} P_n(\lambda_0) \xrightarrow{p} - \lim_{n \to \infty} \frac{h_n}{n} \left[ tr(G_{n}G'_{n}) + tr(G_{n}^2) \right].
\]

12
Proposition 2. Under the regularity conditions of Assumptions 1-5,

\[ \frac{h_n}{n} \left[ \frac{\partial^2 \ln L_n(\hat{\lambda}_n)}{\partial \lambda^2} - P_n(\lambda_0) \right] \xrightarrow{p} 0, \]

for any \( \hat{\lambda}_n \) which converges in probability to \( \lambda_0 \).

An immediate consequence of Propositions 1 and 2 is that

Proposition 3. Under the assumed regularity conditions,

\[ -\frac{h_n}{n} \frac{\partial^2 \ln L_n(\hat{\lambda}_n)}{\partial \lambda^2} \xrightarrow{p} \lim_{n \to \infty} \frac{h_n}{n} [tr(G_n G'_n) + tr(G^2_n) - \frac{2}{n} tr^2(G_n)]. \tag{4.8} \]

In particular, if \( \lim_{n \to \infty} h_n = \infty \), then

\[ -\frac{h_n}{n} \frac{\partial^2 \ln L_n(\hat{\lambda}_n)}{\partial \lambda^2} \xrightarrow{p} \lim_{n \to \infty} \frac{h_n}{n} [tr(G_n G'_n) + tr(G^2_n)]. \]

We note that \( tr(G_n G'_n) + tr(G^2_n) - \frac{2}{n} tr^2(G_n) = \frac{1}{4} tr[(C'_n + C_n)(C'_n + C_n)] \)

The asymptotic distribution of this first-order derivative will depend on the asymptotic distribution of the random quadratic form \( Q_n \). The mean of \( Q_n \) is zero because \( tr(C_n) = 0 \) and its variance is \( \sigma^2_{Q_n} = (\mu_4 - 3\sigma^4_0) \sum_{i=1}^n C^2_{n,ii} + \sigma^4_0 [tr(C_n C'_n) + tr(C^2_n)] \) by Lemma A.12. This random form \( Q_n \), after being normalized, may converge in distribution to a standard normal variable. Limiting distributions of quadratic random forms have been obtained in Whittle (1964), Beran (1972), Sen (1976), Giraitis and Taqqu (1998), and Kelejian and Prucha (1999). The original theorem of Kelejian and Prucha (1999) can be slightly modified to take into account the possible lower than \( \sqrt{n} \)-rate convergence of our \( Q_n \) function. For this purpose, Assumption 3 needs to be slightly strengthened.

Assumption 3'. \( \frac{h_n^{1+\eta}}{n} \rightarrow 0 \) for some \( \eta > 0 \) tends to zero as \( n \) goes to infinity.

Proposition 4. Under the Assumptions 1-2, 3' and 4-6, \( \frac{Q_n}{\sigma_{Q_n}} \xrightarrow{D} N(0, 1) \).

Assumption 7. The limit of \( \frac{h_n}{n} \sum_{i=1}^n C^2_{n,ii} \) exists.

As \( tr(C_n C'_n) + tr(C^2_n) = tr(G_n G'_n) + tr(G^2_n) - \frac{2}{n} tr^2(G_n) = O(\frac{h_n}{n}) \), a consequence of Proposition 4 with the added Assumption 7 is that \( \sqrt{\frac{h_n}{n} \frac{Q_n}{\sigma_{Q_n}}} \xrightarrow{D} N(0, \lim_{n \to \infty} \frac{h_n}{n} [\frac{\mu_4 - 3\sigma^4_0}{\sigma^4_0} \sum_{i=1}^n C^2_{n,ii} + tr(C_n C'_n) + tr(C^2_n)] \).

Because \( \sigma_{Q_n} \) is i.i.d. and \( \sigma^2_{Q_n}(\lambda_0) = \frac{1}{n} \sum_{i=1}^n \sigma_1^2 \), the law of large numbers for i.i.d. random variables implies that \( \sigma^2_{Q_n}(\lambda_0) \xrightarrow{p} \sigma^2_0 \). Consequently, the Slutsky lemma implies the following result.

Proposition 5. Under the assumed regularity conditions,

\[ \sqrt{\frac{h_n}{n} \frac{\partial \ln L_n(\lambda_0)}{\partial \lambda}} = \frac{\sigma^2_0}{\sigma^2_{n}(\lambda_0)} \cdot \sqrt{\frac{h_n}{n} \frac{Q_n}{\sigma_{Q_n}}} \xrightarrow{D} N(0, \lim_{n \to \infty} \frac{h_n}{n} [\frac{\mu_4 - 3\sigma^4_0}{\sigma^4_0} \sum_{i=1}^n C^2_{n,ii} + tr(C_n C'_n) + tr(C^2_n)]). \tag{4.9} \]
For the cases that $v_i$’s are normally distributed or $\lim_{n \to \infty} h_n = \infty$,

$$
\sqrt{\frac{h_n}{n}} \frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} \rightarrow N(0, \lim_{n \to \infty} \frac{h_n}{n}[tr(C_nC_n') + tr(C_n^2)]).
$$

The rate of convergence of the QMLE $\hat{\lambda}_n$ depends on $n$ and $h_n$. The usual $\sqrt{n}$-rate of convergence is achievable only when the sequence $\{h_n\}$ is bounded. That is the case when each unit might be influenced by a few neighboring units even when the total number of cross-sectional units is large. Otherwise, the rate of convergence may be slower than the $\sqrt{n}$-rate. Anyway, the proper normalization rate for $\hat{\lambda}_n$ shall be the square root of $\frac{n}{h_n}$. The’ effective’ sample size in that example is the number of districts $r$. Even though the QMLE of $\lambda$ might converge at a slower rate, the QMLE of $\sigma^2$ is convergent at the usual $\sqrt{n}$-rate. This is shown in the subsequent theorem. Denote $\Sigma^{**}_{\lambda \lambda} = \Sigma_{\lambda \lambda} + \Sigma_{\lambda \sigma} \Omega \Sigma_{\sigma \sigma}$.

**Theorem 6.** Under the assumed regularity conditions,

$$
\sqrt{\frac{n}{h_n}} (\hat{\lambda}_n - \lambda_0) \overset{D}{\rightarrow} N(0, \Sigma),
$$

where $\Sigma = \begin{pmatrix} \Sigma^{**}_{\lambda \lambda} & \Sigma^{**}_{\lambda \sigma} \\ \Sigma^{**}_{\sigma \lambda} & \Sigma^{**}_{\sigma \sigma} \end{pmatrix}$ with $\Sigma_{\sigma \lambda} = \psi \Sigma^{**}_{\lambda \lambda}$, $\Sigma_{\sigma \sigma} = 4\psi^2 \Sigma^{**}_{\lambda \lambda} + (\mu_4 - \sigma_4^4)$, $\psi = \lim_{n \to \infty} \sqrt{n} \frac{\psi}{h_n} tr(G_n)$, and $\mu_4 = E(v^4)$. 

Theorem 5. Under the assumed regularity conditions,

$$
\sqrt{\frac{n}{h_n}} (\hat{\lambda}_n - \lambda_0) \overset{D}{\rightarrow} N(0, \Sigma_{\lambda \lambda} + \Sigma_{\lambda \sigma} \Omega \Sigma_{\sigma \sigma}),
$$

where $\Sigma_{\lambda \lambda} = \left(\lim_{n \to \infty} \frac{h_n}{n} [tr(C_nC_n') + tr(C_n^2)]\right)^{-1}$ and $\Omega = \left(\frac{\mu_4 - 3\sigma_4^4}{\sigma_4^4}\right) \lim_{n \to \infty} \frac{h_n}{n} \sum_{i=1}^n C_{n,ii}$. For the cases that $v_i$’s are normally distributed or $\lim_{n \to \infty} h_n = \infty$,

$$
\sqrt{\frac{n}{h_n}} (\hat{\lambda}_n - \lambda_0) \overset{D}{\rightarrow} N(0, \Sigma_{\lambda \lambda}).
$$
For the case that \( \lim_{n \to \infty} h_n = \infty \), \( \Sigma_{\lambda\lambda}^* = \Sigma_{\lambda\lambda}, \Sigma_{\lambda\sigma} = 0 \) and \( \Sigma_{\sigma\sigma} = (\mu_4 - \sigma_4^2) \).

From Theorem 6, we see that even though the possible slower rate of convergence of \( \hat{\lambda}_n \) does not reduce the usual \( \sqrt{n} \)-rate of convergence of the QMLE \( \hat{\sigma}_n^2(\hat{\lambda}_n) \) of \( \sigma^2 \), the QMLE \( \hat{\lambda}_n \) may have effect on the asymptotic variance of the latter unless \( \psi \) was zero. \( \psi \) will not vanish in general if \( \{h_n\} \) is bounded because \( \sqrt{h_n} tr(G_n) = O(1) \). In that case, the QMLEs \( \hat{\lambda}_n \) and \( \hat{\sigma}_n^2(\hat{\lambda}_n) \) (after proper normalization on their rates of convergence) can be correlated. For the case that \( \lim_{n \to \infty} h_n = \infty \), \( \psi = 0 \) and \( \hat{\lambda}_n \) and \( \hat{\sigma}_n^2(\hat{\lambda}_n) \) become asymptotically independent as \( \Sigma_{\lambda\sigma} = 0 \). These properties are summarized in the following theorem.

**Theorem 7.** Under the assumed regularity conditions, if \( \lim_{n \to \infty} h_n = \infty \), then

\[
\sqrt{n}(\hat{\sigma}_n^2(\hat{\lambda}_n) - \sigma_0^2) \xrightarrow{D} N(0, \mu_4 - \sigma_0^4)
\]

and \( \sqrt{n}(\hat{\sigma}_n^2(\hat{\lambda}_n) - \sigma_0^2) \) is asymptotically independent of \( \sqrt{h_n}(\hat{\lambda}_n - \lambda_0) \).

The QMLE \( \hat{\lambda}_n \) can be derived by maximizing the concentrated log likelihood function (3.6). Its computation can be simplified if an optimization subroutine can start its iterations with a good initial estimator. For the latter, computationally simple estimators based on the method of moments have been suggested by Kelejian and Prucha (1999a). In Appendix C, related moments estimators are described for the autoregressive process under our framework. It is shown that those moment estimators are \( \sqrt{h_n} \)-consistent and are asymptotically normal. With these consistent estimators, a computationally simpler second round (one-step) estimator of \( \lambda \), which has the same asymptotic distribution as the QMLE (MLE) \( \hat{\lambda}_n \), is also available.

The moments estimators \( \tilde{\lambda}_{n1} \) or \( \tilde{\lambda}_{n2} \) can be used as an initial consistent estimator for the derivation of the QMLE \( \hat{\lambda}_n \). A computationally simpler second-round (one-step) estimator can also be derived. Let \( \tilde{\lambda}_n \) be an initial consistent estimator. A second-round estimator can be defined from Newton’s algorithm based on the concentrated log likelihood function in (3.6):

\[
\tilde{\lambda}_{ns} = \tilde{\lambda}_n - \left( \frac{\partial^2 \ln L_n(\tilde{\lambda}_n)}{\partial \lambda^2} \right)^{-1} \frac{\partial \ln L_n(\tilde{\lambda}_n)}{\partial \lambda}.
\]

As \( \tilde{\lambda}_n \) is a \( \sqrt{h_n} \)-consistent estimator, the second-round estimator can have the same limiting distribution as the QMLE \( \hat{\lambda}_n \). This is so as follows. By the Taylor expansion,

\[
\frac{\partial \ln L_n(\tilde{\lambda}_n)}{\partial \lambda} = \frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} + \frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2}(\tilde{\lambda}_n - \lambda_0),
\]

15
where \( \bar{\lambda}_n \) lies between \( \lambda_n \) and \( \lambda_0 \). Hence,

\[
\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = \sqrt{\frac{n}{h_n}}(\lambda_n - \lambda_0) - \left( \frac{h_n}{n} \frac{\partial^2 \ln L_n(\hat{\lambda}_n)}{\partial \lambda^2} \right)^{-1} \left( \frac{h_n}{n} \frac{\partial^2 \ln L_n(\hat{\lambda}_n)}{\partial \lambda^2} \right) - \left( \frac{h_n}{n} \frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2} \right)^{-1} \left( \frac{h_n}{n} \frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2} \right)
\]

\[
= \left( I_n - \left( \frac{h_n}{n} \frac{\partial^2 \ln L_n(\bar{\lambda}_n)}{\partial \lambda^2} \right)^{-1} \left( \frac{h_n}{n} \frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2} \right) \right) \sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) - \left( \frac{h_n}{n} \frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2} \right)^{-1} \left( \frac{h_n}{n} \frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2} \right)
\]

\[
= \left( - \frac{h_n}{n} \frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2} \right)^{-1} \left( \frac{h_n}{n} \frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2} \right) \sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) + o_P(1),
\]

by Propositions 1 and 2. This shows that the second-round estimator has the same asymptotic distribution as \( \hat{\lambda}_n \) from (3.5).

The possible slower rate of convergence of \( \hat{\lambda}_n \) in Theorem 5 implies that, for statistical inference, one shall take into account the factor \( h_n \) in addition to the sample size \( n \). However, some practical formulas for classical inference statistics are valid when the disturbances are normally distributed. For example, the ‘t’ statistic formula for testing \( \lambda \) as a specific constant, say \( \lambda_c \), is asymptotically valid. Let \( \hat{\omega}^2_n = -\left( \frac{\partial^2 \ln L_n(\hat{\lambda}_n)}{\partial \lambda^2} \right)^{-1} \). This statistic is also valid 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

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

under the null hypothesis. In addition to the Wald-type statistic, the conventional likelihood ratio test statistic is also valid for testing \( \lambda_0 = \lambda_c \). This is so, because

\[
2[\ln L_n(\hat{\lambda}_n) - \ln L_n(\lambda_c)] = - \frac{\partial^2 \ln L_n(\lambda_c)}{\partial \lambda^2} (\hat{\lambda}_n - \lambda_0)^2
\]

\[
= \sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0)\Sigma_{\lambda\lambda}^{-1} \sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) + o_P(1) \xrightarrow{D} \chi^2(1),
\]

which is asymptotically chi-square distributed with one degree of freedom under the null hypothesis. For testing the null hypothesis \( \lambda_0 = \lambda_c \), Rao’s efficient score test statistic based on the concentrated log likelihood function is also valid. The efficient score statistic is \( \frac{\partial \ln L_n(\lambda_c)}{\partial \lambda} \left( - \frac{\partial^2 \ln L_n(\lambda_c)}{\partial \lambda^2} \right)^{-1} \frac{\partial \ln L_n(\lambda_c)}{\partial \lambda} \). This statistic is asymptotically chi-square distributed by Proposition 5 and because \( - \frac{h_n}{n} \frac{\partial^2 \ln L_n(\lambda_c)}{\partial \lambda^2} \) is a consistent estimate of the limiting variance of \( \sqrt{\frac{h_n}{n} \frac{\partial \ln L_n(\lambda_c)}{\partial \lambda}} \) under the null hypothesis. From our results, we note that these statistics based on the concentrated likelihood are robust as they are asymptotically valid, even when \( \{v_i\} \) are not normally distributed, as long as either \( \lim_{n \to \infty} h_n = \infty \) or \( \mu_4 = 4\sigma^3 \) holds.
Theorem 5 indicates that the rate of convergence of \( \hat{\lambda}_n \) to \( \lambda_0 \) can be very slow if the rate \( h_n \) is only slightly slower than the \( n \)-rate. Consequently, one may suspect that for cases with \( h_n = n \), the QMLE \( \hat{\lambda}_n \) could be inconsistent. The following example illustrates this consequence. For simplicity, consider the process (2.1) with a known \( \sigma^2 = 1 \) and the spatial weights matrix \( W_{n,n} \) with \( w_{n,ij} = \frac{1}{n-1} \) for all \( i \neq j \). The column and row sums of such weights matrices are uniformly bounded. As \( W_{n,n} = \frac{1}{n^2} (I_n l_n' - I_n) \), \( S^{-1}_n(\lambda) = \frac{1}{(1+\lambda)}(I_n + \frac{\lambda l_n'}{n-1}) \). It is apparent that the column and row sums of \( S^{-1}_n(\lambda) \) are uniformly bounded as long as \( \lambda \) is bounded and is bounded away from 1. With this weights matrix, one has

\[
W_{n,n} S^{-1}_n(\lambda) = \frac{1}{(n-1+\lambda)}(I_n + \frac{l_n'}{1-\lambda} - I_n),
\]

\[
(W_{n,n} S^{-1}_n(\lambda))^2 = \frac{1}{(n-1+\lambda)^2} \left( \frac{n-2(1-\lambda)}{(1-\lambda)^2} l_n' + I_n \right),
\]

\[
tr[W_{n,n} S^{-1}_n(\lambda)] = \frac{\lambda}{1-\lambda}(1 - \frac{1-\lambda}{n})^{-1},
\]

and

\[
tr[(W_{n,n} S^{-1}_n(\lambda))^2] = \left( \frac{n}{n-1+\lambda} \right)^2 \left( \frac{1-2(1-\lambda)/n}{(1-\lambda)^2} + \frac{1}{n} \right).
\]

Because \( W_{n,n} \) is symmetric, its first-order derivative of the concentrated log likelihood with respect to \( \lambda \) at \( \lambda_0 \) is

\[
\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} = -tr(W_{n,n} S^{-1}_n(\lambda)) + V_n' S^{-1}_n W_{n,n} V_n
\]

\[
= -\frac{\lambda_0}{1-\lambda_0} \left( 1 - \frac{1-\lambda_0}{n} \right)^{-1} + \frac{1}{n-1+\lambda_0} V_n' \left( \frac{l_n'}{1-\lambda_0} - I_n \right) V_n.
\]

As \( \frac{1}{n} V_n' \left( \frac{l_n'}{1-\lambda_0} - I_n \right) V_n - \lambda_0 = \frac{1}{1-\lambda_0} \left( \sum_{i=1}^{n} \frac{v_i}{n} \right)^2 - 1 + (1 - \sum_{i=1}^{n} \frac{v_i^2}{n}) \) converges in probability to \( \xi^{-1} \frac{\xi - 1}{1-\lambda_0} \) where \( \xi \) is a \( \chi^2(1) \) random variable with one degree of freedom, it follows that

\[
\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} \xrightarrow[\mathbb{P}]{} \frac{\xi - 1}{1-\lambda_0}.
\]

The second-order derivative is

\[
\frac{\partial^2 \ln L_n(\lambda)}{\partial \lambda^2} = -tr[(W_{n,n} S^{-1}_n(\lambda))^2] - V_n' (W_{n,n} S^{-1}_n)^2 V_n
\]

\[
= \left( \frac{n}{n-1+\lambda} \right)^2 \left( 1 \frac{2(1-\lambda)}{(1-\lambda)^2} + \frac{1}{n} \right) \left( 1 \frac{2(1-\lambda)}{(1-\lambda)^2} l_n' + I_n \right) V_n.
\]

The QMLE \( \hat{\lambda}_n \) satisfies the first-order condition that \( \frac{\partial \ln L_n(\lambda)}{\partial \lambda} = 0 \). Hence, by the mean value theorem,

\[
\hat{\lambda}_n = \lambda_0 - \left( \frac{\partial^2 \ln L_n(\lambda_0)}{\partial \lambda^2} \right)^{-1} \frac{\partial \ln L_n(\lambda_0)}{\partial \lambda},
\]
where \( \bar{\lambda}_n \) lies between \( \hat{\lambda}_n \) and \( \lambda_0 \). If \( \hat{\lambda}_n \) were a consistent estimator, \( \bar{\lambda}_n \) would converge in probability to \( \lambda_0 \). Under this situation, one has

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

Thus, for this example, if the QMLE \( \hat{\lambda}_n \) were a consistent estimator, it would imply that \( \hat{\lambda}_n - \lambda_0 \xrightarrow{D} (1 - \lambda_0)\frac{\xi - 1}{\xi + 1} \). The latter is a contradiction as \( (1 - \lambda_0)\frac{\xi - 1}{\xi + 1} \) does not have a degenerate distribution (at zero). So \( \hat{\lambda}_n \) could not be a consistent estimator of \( \lambda_0 \).

This example is revealing in terms of the asymptotic scenarios of adding new sample observations (as \( n \) increases). The above example corresponds to a sample obtained from a single district. By increasing \( n \), it means increasing spatial units in the same district. That will correspond to the notion of 'infill asymptotics' (Cressie 1993, p.101). The above example shows that the QMLE under infill asymptotics alone may not be consistent. As in Case (1991, 1992), if there are many districts from which samples are obtained, the QMLEs can be consistent if the number of districts, i.e., \( \frac{n}{m} \), increases. The scenario of increasing numbers of districts 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 rate of convergence of the QMLE under the increasing-domain asymptotics alone can be the usual \( \sqrt{n} \)-rate. But, when both infill and increasing-domain asymptotics are operating, the rates of convergence of the QMLEs for various parameters can be different and some of them may have a slower rate than the usual one.

5. Conclusions

In this paper, we have investigated the asymptotic distribution of the QMLE and MLE for spatial autoregressive processes. The familiar asymptotic properties of a QMLE or MLE such as the \( \sqrt{n} \)-consistency, where \( n \) is the sample size, and asymptotic normality are casually stated in the existing literature. Our analysis reveals that these properties depend on some features of the spatial weights matrix. Under some situations, the estimator of the spatial effect parameter in a spatial autoregressive process might converge at a slower than \( \sqrt{n} \)-rate even though the limiting distribution is normal.

A spatial weights matrix is typically row-normalized to facilitate the interpretation of the spatial interaction effect parameter. The normalization factors \( \sum_{j=1}^n d_{ij} \) for each row play an important role in the asymptotic properties of the QMLE for spatial autoregressive processes. For models with spatial weights matrices being sparse, \( \sum_{j=1}^n d_{ij} \) would remain finite and bounded as \( n \) goes to infinity. For those cases, the QMLE will converge at the usual \( \sqrt{n} \)-rate and its limiting distribution is normal. However, the rate of
convergence might be different if the sums $\sum_{j=1}^{n} d_{ij}$ do not stabilize as $n$ increases. The latter implies that influences of each unit by other spatial units are uniformly small at a certain rate of order $O(\frac{1}{n})$. The rate of convergence of the QMLE may depend on the rate of divergence $h_n$. We show that for the first-order autoregressive process, if $h_n$ tends to infinity slower than the $n$-rate, the stochastic rate of convergence of the QMLE of the spatial effect parameter is $\sqrt{\frac{n}{h_n}}$, which is lower than the $\sqrt{n}$ by the factor $h_n^{-1/2}$. From this rate, we can see that if $h_n$ were divergent at a rate equal to $n$, the QMLE would not possibly be consistent.

For the spatial scenario considered in Case (1991, 1992), the case of $h_n = O(n)$ corresponds to the notion of infill asymptotics. Our example illustrates that the consistency and asymptotic distributions of the parameter estimates are subject to the requirement of increasing-domain asymptotics. Infill asymptotics alone might not provide consistent estimates. While increasing-domain asymptotic may provide consistent estimates with the usual $\sqrt{n}$-rate of convergence, the interacting of infill and increasing-domain asymptotics may provide low rate of convergences for some parameter estimates.

We investigate possible implications of the low rate of convergence of the MLE on statistical inference. Statistical inferences with classical statistics such as the likelihood ratio, the Wald test statistics and the efficient score test statistics, the lower rate of convergence of the MLE does not invalidate conventional formulas. Empirical researchers should worry about are those where $h_n$ diverges at the rate $n$. For the latter case, as the MLE would not be consistent, a conventional test statistic might not be meaningful.

This paper has analyzed asymptotic properties for the spatial autoregressive processes. When explanatory variables are also incorporated into a spatial autoregressive process, the resulting model is known as the mixed regressive, spatial autoregressive model (Ord 1975, Anselin 1988). When explanatory variables are relevant, they play an interesting role for parameter estimation. For those models, methods of instrumental variables (IV) and, under some circumstances, even OLS method are feasible (Kelejian and Prucha 1998; Lee 1999a). However, it could be shown when all explanatory variables were irrelevant, the IV and OLS methods would not be applicable. For the latter, the asymptotic distributions of various coefficients would have various rates of convergence. We will report our findings on that model in a separate paper.
Appendix A: Some Useful Lemmas

A.1 Uniform Boundedness in Row and Column Sums

Definition: The sequences \( \{ b_{in} \} \) for \( i = 1, \ldots, n \) are of order \( h_n \) uniformly in \( i \), denoted by \( O(h_n) \), if there exists a finite constant \( c_1 \) independent of \( i \) and \( n \) such that \( |b_{in}| \leq c_1 h_n \) for all \( i = 1, \ldots, n \) and for large enough \( n \). They are bounded away from zero uniformly in \( i \) at the rate \( h_n \) if there exists a positive sequence \( \{ h_n \} \) and a constant \( c_2 > 0 \) independent of \( i \) and \( n \) such that \( c_2 h_n \leq |b_{in}| \) for all \( i = 1, \ldots, n \) for large enough \( n \). The sequences are of exact order \( h_n \) uniformly in \( i \), denoted by \( O_c(h_n) \), if they are of order \( h_n \) and are bounded away from zero at the rate \( h_n \) uniformly in \( i \).

Lemma A.1 Suppose that the spatial weights matrix \( W_{n,n} \) is a non-negative matrix with its \((i,j)\)th element being \( w_{n,ij} = \frac{d_{ij}}{\sum_{i=1}^{d} d_{ij}} \) and \( d_{ij} \geq 0 \) for all \( i, j \).

1. If the row sums \( \sum_{j=1}^{n} d_{ij} \) are uniformly bounded away from zero at the rate \( h_n \) uniformly in \( i \), and the column sums \( \sum_{i=1}^{n} d_{ij} \) are of order \( h_n \) uniformly in \( j \), then \( \{ W_{n,n} \} \) are uniformly bounded in column sums.

Proof: (1) Let \( c_1 \) and \( c_2 \) be the constants such that \( c_1 h_n \leq \sum_{j=1}^{n} d_{ij} \) for all \( i \), and \( \sum_{i=1}^{n} d_{ij} \leq c_2 h_n \) for all \( j \). It follows that \( \sum_{j=1}^{n} w_{n,ij} = \sum_{i=1}^{n} \frac{d_{ij}}{\sum_{i=1}^{d} d_{ij}} \leq \frac{1}{c_1 h_n} \sum_{i=1}^{n} d_{ij} \leq \frac{c_2}{c_1} \) for all \( i \).

2. This is a special case of (1) because \( \sum_{i=1}^{n} d_{il} = O(h_n) \) and \( \sum_{i=1}^{n} d_{ij} = \sum_{i=1}^{n} d_{ji} \) imply \( \sum_{i=1}^{n} d_{ij} = O(h_n) \). Q.E.D.

Lemma A.2 Suppose that \( \sup_{n} \| \lambda_0 W_{n,n} \|_1 < 1 \) and \( \sup_{n} \| \lambda_0 W_{n,n} \|_{\infty} < 1 \), then \( \{ S_n^{-1} \} \) is uniformly bounded in both row and column sums.

Proof: For any matrix norm \( \| \cdot \| \), \( \| \lambda_0 W_{n,n} \| < 1 \) implies that \( S_n^{-1} = \sum_{k=0}^{\infty} (\lambda_0 W_{n,n})^k \) (Horn and Johnson 1985, p.301). Let \( c = \sup_{n} \| \lambda_0 W_{n,n} \| \). Then, \( \| S_n^{-1} \| \leq \sum_{k=0}^{\infty} \| \lambda_0 W_{n,n} \|^k = \sum_{k=0}^{\infty} c^k = \frac{1}{1-c} < \infty \) for all \( n \). Q.E.D.

Lemma A.3 Suppose that \( \{ \| W_{n,n} \| \} \) and \( \{ \| S_n^{-1} \| \} \), where \( \| \cdot \| \) is a matrix norm, are bounded. Then \( \{ \| S_n(\lambda)^{-1} \| \} \), where \( S_n(\lambda) = I_n - \lambda W_{n,n} \), are uniformly bounded in a neighborhood of \( \lambda_0 \).

Proof: Let \( c \) be a constant such that \( \| W_{n,n} \| \leq c \) and \( \| S_n^{-1} \| \leq c \) for all \( n \). We note that \( S_n^{-1}(\lambda) = (S_n - (\lambda - \lambda_0)W_{n,n})^{-1} = S_n^{-1}(I_n - (\lambda - \lambda_0)G) \), where \( G_n = W_{n,n} S_n^{-1} \). By the submultiplicative property of a matrix norm, \( \| G_n \| \leq \| W_{n,n} \| \cdot \| S_n^{-1} \| \leq c^2 \) for all \( n \).
Let $B_1(\lambda_0) = \{ \lambda : |\lambda - \lambda_0| < 1/c^2 \}$. It follows that, for any $\lambda \in B_1(\lambda_0)$, $\| (\lambda - \lambda_0)G_n \| \leq |\lambda - \lambda_0| \cdot \| G_n \| < 1$. As $\| (\lambda - \lambda_0)G_n \| < 1$, it follows that $I_n - (\lambda - \lambda_0)G_n$ is invertible and $(I_n - (\lambda - \lambda_0)G_n)^{-1} = \sum_{k=0}^{\infty} (\lambda - \lambda_0)^k G_n^k$. Therefore,

$$
\| (I_n - (\lambda - \lambda_0)G_n)^{-1} \| \leq \sum_{k=0}^{\infty} |\lambda - \lambda_0|^k \cdot \| G_n \|^k \leq \sum_{k=0}^{\infty} |\lambda - \lambda_0|^k e^{2k} = \frac{1}{1 - |\lambda - \lambda_0|e^2} < \infty
$$

for any $\lambda \in B_1(\lambda_0)$. The result follows by taking a close neighborhood $B(\lambda_0)$ contained in $B_1(\lambda_0)$. In $B(\lambda_0)$, $\sup_{\lambda \in B(\lambda_0)} |\lambda - \lambda_0|^2 < 1$, and, hence,

$$
\sup_{\lambda \in B(\lambda_0)} \| S_n^{-1}(\lambda) \| \leq \| S_n^{-1} \| \cdot \sup_{\lambda \in B(\lambda_0)} \| (I_n - (\lambda - \lambda_0)G_n)^{-1} \| \leq \sup_{\lambda \in B(\lambda_0)} \frac{c}{1 - |\lambda - \lambda_0|e^2} < \infty.
$$

Q.E.D.

**Lemma A.4** Suppose that $\| W_{n,n} \| \leq 1$ for all $n$, where $\| \cdot \|$ is a matrix norm, then $\{ \| S_n(\lambda)^{-1} \| \}$, where $S_n(\lambda) = I_n - \lambda W_{n,n}$, are uniformly bounded in any closed subset of $(-1,1)$.

Proof: For any $\lambda \in (-1,1)$, $\| \lambda W_{n,n} \| \leq |\lambda| \cdot \| W_{n,n} \| < 1$ and, hence, $S_n^{-1}(\lambda) = \sum_{k=0}^{\infty} \lambda^k W_{n,n}^k$. It follows that, for any $|\lambda| < 1$, $\| S_n^{-1}(\lambda) \| \leq \sum_{k=0}^{\infty} |\lambda|^k \cdot \| W_{n,n} \|^k \leq \sum_{k=0}^{\infty} |\lambda|^k = \frac{1}{1 - |\lambda|}$. Hence, for any closed subset $B$ of $(-1,1)$, $\sup_{\lambda \in B} \| S_n^{-1}(\lambda) \| \leq \sup_{\lambda \in B} \frac{1}{1 - |\lambda|} < \infty$. Q.E.D.

**A.2 Orders of Some Relevant Quantities**

In this subsection, the assumption that the elements $w_{n,ij}$ of the weights matrix $W_{n,n}$ are of order $O(\frac{1}{n^2})$ uniformly in $i,j$ will be maintained. In general, $W_{n,n}$ may not be row-normalized.

**Lemma A.5** Suppose that the elements of the sequences of vectors $P_n = (p_{n1}, \ldots, p_{nn})'$ and $Q_n = (q_{n1}, \ldots, q_{nn})'$ are uniformly bounded for all $n$.

1) If $\{ A_n \}$ are uniformly bounded in either row or column sums, then $|P_n' A_n Q_n| = O(n)$.

2) If the row sums of $\{ A_n \}$ and $\{ Z_n \}$ are uniformly bounded, $|e_{ni}' A_n P_n| = O(1)$ and $|z_{ni} A_n P_n| = O(1)$ uniformly in $i$, where $e_{ni}$ is the $i$th unit column vector of dimension $n$, and $z_{ni}$ is the $i$th row of $Z_n$.

3) If both the row and column sums of $\{ A_n \}$ are uniformly bounded and $V_n = (v_1, \ldots, v_n)'$ where $v_i$ are $i.i.d.$ with zero mean and finite variance $\sigma^2$, then $\frac{1}{n} P_n' A_n V_n = o_p(1)$.

Proof: Let constants $c_1$ and $c_2$ such that $|p_{ni}| \leq c_1$ and $|q_{ni}| \leq c_2$. For 1), there exists a constant such that $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{n,ij}| \leq c_3$. Hence, $|P_n' A_n Q_n| = |\sum_{i=1}^{n} \sum_{j=1}^{n} a_{n,ij} p_{ni} q_{nj}| \leq c_1 c_2 \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{n,ij}| \leq n c_1 c_2 c_3$. For 2), let $c_4$ be a constant such that $\sum_{j=1}^{n} |a_{n,ij}| \leq c_4$ for all $n$ and $i$. It follows that $|e_{ni}' A_n P_n| = |\sum_{j=1}^{n} a_{n,ij} p_{nj}| \leq c_1 \sum_{i=1}^{n} |a_{n,ij}| \leq c_1 c_4$. Because $\{ Z_n \}$ is uniformly bounded in row sums, $\sum_{j=1}^{n} |z_{ni}| \leq c_2$.
for some constant $c_2$. It follows that $|z_{i,n}A_nP_n| \leq \sum_{j=1}^{n} |z_{n,i,j}| \cdot |e'_{n,j}A_nP_n|((\sum_{j=1}^{n} |z_{n,i,j}|)c_1c_4 \leq c_2c_1c_4$. For 3), $\frac{1}{n} P_n' A_n V_n$ has zero mean and its variance is $E(\frac{1}{n} P_n' A_n V_n)^2 = \sigma^2 \frac{1}{n^2} P_n' A_n A_n' P_n$. Because $A_n A_n'$ is uniformly bounded in both row and column sums, $\frac{1}{n} P_n' A_n A_n' P_n = O(\frac{1}{n})$ from 1). The result 3) follows from Chebyshev’s inequality. Q.E.D.

**Lemma A.6** Suppose that the column sums of $\{A_n\}$ are uniformly bounded. Then, $|w_{i,n} A_n b_n| = O\left(\frac{k}{n} \right)$ and $|w'_{n,i} A_n b_n| = O\left(\frac{k}{n} \right)$ uniformly in $i$ for any column vector $b_n = (b_{n1}, \ldots, b_{nn})'$ such that $\sum_{i=1}^{n} |b_{ni}| = O(k)$. Proof: Let $w_{n,ij} = \frac{c_{n,ij}}{n}$. Because $w_{n,ij} = O\left(\frac{1}{n} \right)$ uniformly in $i$ and $j$, there exists a constant $c_1$ such that $c_{n,ij} \leq c_1$ for all $i, j$ and $n$. Because $\{A_n\}$ is uniformly bounded in column sums, $\max_{1 \leq i \leq n} \sum_{k=1}^{n} |a_{ik}| \leq c_2$ for all $n$. It follows that

$$|w_{i,n} A_n b_n| \leq \frac{1}{n} \left( \sum_{k=1}^{n} c_{n,ik} a_{n,k1}, \ldots, \sum_{k=1}^{n} c_{n,ik} a_{n,kn} \right) b_n| \leq \frac{c_1 c_2}{n} \sum_{l=1}^{n} |b_{nl}| \leq O\left(\frac{k}{n} \right).$$

Similarly, $|w'_{n,i} A_n b_n| = O\left(\frac{k}{n} \right)$. Q.E.D.

We note that the order of terms in Lemma A.6 is the at-most order. For some special cases, other orders can be sharper. For example, if $\{W_{n,n}\}$ is uniformly bounded in row sums and the elements of the sequence of vectors $A_n b_n$ are uniformly bounded for all $n$, $|w_{i,n} A_n b_n| = O(1)$ from Lemma A.5. The usefulness of the orders of Lemma A.6 depends on the order $k_n$ of $b_n$ as in the following Corollary.

**Lemma A.7** Suppose that the column sums of $\{A_n\}$ are uniformly bounded.

1) $w_{i,n} A_n e_{nj} = O\left(\frac{1}{n} \right)$ and $w'_{n,i} A_n e_{nj} = O\left(\frac{1}{n} \right)$ uniformly in $i$ and $j$, where $e_{nj}$ is the $j$th unit column vector of the $n$-dimensional Euclidean space.

2) $w_{i,n} A_n b'_{jn} = O\left(\frac{1}{n} \right)$ and $w'_{n,i} A_n b'_{jn} = O\left(\frac{1}{n} \right)$ uniformly in $i$ and $j$, where $b_{jn}$ is the $j$th row of $B_n$, when $\{B_n\}$ is uniformly bounded in row sums.

Proof: The result 1) follows because $l_n' e_{nj} = 1$ where $l_n = (1, \ldots, 1)'$. The result 2) holds because $\{B_n\}$ being uniformly bounded in row sums is equivalent to $\{B_n'\}$ being uniformly bounded in column sums. Q.E.D.

**Lemma A.8** Suppose that $\{W_{n,n}\}$ and $\{S^{-1}_{n}\}$ are uniformly bounded in row sums, and $\{A_n\}$ is uniformly bounded in column sums. Then, the matrix $H_n = S^{-1}_{n} W_{n,n} A_n$ has the properties that $e'_{nj} H_n^m e_{nj} = O\left(\frac{1}{n} \right)$ uniformly in $i$ and $j$, and $tr(H_n^m) = tr(H_n'^m) = O\left(\frac{1}{n} \right)$ for any integer $m \geq 1$. 

22
Furthermore, if \( \{W_{n,n}\}, \{S_n^{-1}\} \) and \( \{A_n\} \) are uniformly bounded in both row and column sums, then
\[
e'_{ni}(H_nH'_n)^m e_{nj} = O(\frac{1}{n}) \quad \text{uniformly in } i \text{ and } j, \quad \text{and} \quad tr[(H_nH'_n)^m] = tr[(H'_nH_n)^m] = O(\frac{1}{n}m).
\]
In addition, for \( G_n = W_{n,n}S_n^{-1}, e'_{ni}(G'_nG_n)^m e_{nj} = O(\frac{1}{n}). \)

Proof: Because \( S_n^{-1}W'_{n,n} = W'_nS_n^{-1}, \) it follows that \( H_n = W'_nS_n^{-1}A_n. \) As \( \{W_{n,n}\}, \{S_n^{-1}\} \) and \( \{A_n\} \) are all uniformly bounded in column sums, \( \{H'_n\} \) for any integer \( m \geq 1 \) are uniformly bounded in column sums. Therefore, \( e'_{ni}H_n^m e_{nj} = w'_{n,i}S_n^{-1}A_nH_n^{-1}e_{nj} = O(\frac{1}{n}) \) by Lemma A.7, and \( tr(H_n^m) = \sum_{i=1}^{n} e'_{ni}H_n^m e_{ni} = O(\frac{1}{n}). \)

When \( \{W_{n,n}\}, \{S_n^{-1}\} \) and \( \{A_n\} \) are uniformly bounded in both row and column sums, \( \{H_n\}, \{H'_n\} \) and \( \{H_nH'_n\} \) for any integer \( m \) will uniformly be bounded in column sums. Therefore,
\[
e'_{ni}(H_nH'_n)^m e_{nj} = e'_{ni}H_nH'_n(H_nH'_n)^m e_{nj} = w'_{n,i}S_n^{-1}A_nH'_n(H_nH'_n)^m e_{nj} = O(\frac{1}{n}).
\]
by Lemma A.7 and, hence, \( tr[(H_nH'_n)^m] = O(\frac{1}{n}). \)

With the trace operator,
\[
th(r[(H_nH'_n)^m]) = tr[H_nH'_n(H_nH'_n)^m] = tr[H'_n(H_nH'_n)^m] = tr[(H'_nH_n)^m].
\]
Finally, for \( G_n = W_{n,n}S_n^{-1}, \) because \( G'_n = S_n^{-1}W'_{n,n} = W'_nS_n^{-1}, \)
\[
e'_{ni}(G'_nG_n)^m e_{nj} = e'_{ni}G'_nG_n(G'_nG_n)^m e_{nj} = w'_{n,i}S_n^{-1}G_n(G'_nG_n)^m e_{nj} = O(\frac{1}{n}).
\]
Q.E.D.

**Lemma A.9** Suppose \( \{A_n\} \) are uniformly bounded either in row sums or in column sums. Then,

1) elements \( a_{n,ij} \) of \( A_n \) are uniformly bounded in \( i \) and \( j, \)

2) \( tr(A_n^m) = O(n) \) for \( m \geq 1, \) and

3) \( tr(A_nA'_n) = O(n). \)

Proof: If \( A_n \) is uniformly bounded in row sums, let \( c_1 \) be the constant such that \( \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{n,ij}| \leq c_1 \) for all \( n. \) On the other hand, if \( A_n \) is uniformly bounded in column sums, let \( c_2 \) be the constant such that \( \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{n,ij}| \leq c_2 \) for all \( n. \) Therefore, \( |a_{n,ij}| \leq \sum_{i=1}^{n} |a_{n,il}| \leq c_1 \) if \( A_n \) is uniformly bounded in row sums; otherwise, \( |a_{n,ij}| \leq \sum_{k=1}^{n} |a_{n,kj}| \leq c_2 \) if \( A_n \) is uniformly bounded in column sums. The result 1) implies immediately that \( tr(A_n) = O(n). \) If \( A_n \) is uniformly bounded in row (column) sums, then \( A_n^m \) for \( m \geq 2 \) is uniformly bounded in row (column) sums. Therefore, 1) implies also that \( tr(A_n^m) = O(n). \) Finally, as \( tr(A_nA'_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{n,ij}^2, \) it follows that \( |tr(A_nA'_n)| \leq \sum_{i=1}^{n} (\sum_{j=1}^{n} |a_{n,ij}|)^2 \leq n c_1^2 \) if \( A_n \) is uniformly bounded in row sums; otherwise \( |tr(A_nA'_n)| \leq \sum_{j=1}^{n} (\sum_{i=1}^{n} |a_{n,ij}|)^2 \leq n c_2^2. \) Q.E.D.

23
A.3 First and Second Moments of Quadratic Forms and Limiting Distribution

**Lemma A.10** Let $A_n = [a_{ij}]$ be an $n$-dimensional square matrix. Suppose that $v_1, \ldots, v_n$ are i.i.d. with zero mean, variance $\sigma^2$ and finite fourth moment $\mu_4$. Then

1) $E(V_n' A_n V_n) = \sigma^2 \text{tr}(A_n)$,

2) $E(V_n' A_n V_n)^2 = (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \sigma^4 \{ \text{tr}^2(A_n) + \text{tr}(A_n A_n') + \text{tr}(A_n^2) \}$, and

3) $\text{var}(V_n' A_n V_n) = (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \sigma^4 \{ \text{tr}(A_n A_n') + \text{tr}(A_n^2) \}$.

In particular, if $v_i$s are normally distributed, then $E(V_n' A_n V_n)^2 = \sigma^4 \{ \text{tr}^2(A_n) + \text{tr}(A_n A_n') + \text{tr}(A_n^2) \}$ and $\text{var}(V_n' A_n V_n) = \sigma^4 \{ \text{tr}(A_n A_n') + \text{tr}(A_n^2) \}$.

Proof: The result in 1) is trivial. For the second moment,

$$E(V_n' A_n V_n)^2 = E(\sum_{i=1}^n \sum_{j=1}^n a_{ij} v_i v_j)^2 = E(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij} a_{kl} v_i v_j v_k v_l).$$

Because $v_i$s are i.i.d. with zero mean, $E(v_i v_j v_k v_l)$ will not vanish only when $i = j = k = l$, $(i = j) \neq (k = l)$, $(i = k) \neq (j = l)$, and $(i = l) \neq (j = k)$. Therefore,

$$E(V_n' A_n V_n)^2 = \sum_{i=1}^n a_{ii}^2 E(v_i^2) + \sum_{i=1}^n \sum_{j \neq i} a_{ii} a_{jj} E(v_i^2 v_j^2) + \sum_{i=1}^n \sum_{j \neq i} a_{ii}^2 E(v_i^2 v_j^2) + \sum_{i=1}^n \sum_{j \neq i} a_{ij} a_{ji} E(v_i^2 v_j^2)$$

$$= (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \sigma^4 \{ \sum_{i=1}^n \sum_{j \neq i} a_{ii} a_{jj} + \sum_{i=1}^n \sum_{j=1}^n a_{ii}^2 + \sum_{i=1}^n \sum_{j=1}^n a_{ij} a_{ji} \}$$

$$= (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \sigma^4 \{ \text{tr}^2(A_n) + \text{tr}(A_n A_n') + \text{tr}(A_n^2) \}. $$

The result 3) follows from $\text{var}(V_n' A_n V_n) = E(V_n' A_n V_n)^2 - E^2(V_n' A_n V_n)$ and the results of 1) and 2). When $v_i$s are normally distributed, $\mu_4 = 3\sigma^2$. Q.E.D.

**Lemma A.11** Under the assumptions in Lemma A.10, if $\{A_n\}$ are uniformly bounded either in row sums or in column sums, then

1) $E(V_n' A_n V_n) = O(n)$,

2) $E(V_n' A_n V_n)^2 = O(n^2)$,

3) $\text{var}(V_n' A_n V_n) = O(n)$,

4) $V_n' A_n V_n = O_P(n)$, and

5) $\frac{1}{n} V_n' A_n V_n - \frac{1}{n} E(V_n' A_n V_n) = o_P(1)$.

Proof: The results 1)-3) follow immediately from Lemmas A.9 and A.10. For any $\epsilon > 0$, by Chebyshev’s inequality and 3), $P(|\frac{1}{n} V_n' A_n V_n - E(\frac{1}{n} V_n' A_n V_n)| > \epsilon) \leq \frac{\text{var}(\frac{1}{n} V_n' A_n V_n)}{\epsilon^2} = O(\frac{1}{n})$. That is, $\frac{1}{n} V_n' A_n V_n = E(\frac{1}{n} V_n' A_n V_n) + o_P(1) = O_P(1)$ by 1), and $\frac{1}{n} V_n' A_n V_n - E(\frac{1}{n} V_n' A_n V_n) = o_P(1)$ Q.E.D.
Lemma A.12 Assume that \( \{W_{n,n}\}, \{S_{n}^{-1}\} \) and \( \{A_n\} \) are uniformly bounded in both row and column sums. The \( v_1, \ldots, v_n \) are i.i.d. with zero mean and finite variance \( \sigma^2 \).

1) Let \( H_n = S_{n}^{-1}W_{n,n}A_n \). Then \( E(V_n'H_nV_n) = O(\frac{1}{n}) \), \( \text{var}(V_n'H_nV_n) = O(\frac{1}{n}) \), and \( V_n'H_nV_n = O_P(\frac{1}{n}) \). Furthermore, if \( \lim_{n \to \infty} \frac{b_n}{n} = 0 \), then \( \frac{b_n}{n}V_n'H_nV_n - \frac{b_n}{n}E(V_n'H_nV_n) = o_P(1) \).

2) With \( Y_n = S_{n}^{-1}V_n \), for any nonnegative integers \( r_1 \) and \( r_2 \) such that \( r_1 + r_2 \geq 1 \), \( E(Y_n^rW_{n,n}^rY_n^{r_2}) = O(\frac{1}{n}) \), \( \text{var}(Y_n^rW_{n,n}^rY_n^{r_2}) = O(\frac{1}{n}) \), and \( Y_n^rW_{n,n}^rY_n^{r_2} = O_P(\frac{1}{n}) \). Furthermore, \( \frac{b_n}{n}Y_n^rW_{n,n}^rY_n^{r_2} = o_P(1) \) if \( \lim_{n \to \infty} \frac{b_n}{n} = 0 \).

Proof: Lemma A.8 implies that \( E(V_n'H_nV_n) = \sigma^2 \text{tr}(H_n) = O(\frac{1}{n}) \). By Lemmas A.10 and A.8, \[
\text{var}(V_n'H_nV_n) = (\mu_4 - 3\sigma^4) \sum_{i=1}^{n} (e_{ni}H_ne_{ni})^2 + \sigma^4[\text{tr}(H_n^2) + \text{tr}(H_n^2)] = O(\frac{n}{b_n}).
\]

As \( E((V_n'H_nV_n)^2) = \text{var}(V_n'H_nV_n) + E^2(V_n'H_nV_n) = O(\frac{1}{n}) \), the generalized Chebyshev inequality implies that \( P(\frac{b_n}{n}V_n'H_nV_n \geq M) \leq \frac{1}{M^2}(\frac{b_n}{n})^2E((V_n'H_nV_n)^2) = \frac{1}{M^2}O(1) \) and, hence, \( \frac{b_n}{n}V_n'H_nV_n = O_P(1) \). Finally, because \( \text{var}(\frac{b_n}{n}V_n'H_nV_n) = O(\frac{b_n}{n}) \), the Chebyshev inequality implies that \( \frac{b_n}{n}V_n'H_nV_n - \frac{b_n}{n}E(V_n'H_nV_n) = o_P(1) \) when \( \lim_{n \to \infty} \frac{b_n}{n} = 0 \). These prove the results in 1). The results of 2) follow from 1) by letting \( H_n = S_{n}^{-1}W_{n,n}^rW_{n,n}^rS_{n}^{-1} \), as \( Y_n^rW_{n,n}^rY_n^{r_2} = V_n^rH_nV_n \). Q.E.D.

Lemma A.13 Suppose that \( \{A_n\} \) is a sequence of symmetric matrices with row and column sums uniformly bounded in absolute value. The \( v_1, \ldots, v_n \) are i.i.d. random variables with zero mean and finite variance \( \sigma^2 \), and its moment \( E(|v|^{4+2\delta}) \) for some \( \delta > 0 \) exists. Let \( \sigma^2_{Q_n} \) be the variance of \( Q_n \) where \( Q_n = V_n'A_nV_n - \sigma^2 \text{tr}(A_n) \). Assume that the variance \( \sigma^2_{Q_n} \) is bounded away from zero at the rate \( \frac{1}{n} \) and the entries \( a_{n,ij} \) of \( A_n \) are of order \( O(\frac{1}{n}) \). If \( \lim_{n \to \infty} \frac{b_n}{n} = 0 \), then \( \frac{Q_n}{\sigma_{Q_n}} \stackrel{D}{\to} N(0,1) \).

Proof: The asymptotic distribution of the quadratic random form \( Q_n \) can be established via the martingale central limit theorem. Our proof of this Lemma follows closely the original arguments in Kelejian and Prucha (1999b). In their paper, \( \sigma^2_{Q_n} \) is assumed to be bounded away from zero with the \( n \)-rate. Our subsequent arguments modify theirs to take into account the different rate of \( \sigma^2_{Q_n} \).

\( Q_n \) can be expanded into \( Q_n = \sum_{i=1}^{n} a_{n,ii}v_i^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{i-1} a_{n,ij}v_i^j - \sigma^2 \text{tr}(A_n) = \sum_{i=1}^{n} Z_{ni} \), where \( Z_{ni} = a_{n,ii}(v_i^2 - \sigma^2) + 2v_i \sum_{j=1}^{i-1} a_{n,ij}v_j \). Define \( \sigma \)-fields \( \mathcal{J}_i = \langle v_1, \ldots, v_i \rangle \) generated by \( v_1, \ldots, v_i \). Because \( v_s \) are i.i.d. with zero mean and finite variance, \( E(Z_{ni}^2) = a_{n,ii}(E(v_i^2) - \sigma^2) + 2E(v_i) \sum_{j=1}^{i-1} a_{n,ij}v_j = 0 \). The \( \{(Z_{ni}, \mathcal{J}_i) \mid 1 \leq i \leq n \} \) forms a martingale difference double array. We note that \( \sigma^2_{Q_n} = \sum_{i=1}^{n} E(Z_{ni}^2) \) as \( Z_{ni} \) are martingale differences. Also \( \frac{b_n}{n} \sigma^2_{Q_n} = O(1) \). Define the normalized variables \( Z_{ni} = Z_{ni}/\sigma_{Q_n} \).
The \{(Z_{ni}^*, J_i) | 1 \leq i \leq n, 1 \leq n \} is a martingale difference double array and \(\frac{\partial}{\partial x_n} = \sum_{i=1}^{n} Z_{ni}^*\). In order for the martingale central limit theorem to be applicable, we would show that there exists a \(\delta > 0\) such that \(\sum_{i=1}^{n} E|Z_{ni}^*|^{2+\delta}\) tends to zero as \(n\) goes to infinity. Secondly, it will be shown that \(\sum_{i=1}^{n} E(Z_{ni}^* | J_{i-1}) \leq 1\).

For any positive constants \(p\) and \(q\) such that \(\frac{1}{p} + \frac{1}{q} = 1\), \(|Z_{ni}|^q \leq (|a_{n,ii}| \cdot |v_i^2 - \sigma^2| + 2|v_i| \cdot \sum_{j=1}^{n-1} |a_{n,ij}| \cdot |v_j|^q) = (|a_{n,ii}| \cdot |v_i^2 - \sigma^2| + \sum_{j=1}^{n-1} |a_{n,ij}| \cdot |v_j|^q)^q.\) The Holder inequality for inner products applied to the last term implies that

\[
|Z_{ni}|^q \leq \left\{ \left[ \sum_{j=1}^{i} (|a_{n,ij}| \cdot |v_i^2 - \sigma^2|)^{p} \right]^{\frac{1}{p}} \left[ \sum_{j=1}^{i} (|a_{n,ij}| \cdot |v_j|^q)^{q} \right]^{\frac{1}{q}} \right\}^q = \left[ \sum_{j=1}^{i} |a_{n,ij}| \cdot |v_i^2 - \sigma^2| + \sum_{j=1}^{i} |a_{n,ij}| \cdot |v_j|^q \right].
\]

As \(\{A_n\}\) are uniformly bounded in row sums, there exists a constant \(c_1\) such that \(\sum_{i=1}^{n} |a_{n,ij}| \leq c_1\) for all \(i\) and \(n\). Hence \(|Z_{ni}|^q \leq 2c_1 \sum_{i=1}^{n} |a_{n,ij}| \cdot |v_i^2 - \sigma^2|^q + |v_j|^q \sum_{j=1}^{n-1} |a_{n,ij}| \cdot |v_j|^q\). Take \(q = 2 + \delta\). Let \(c_q > 1\) be a finite constant such that \(E(|v|^q) \leq c_q\) and \(E(|v^2 - \sigma^2|^q) \leq c_q\). Such a constant exists under the moment conditions of \(v\). It follows that \(\sum_{i=1}^{n} E|Z_{ni}|^q \leq 2c_1 \sum_{i=1}^{n} \sum_{j=1}^{i} |a_{n,ij}| = O(n)\). As \(\sum_{i=1}^{n} E|Z_{ni}|^{2+\delta} = \frac{1}{\sigma_{Q_n}} \sum_{i=1}^{n} \sum_{j=1}^{i} E|Z_{ni}|^{2+\delta} = \sigma_{Q_n}^{2+\delta} (\frac{1}{\sqrt{n}})^{\frac{1}{2}} \cdot (\frac{n}{\sqrt{n}})^{\frac{1}{2}} = O(\sqrt{n})\), \(\sum_{i=1}^{n} E|Z_{ni}|^{2+\delta} = O(\frac{1}{n^{\frac{1}{2}}})\), which goes to zero as \(n\) tends to infinity.

It remains to show that \(\sum_{i=1}^{n} E(Z_{ni}^* | J_{i-1}) \leq 0\). As \(E(Z_{ni}^* | J_{i-1}) = (\mu_4 - \sigma^4) a_{n,ji} + 4\sigma^2 (\sum_{j=1}^{n-1} a_{n,ij} v_j)^2 + 2\mu_3 a_{n,ii} (\sum_{j=1}^{n-1} a_{n,ij} v_j)\), and \(E(Z_{ni}^* \cdot \sigma_{Q_n}^{2+\delta} (H_n + 2H_n) + H_n^2 H_3n\), where

\[
H_{1n} = \frac{h_n}{n} \sum_{i=1}^{n} \sum_{j=1}^{n-1} a_{n,ij} a_{n,ik} v_j v_k, \quad H_{2n} = \frac{h_n}{n} \sum_{i=1}^{n} \sum_{j=1}^{n-1} a_{n,ij} (v_j^2 - \sigma^2),
\]

and \(H_{3n} = \frac{h_n}{n} \sum_{i=1}^{n} \sum_{j=1}^{n-1} a_{n,ii} a_{n,ij} v_j \cdot v_j\). We would like to show that \(H_{jn}\) for \(j = 1, 2, 3\) converge in probability to zero. \(E(H_{3n}) = 0\). By exchanging summations, \(\sum_{i=1}^{n} \sum_{j=1}^{n-1} a_{n,ii} a_{n,ij} v_j = \sum_{j=1}^{n-1} (\sum_{i=j+1}^{n} a_{n,ii} a_{n,ij}) v_j\).

Thus, \(E(H_{3n}^2) = \frac{h_n^2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n-1} (a_{n,ii} a_{n,ij})^2 \leq \frac{\sigma^2 h_n^2}{n} \max_{1 \leq i \leq n} |a_{n,ii}|^2 \cdot \sum_{i=1}^{n-1} (\sum_{j=i+1}^{n} a_{n,ij})^2 = O(\frac{1}{n})\) because \(\max_{i,j} |a_{n,ij}|^2 = O(\frac{1}{n^2})\) and \(\sum_{i=1}^{n-1} \sum_{j=1}^{n} |a_{n,ij}|^2 = O(n)\). \(E(H_{2n}) = 0\) and \(H_{2n}\) can be rewritten into \(H_{2n} = \frac{h_n}{n} \sum_{i=1}^{n-1} (\sum_{j=i+1}^{n} a_{n,ii}^2) (v_j^2 - \sigma^2)\). Thus

\[
E(H_{2n}^2) = (\frac{h_n}{n})^2 (\mu_4 - \sigma^4) \sum_{j=1}^{n-1} \sum_{i=1}^{n} a_{n,ii}^2 \leq (\frac{h_n}{n})^2 (\mu_4 - \sigma^4) \max_{1 \leq i,j \leq n} |a_{n,ij}|^2 \cdot \sum_{j=1}^{n} \sum_{i=1}^{n} (\sum_{j=i+1}^{n} a_{n,ij})^2 = O(\frac{1}{n}).
\]
We conclude that $H_{3n} = o_P(1)$ and $H_{2n} = o_P(1)$. $E(H_{1n}) = 0$ but its variance is relatively more complex than that of $H_{2n}$ and $H_{3n}$. By rearranging terms, $H_{1n} = \frac{h_n}{n} \sum_{j=1}^{n-1} \sum_{k \neq j}^{n-1} a_{n,ij} a_{n,ik} v_j v_k = \frac{h_n}{n} \sum_{j=1}^{n-1} \sum_{k \neq j}^{n-1} \bar{S}_{n,jk} v_j v_k$, where $\bar{S}_{n,jk} = \sum_{i=\max(j,k)+1}^{n} a_{n,ij} a_{n,ik}$. The variance of $H_{1n}$ is

$$E(H_{1n}^2) = \left(\frac{h_n}{n}\right)^2 \sum_{j=1}^{n-1} \sum_{k \neq j}^{n-1} \sum_{i=1}^{n} \sum_{i' = 1}^{n} |a_{n,ij} a_{n,i'k} a_{n,i'j} a_{n,ik}|$$

As $k \neq j$ and $r \neq s$, $E(v_j v_k v_r v_s) \neq 0$ only for the cases that $(j = r) \neq (k = s)$ and $(j = s) \neq (k = r)$. The variance of $H_{1n}$ can be simplified and

$$E(H_{1n}^2) = 2\sigma^4 \left(\frac{h_n}{n}\right)^2 \sum_{j=1}^{n-1} \sum_{k \neq j}^{n-1} \bar{S}_{n,jk}^2 \leq 2\sigma^4 \left(\frac{h_n}{n}\right)^2 \sum_{j=1}^{n-1} \sum_{k \neq j}^{n-1} \left(\sum_{i=1}^{n} \sum_{i' = 1}^{n} |a_{n,ij} a_{n,i'k} a_{n,i'j} a_{n,ik}| \right)$$

$$\leq 2\sigma^4 \left(\frac{h_n}{n}\right)^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i' = 1}^{n} \sum_{i' = 1}^{n} |a_{n,ij} a_{n,i'k} a_{n,i'j} a_{n,ik}|$$

$$\leq 2\sigma^4 \left(\frac{h_n}{n}\right)^2 \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{n,ij}| \cdot \max_{1 \leq k \leq n} \sum_{i=1}^{n} |a_{n,ik}|$$

because $A_n$ is uniformly bounded in row and column sums and $a_{n,ij} = O\left(\frac{1}{n}\right)$ uniformly in $i$ and $j$. Thus, $H_{1n} = o_P(1)$ as $\lim_{n \to \infty} \frac{h_n}{n} = 0$ implied by the condition $\lim_{n \to \infty} \frac{1}{h_n} = 0$.

As $H_{jn}, j = 1, 2, 3$, are $o_P(1)$ and $\lim_{n \to \infty} \frac{h_n}{n} \sigma_Q^2 > 0$, $\sum_{i=1}^{n} E(Z_{ni}^2 | J_{i-1})$ converges in probability to 1. The central limit theorem for the martingale difference double array is thus applicable (see, Hall and Heyde, 1980; Potscher and Prucha, 1997) to establish the result. \textbf{Q.E.D.}

\textbf{Lemma A.14} Suppose that $A_n$ is a square matrix with its column sums being uniformly bounded and elements of the $n \times k$ matrix $C_n$ are uniformly bounded. Then, $\frac{1}{\sqrt{n}} C_n' A_n V_n = O_P(1)$. Furthermore, if the limit of $\frac{1}{\sqrt{n}} C_n A_n' C_n$ exists and it is positive definite, then $\frac{1}{\sqrt{n}} C_n' A_n V_n \overset{D}{\to} N(0, \sigma^2 \lim_{n \to \infty} \frac{1}{\sqrt{n}} C_n' A_n' C_n)$.

Proof: This is Lemma A.2 in Lee (1999b). These results can be established by Chebyshev’s inequality and Liapounov double array central limit theorem. \textbf{Q.E.D.}
Appendix B: Proofs of Theorems

Proof of Theorem 1. We will show that $\frac{1}{n}\ln L_n(\theta) - E(\frac{1}{n}\ln L_n(\theta)) = o_P(1)$ at each $\theta$ and is stochastically equicontinuous. Let $F_n(\lambda) = S_n^{-1}S_n'\lambda S_n(\lambda)S_n^{-1}$. As

$$E\left(\frac{1}{n}\ln L_n(\theta)\right) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln \sigma^2 + \frac{1}{n} \ln |S_n(\lambda)| - \frac{\sigma^2}{2\sigma^2 n} tr(F_n(\lambda)),$$

it follows that

$$\frac{1}{n}\ln L_n(\theta) - E\left(\frac{1}{n}\ln L_n(\theta)\right) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln \sigma^2 + \frac{1}{n} \ln |S_n(\lambda)| - \frac{\sigma^2}{2\sigma^2 n} [tr(F_n(\lambda)) - tr(F_n(\lambda))].$$

With the formula of Lemma A.10, var$\left(\frac{1}{n}\ln L_n(\theta)\right) = \frac{1}{2\sigma^4 n} \left\{ (\mu_4 - 3\sigma_4^4) \sum_{i=1}^{n} F_{n,i}(\lambda) + 2\sigma_4^4 tr(F_n(\lambda)F_n(\lambda)) \right\}$. By Lemma A.9, $F_{n,i}(\lambda) = O(1)$ and $tr(F_n(\lambda)F_n(\lambda)) = O(n)$. Therefore, var$\left(\frac{1}{n}\ln L_n(\theta)\right) = O\left(\frac{1}{n}\right)$. The Chebyshev inequality implies that $\frac{1}{n}\ln L(\theta) - E\left(\frac{1}{n}\ln L(\theta)\right) = o_P(1)$ at each $\theta$.

Let $\theta$ and $\tilde{\theta}$ be two values of the parameter vector, $T_{n1} = S_n^{-1}W_{n,n}S_n^{-1}$ and $T_{n2} = S_n^{-1}W_{n,n}W_{n,n}S_n^{-1}$. We have

$$\frac{1}{\sigma^2} Y_n'S_n(\lambda)S_n(\lambda)Y_n - \frac{1}{\sigma^2} Y_n'S_n(\tilde{\lambda})S_n(\tilde{\lambda})Y_n$$

$$= \frac{(\tilde{\sigma}^2 - \sigma^2)}{\sigma^2 \sigma^2} Y_n'S_n(\lambda)S_n(\lambda)Y_n + \frac{(\tilde{\lambda} - \lambda)}{\sigma^2} \{2Y_n'W_{n,n}Y_n - (\lambda + \tilde{\lambda})Y_n'W_{n,n}W_{n,n}Y_n\},$$

and

$$\frac{1}{\sigma^2} tr(F_n(\lambda)) - \frac{1}{\sigma^2} tr(F_n(\tilde{\lambda})) = \frac{1}{\sigma^2} [tr(F_n(\lambda)) - tr(F_n(\tilde{\lambda}))] + \frac{1}{\sigma^2 \sigma^2} (\tilde{\sigma}^2 - \sigma^2) tr(F_n(\tilde{\lambda}))$$

$$= \frac{1}{\sigma^2} (\tilde{\lambda} - \lambda) \{2tr(T_{n1}) - (\lambda + \tilde{\lambda}) tr(T_{n2})\} + \frac{1}{\sigma^2 \sigma^2} (\tilde{\sigma}^2 - \sigma^2) tr(F_n(\tilde{\lambda})).$$

Hence,

$$\left| \frac{1}{n}\ln L_n(\theta) - E\left(\frac{1}{n}\ln L_n(\theta)\right) \right| - \left| \frac{1}{n}\ln L_n(\tilde{\theta}) - E\left(\frac{1}{n}\ln L_n(\tilde{\theta})\right) \right|$$

$$\leq \frac{1}{2\sigma^4 n} \left\{ \frac{1}{n} Y_n'S_n(\lambda)S_n(\lambda)Y_n - \frac{1}{2\sigma^2} Y_n'S_n(\tilde{\lambda})S_n(\tilde{\lambda})Y_n \right\} + \frac{\sigma^2}{2\sigma^4 n} tr(F_n(\lambda)) - \frac{\tilde{\sigma}^2}{2\sigma^4 n} tr(F_n(\tilde{\lambda}))$$

$$\leq \frac{(\tilde{\lambda} - \lambda)}{\sigma^2} \left\{ \frac{1}{n} Y_n'W_{n,n}Y_n + \frac{1}{n} Y_n'W_{n,n}W_{n,n}Y_n \right\} + 2\frac{\tilde{\lambda} - \lambda}{\sigma^2} \left\{ \frac{1}{n} Y_n'W_{n,n}Y_n + \frac{1}{n} Y_n'W_{n,n}W_{n,n}Y_n \right\} + 2\frac{(\tilde{\sigma}^2 - \sigma^2)}{\sigma^2 \sigma^2} \left\{ \frac{1}{n} tr(T_{n1}) + \frac{1}{n} tr(T_{n2}) \right\}$$

Lemma A.11 implies that $|\frac{1}{n} Y_n'Y_n| = |\frac{1}{n} Y_n'S_n^{-1}S_n^{-1}V_n| = O_P(1)$, $|\frac{1}{n} Y_n'W_{n,n} + W_{n,n}Y_n| = O_P(\frac{1}{n})$ and $|\frac{1}{n} Y_n'W_{n,n}W_{n,n}Y_n| = O_P(\frac{1}{n})$ by Lemma A.12. Also, $\frac{1}{n} tr[S_n^{-1}S_n^{-1}] = O(1)$ from Lemma A.9. Lemma A.8 implies that $|\frac{1}{n} tr(T_{n1})| = O(\frac{1}{n})$ and $|\frac{1}{n} tr(T_{n2})| = O(\frac{1}{n})$. These imply the stochastic equicontinuity of $\frac{1}{n}\ln L_n(\theta) - E(\frac{1}{n}\ln L_n(\theta))$. 

28
With the pointwise probability convergence and stochastic equicontinuity, \( \frac{1}{n} \ln L_n(\theta) - E(\frac{1}{n} \ln L_n(\theta)) \) \( \overset{p}{\longrightarrow} 0 \) uniformly in \( \theta \in \Theta \) (Davidson, 1994, Theorem 21.9). Q.E.D.

**Proof of Theorem 2.** Because \( \frac{h_n}{n}(\ln L_n(\lambda) - Q_n(\lambda)) = -\frac{h_n}{2\sigma^2_n(\lambda)}(\hat{\sigma}^2_n(\lambda) - \sigma^2_n(\lambda)) \), it follows by the mean value theorem that

\[
\frac{h_n}{n}(\ln L_n(\lambda) - Q_n(\lambda)) = -\frac{h_n}{2\sigma^2_n(\lambda)}(\hat{\sigma}^2_n(\lambda) - \sigma^2_n(\lambda)),
\]

where \( \hat{\sigma}^2_n(\lambda) \) lies between \( \sigma^2_n(\lambda) \) and \( \hat{\sigma}^2_n(\lambda) \). Because

\[
\begin{align*}
\frac{h_n}{n}[(\hat{\sigma}^2_n(\lambda) - \sigma^2_n(\lambda))] &= \left\{ \frac{h_n}{n}V_n' S_n^{-1} V_n - \sigma^2_0 \frac{h_n}{n} \text{tr}(S_n'^{-1} S_n^{-1}) \right\} \\
&\quad - \lambda \left\{ \frac{h_n}{n}V_n' S_n^{-1} (W_n' + W_n) S_n^{-1} V_n - \sigma^2_0 \frac{h_n}{n} \text{tr}[S_n'^{-1}(W_n' + W_n) S_n^{-1}] \right\} \\
&\quad + \lambda^2 \left\{ \frac{h_n}{n}V_n' G_n' G_n V_n - \sigma^2_0 \frac{h_n}{n} \text{tr}[G_n' G_n] \right\},
\end{align*}
\]

Lemma A.12 implies that \( \sup_\lambda |\hat{\sigma}^2_n(\lambda) - \sigma^2_n(\lambda)| = o_P(1) \) on any bounded set of \( \lambda \). In particular, this implies that \( \hat{\sigma}^2_n(\lambda) - \sigma^2_n(\lambda) \) converges to zero in probability uniformly in \( \lambda \) in any bounded set. As \( \hat{\sigma}^2_n(\lambda) \) lies between \( \hat{\sigma}^2_n(\lambda) \) and \( \sigma^2_n(\lambda) \), it follows that \( \frac{1}{\sigma^2_n(\lambda)} \leq \max\{\frac{1}{\hat{\sigma}^2_n(\lambda)}, \frac{1}{\hat{\sigma}^2_n(\lambda)}\} \leq \frac{1}{\sigma^2_n(\lambda)} \) and

\[
\frac{1}{h_n}(\ln L_n(\lambda) - Q_n(\lambda))) \leq \frac{1}{2} \frac{1}{\sigma^2_n(\lambda)} |\hat{\sigma}^2_n(\lambda) - \sigma^2_n(\lambda)|.
\]

One can consider the stochastic boundedness of \( \frac{1}{\sigma^2_n(\lambda)} \) and \( \frac{1}{\hat{\sigma}^2_n(\lambda)} \). Because \( S_n(\lambda) = S_n - (\lambda - \lambda_0) W_n, \)

\[
\sigma^2_n(\lambda) = \sigma^2_n(\lambda) \text{tr}(S_n'^{-1} S_n(\lambda) S_n(\lambda) S_n'^{-1}) = \sigma^2(1 - 2(\lambda - \lambda_0) \frac{\text{tr}(G_n)}{n} + (\lambda - \lambda_0)^2 \frac{\text{tr}(G_n' G_n)}{n}),
\]

i.e.,

\[
\sigma^2_n(\lambda) - \sigma^2(\lambda) = \frac{\sigma^2}{n} [2(\lambda - \sigma^2) \text{tr}(G_n) + (\lambda - \lambda_0)^2 \text{tr}(G_n' G_n)].
\]

It follows that \( |\sigma^2_n(\lambda) - \sigma^2(\lambda)| \leq O(1) \cdot |\lambda - \lambda_0| \) on any bounded set of \( \lambda \). As \( \sigma^2(\lambda) > 0, \sup_{\lambda \in \Lambda} \frac{1}{\sigma^2_n(\lambda)} < \infty \) on a small neighborhood \( \Lambda \) of \( \lambda_0 \). Furthermore, because \( \hat{\sigma}^2_n(\lambda) - \sigma^2_n(\lambda) = o_P(1) \) uniformly in \( \lambda \) in any bounded set, one has that \( \sup_{\lambda \in \Lambda} \frac{1}{\sigma^2_n(\lambda)} = O_P(1) \). Thus, \( \frac{h_n}{n}(\ln L_n(\lambda) - Q_n(\lambda)) \) converges to zero in probability uniformly in \( \lambda \) in \( \Lambda \).

Since \( \sigma^2_n(\lambda) \) is a quadratic function of \( \lambda \), it is easy to see that it has a minimum at \( \lambda = \lambda_0 + \frac{\text{tr}(G_n)}{\text{tr}(G_n' G_n)} \) and the minimum is \( \sigma^2_{n, \min} = \sigma^2(1 - \frac{\text{tr}(G_n)}{\text{tr}(G_n' G_n)}) \). Under the assumption that \( \lim_{n \to \infty} \frac{\text{tr}(G_n)}{\text{tr}(G_n' G_n)} < 1, \)

\( \sigma^2_{n, \min} \) is bounded away from zero for large enough \( n \). Because \( \hat{\sigma}^2_n(\lambda) - \sigma^2_n(\lambda) \) converges to zero in probability uniformly in \( \lambda, \sigma_n(\lambda) \) is bounded away from zero in probability on any bounded set \( \Lambda \). For the case that \( \lim_{n \to \infty} h_n = \infty \), because \( \frac{\text{tr}(G_n)}{\text{tr}(G_n' G_n)} = O(\frac{1}{h_n}) \), the specified assumption is always satisfied. Hence, \( \sup_{\lambda \in \Lambda} \frac{1}{\sigma^2_n(\lambda)} = O(1) \) and \( \sup_{\lambda \in \Lambda} \frac{1}{\hat{\sigma}^2_n(\lambda)} = O_P(1) \). Q.E.D.
Proof of Theorem 3. As \( \frac{h_n}{n} (Q_n(\lambda) - Q_n(\lambda_0)) = -\frac{h_n}{n} (\ln \sigma_n^2(\lambda) - \ln \sigma_0^2) + \frac{h_n}{n} (\ln |S_n(\lambda)| - \ln |S_n(\lambda_0)|) \), by expanding this function at \( \lambda_0 \) up to the third order terms,
\[
\frac{h_n}{n} (Q_n(\lambda) - Q_n(\lambda_0)) = \frac{h_n}{2n} \frac{2\text{tr}^2(G_n)}{n} - \text{tr}(G_n G_n) - \frac{tr(G_n^2)}{n} (\lambda - \lambda_0)^2 + \frac{8\sigma_0^3}{\sigma_n^2(\lambda)} \frac{h_n}{n} \frac{tr^3(G_n' S_n(\lambda) S_n^{-1})}{3!} + \frac{6\sigma_0^4}{\sigma_n^4(\lambda)} \frac{h_n}{n} \frac{tr(G_n' S_n(\lambda) S_n^{-1})}{n} \frac{tr(G_n G_n)}{n} + \frac{2h_n}{n} \frac{tr((W_{n,0} S_n^{-1}(\lambda)) )}{3!} (\lambda - \lambda_0)^3.
\]
Because \( \text{tr}(G_n' S_n(\lambda) S_n^{-1}) \) is linear in \( \lambda \), it is of order \( O\left(\frac{h_n^2}{n^2}\right) \) uniformly in \( \lambda \) in any bounded set of \( \lambda \). From Lemma A.8, \( \text{tr}((W_{n,0} S_n^{-1}(\lambda)) ) \) is of order \( O\left(\frac{h_n}{n^2}\right) \) uniformly in \( \lambda \) in a neighborhood \( \Lambda_1 \) of \( \lambda_0 \). Therefore, there exists a constant \( c_1 > 0 \) such that, for large enough \( n \),
\[
\frac{h_n}{n} (Q_n(\lambda) - Q_n(\lambda_0)) \leq \frac{h_n}{2n} \frac{2\text{tr}^2(G_n)}{n} - \text{tr}(G_n G_n) - \frac{tr(G_n^2)}{n} (\lambda - \lambda_0)^2 + c_1 |\lambda - \lambda_0|^3.
\]
On the other hand, \( \frac{h_n}{2n} \frac{(2\text{tr}^2(G_n))}{n} \frac{tr(G_n G_n) - tr(G_n^2)}{n} \leq -c_2 \) for some \( c_2 > 0 \) for large \( n \). Thus, let \( \Lambda_2 = \{ \lambda : |\lambda - \lambda_0| < \frac{c_2}{c_1}\} \cap \Lambda_1 \) and one has
\[
\frac{h_n}{n} (Q_n(\lambda) - Q_n(\lambda_0)) < -\frac{c_2}{2} (\lambda - \lambda_0)^2.
\]
The identification uniqueness condition holds on \( \Lambda_2 \). Theorem 2 implies that \( \frac{h_n}{n} (\ln L_n(\lambda) - \ln L_n(\lambda_0)) - \frac{h_n}{n} (Q_n(\lambda) - Q_n(\lambda_0)) \) converges to zero in probability uniformly in a neighborhood \( \Lambda_3 \) of \( \lambda_0 \). By taking \( \Lambda = \Lambda_2 \cap \Lambda_3 \), the consistency of \( \hat{\lambda}_n \) follows from the uniform convergence and the identification uniqueness condition (White 1994). Q.E.D.

Proof of Theorem 4. Because \( Q_n(\lambda) - Q_n(\lambda_0) = -\frac{h_n}{n} (\ln \sigma_n^2(\lambda) - \ln \sigma_0^2) + \frac{h_n}{n} (\ln |S_n(\lambda)| - \ln |S_n(\lambda_0)|) \),
\[
\frac{\partial^2}{\partial \lambda^2} \frac{h_n}{n} (Q_n(\lambda) - Q_n(\lambda_0)) = \frac{h_n}{n} \left[ \frac{2\sigma_0^2}{\sigma_n^2(\lambda)} \frac{tr(G_n' S_n(\lambda) S_n^{-1})}{n} - \frac{\sigma_0^2}{\sigma_n^2(\lambda)} \frac{tr(G_n G_n)}{n} + \frac{6\sigma_0^2}{\sigma_n^4(\lambda)} \frac{h_n}{n} \frac{tr(G_n G_n)}{n} \right] (\lambda - \lambda_0)^2.
\]
As \( \sigma_n^2(\lambda) - \sigma_0^2 = O\left(\frac{1}{h_n}\right) \) and \( \frac{h_n}{n} tr^2(G_n' S_n(\lambda) S_n^{-1}) = O\left(\frac{1}{h_n}\right) \) uniformly in any bounded set of \( \lambda \), it follows that as \( \lim_{n \to \infty} h_n = \infty \),
\[
\limsup_{n \to \infty} \frac{\partial^2}{\partial \lambda^2} \frac{h_n}{n} (Q_n(\lambda) - Q_n(\lambda_0)) = \limsup_{n \to \infty} \frac{h_n}{n} \left[ \frac{2\sigma_0^2}{\sigma_n^2(\lambda)} \frac{tr(G_n' G_n)}{n} + \frac{6\sigma_0^2}{\sigma_n^4(\lambda)} \frac{h_n}{n} \frac{tr(G_n G_n)}{n} \right] (\lambda - \lambda_0)^2.
\]
uniformly in \( \lambda \) in any bounded set.

Note that \( W_{n,0} S_n(\lambda) = S_n(\lambda) W_{n,0} \) implies \( S_n^{-1}(\lambda) W_{n,0} = W_{n,0} S_n^{-1}(\lambda) \). When \( W_{n,0} = W_{n,0} \), one has \( S_n'(\lambda) = S_n(\lambda) \) and
\[
[W_{n,0} S_n^{-1}(\lambda)]' = [S_n^{-1}(\lambda) W_{n,0}]' = W_{n,0} S_n^{-1}(\lambda) = W_{n,0} S_n^{-1}(\lambda),
\]
30
that, in turn, implies $\text{tr}[(W_{n,n}S_n^{-1}(\lambda))^2] = \text{tr}[[W_{n,n}S_n^{-1}(\lambda)] [W_{n,n}S_n^{-1}(\lambda)]] \geq 0$. We conclude that when $W_{n,n}$ is symmetric and $\lim_{n \to \infty} h_n = \infty$, $\frac{h_n}{n} [Q_n(\lambda) - Q_n(\lambda_0)]$ is a concave function in $\lambda$ on any bounded set of $\lambda$ for large enough $n$.

When $W_{n,n} = \Lambda_n^{-1} D_n$, one has $S_n^{-1}(\lambda) = (I_n + \lambda \Lambda_n^{-1} D_n)^{-1} = (\Lambda_n - \lambda D_n)^{-1} \Lambda_n$. Therefore,

$$\text{tr}[(W_{n,n}S_n^{-1}(\lambda))^2] = \text{tr}(\Lambda_n^{-1} D_n (\Lambda_n - \lambda D_n)^{-1} \Lambda_n \cdot \Lambda_n^{-1} D_n (\Lambda_n - \lambda D_n)^{-1} \Lambda_n)$$

$$= \text{tr}(D_n (\Lambda_n - \lambda D_n)^{-1} \cdot D_n (\Lambda_n - \lambda D_n)^{-1}).$$

Also, $S_n^{-1}(\lambda) W_{n,n} = W_{n,n} S_n^{-1}(\lambda)$ implies that $(\Lambda_n - \lambda D_n)^{-1} D_n = \Lambda_n^{-1} D_n (\Lambda_n - \lambda D_n)^{-1} \Lambda_n$. It follows that

$$\text{tr}[(W_{n,n}S_n^{-1}(\lambda))^2] = \text{tr}(D_n \Lambda_n^{-1} D_n (\Lambda_n - \lambda D_n)^{-1} \Lambda_n (\Lambda_n - \lambda D_n)^{-1})$$

$$= \text{tr}(\Lambda_n^{-1/2} D_n (\Lambda_n - \lambda D_n)^{-1/2} \Lambda_n (\Lambda_n - \lambda D_n)^{-1/2} D_n \Lambda_n^{-1/2}).$$

Denote $E_n = \Lambda_n^{-1/2} D_n (\Lambda_n - \lambda D_n)^{-1/2} \Lambda_n$. As $D'_n = D_n$, $E'_n = \Lambda_n^{1/2} (\Lambda_n - \lambda D'_n)^{1/2} \Lambda_n^{1/2} = \Lambda_n^{1/2} (\Lambda_n - \lambda D_n)^{-1/2} D_n \Lambda_n^{-1/2}$. Consequently, $\text{tr}[(W_{n,n}S_n^{-1}(\lambda))^2] = \text{tr}(E_n E'_n) \geq 0$. For large enough $n$, $\frac{h_n}{n} [Q_n(\lambda) - Q_n(\lambda_0)]$ is also a concave function on any bounded set of $\lambda$ for this case. $\lambda_0$ is the global maximizer as $Q_n(\lambda) - Q_n(\lambda_0) < 0$ for any $\lambda \neq \lambda_0$ by Jensen’s inequality. Hence, the local identification of $\lambda_0$ implies the global identification of $\lambda_0$.

The consistency of $\hat{\lambda}_n$ follows from the global identification and the uniform convergence in Theorem 2.

Q.E.D.

**Proof of Proposition 1.** By Lemma A.10, $E(V'_n G_n V_n)^2 = (\mu_4 - 3\sigma_0^4) \sum_{i=1}^n G_{n,ii}^2 + \sigma_0^4 \text{tr}(G_n) + \text{tr}(G_n G_n') + \text{tr}(G_n^2)].$ Therefore,

$$E(P_n(\lambda_0)) = \frac{2}{n\sigma_0^4} E(V'_n G_n V_n)^2 - \frac{1}{\sigma_0^4} E(V'_n G_n' G_n V_n) - \text{tr}(G_n^2)$$

$$= \frac{2}{n}(\mu_4 - 3\sigma_0^4) \sum_{i=1}^n G_{n,ii}^2 + [\text{tr}(G_n) + \text{tr}(G_n' G_n) + \text{tr}(G_n^2)] - \text{tr}(G_n' G_n) - \text{tr}(G_n^2)$$

$$= \frac{2}{n} \text{tr}(G_n) - \text{tr}(G_n' G_n) - \text{tr}(G_n^2) + \frac{2}{n} \text{tr}(G_n G_n') + \text{tr}(G_n^2) + \frac{2}{n}(\frac{\mu_4 - 3\sigma_0^4}{\sigma_0^4}) \sum_{i=1}^n G_{n,ii}^2.$$ 

By Lemma A.8, $\sum_{i=1}^n G_{n,ii}^2 = O\left(\frac{1}{n}\right)$, and $\text{tr}(G_n^2)$ and $\text{tr}(G_n G_n')$ are of order $O\left(\frac{1}{n}\right)$. These give the first part of the results.

Eq.(4.6) can be rewritten as $\frac{h_n}{n} P_n(\lambda_0) = \frac{2}{\sigma_0^4}(\frac{h_n}{n} V'_n G_n V_n)^2 - \frac{1}{\sigma_0^4} \frac{h_n}{n} V'_n G_n' G_n V_n - \frac{h_n}{n} \text{tr}(G_n^2)$. The probability limit of $\frac{h_n}{n} P_n(\lambda_0)$ will follow from those of $\frac{h_n}{n} V'_n G_n V_n$ and $\frac{h_n}{n} V'_n G_n' G_n V_n$. The term $\frac{h_n}{n} (V'_n G_n V_n - \sigma_0^2 \text{tr}(G_n))$ has zero mean and its variance is

$$\text{var}\left(\frac{h_n}{n} (V'_n G_n V_n)\right) = \frac{h_n}{n^2} ((\mu_4 - 3\sigma_0^4) \sum_{i=1}^n G_{n,ii}^2 + \sigma_0^4 \text{tr}(G_n G_n') + \text{tr}(G_n^2)) = O\left(\frac{1}{n}\right).$$
by Lemmas A.10 and A.8. By the Chebyshev inequality, \( \frac{h_n}{n} V_n'G_nV_n - \sigma_0^2 \frac{h_n}{n} \text{tr}(G_n) \overset{P}{\to} 0. \) By the continuity mapping theorem (see, White 1984, p.25), \( \left( \frac{h_n}{n} V_n'G_nV_n \right)^2 - \left( \sigma_0^2 \frac{h_n}{n} \text{tr}(G_n) \right)^2 \overset{P}{\to} 0. \) The mean of \( \frac{h_n}{n} V_n'G_nG_nV_n \) is \( \sigma_0^2 \frac{h_n}{n} \text{tr}(G_n'G_n) \) and its variance is

\[
\text{var}\left[ \frac{h_n}{n} (V_n'G_nG_nV_n) \right] = \left( \frac{h_n}{n} \right)^2 \sum_{i=1}^{n} (G_n'G_n)_{ii}^2 + 2\sigma_0^4 \text{tr}((G_n'G_n)^2) = O\left( \frac{h_n}{n} \right).
\]

Hence, \( \frac{h_n}{n} V_n'G_n'G_nV_n - \sigma_0^2 \frac{h_n}{n} \text{tr}(G_n'G_n) \overset{P}{\to} 0. \) The probability limit of \( \frac{h_n}{n} P_n(\lambda_0) \) follows by combining these terms. Because \( \text{tr}(G_n) = O\left( \frac{n}{h_n} \right), \frac{h_n}{n} \text{tr}(G_n) = O\left( \frac{1}{h_n} \right) \) goes to zero if \( \lim_{n \to \infty} h_n = \infty. \quad \text{Q.E.D.} \)

**Proof of Proposition 2.** From (4.3) and (4.6), \( \frac{h_n}{n} \left[ \frac{\partial^2 \ln L_n(\tilde{\lambda}_n)}{\partial \lambda^2} - P_n(\lambda_0) \right] = 2\Delta_n + \Delta_{n2} + \Delta_{n3} \) where

\[
\Delta_n = \frac{1}{\sigma_0^2(\lambda_n)} \left[ \frac{1}{n} Y_n'W_n'W_nS(\tilde{\lambda}_n)Y_n - \frac{1}{\sigma_0^2} \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n - \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right) \right]
\]

\[
\Delta_{n2} = \frac{1}{\sigma_0^2(\lambda_n)} \left( \frac{1}{n} Y_n'W_n'W_nS(\tilde{\lambda}_n)Y_n - \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right), \quad \Delta_{n3} = \frac{1}{n} \left( \frac{1}{n} Y_n'W_n'W_nS(\tilde{\lambda}_n)Y_n - \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right).
\]

\[
\Delta_{n1} = (\frac{1}{\sigma_0^2(\lambda_n)} - \frac{1}{\sigma_0^2}) \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n\right), \quad \Delta_{n12} = \frac{1}{\sigma_0^2(\lambda_n)} \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n\right).
\]

We would like to show that all these terms converge in probability to zero.

As \( \hat{\sigma}_n^2(\tilde{\lambda}_n) = \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n = \frac{1}{n} Y_n'Y_n - 2\tilde{\lambda}_n \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right) + \tilde{\lambda}_n^2 \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right) \) and that \( \tilde{\lambda}_n \overset{P}{\to} \lambda_0 \) imply

\[
\hat{\sigma}_n^2(\tilde{\lambda}_n) - \sigma_0^2(\lambda_0) = 2(\lambda_0 - \tilde{\lambda}_n) \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right) + (\tilde{\lambda}_n^2 - \lambda_0^2) \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right) = o_P(1).
\]

As \( \hat{\sigma}_n^2(\lambda_0) = \frac{1}{n} Y_n'V_n \overset{P}{\to} \sigma_0^2 \) by the strong law of large numbers for i.i.d. variables, we conclude that \( \hat{\sigma}_n^2(\tilde{\lambda}_n) \overset{P}{\to} \sigma_0^2. \) Similarly, \( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n = \frac{1}{n} Y_n'W_nS(\lambda_0)Y_n - (\tilde{\lambda}_n - \lambda_0) \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n = O_P(\frac{1}{n}) \)

and \( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n = \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n - (\tilde{\lambda}_n - \lambda_0) \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n = O_P(1). \) It follows by the Slutsky lemma that \( \Delta_{n11} \overset{P}{\to} 0. \) By expanding the products in \( \Delta_{n12}, \)

\[
\Delta_{n12} = \left[ 2(\lambda_0 - \tilde{\lambda}_n) \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right) + (\tilde{\lambda}_n^2 - \lambda_0^2) \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right) \right] \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right) = o_P(1).
\]

Thus, \( \Delta_{n1} \overset{P}{\to} 0, \Delta_{n2} \overset{P}{\to} 0 \) because \( \Delta_{n2} = (\frac{1}{\sigma_0^2(\lambda_n)} - \frac{1}{\sigma_0^2}) \left( \frac{1}{n} Y_n'W_nS(\tilde{\lambda}_n)Y_n \right) = o_P(1). \) By the mean-value theorem, \( \text{tr}(G_n^2(\tilde{\lambda}_n)) = \text{tr}(G_n^2) + 2\text{tr}(G_n^3(\tilde{\lambda}_n)) \cdot (\tilde{\lambda}_n - \lambda_0), \) where \( \tilde{\lambda}_n \) lies between \( \tilde{\lambda}_n \) and \( \lambda_0. \) As \( \{S_n^{-1}(\lambda)\} \) is
uniformly bounded in row sums uniformly in $\lambda$ in a neighborhood of $\lambda_0$, $G_n^3(\tilde{\lambda}_n)$ is uniformly bounded in row sums. It follows from Lemma A.8 that $tr(G_n^3(\tilde{\lambda}_n)) = O_P(\frac{1}{n})$. Hence $\Delta_{n3} = 2\frac{h_n}{n}tr(G_n^3(\tilde{\lambda}_n)) \cdot (\tilde{\lambda}_n - \lambda_0) = o_P(1)$. Q.E.D.

**Proof of Proposition 3.** It follows from Propositions 1 and 2 that
\[
\frac{h_n}{n} \frac{\partial^2 \ln L_n(\tilde{\lambda}_n)}{\partial \lambda^2} = \frac{h_n}{n} \left[ \frac{\partial^2 \ln L_n(\tilde{\lambda}_n)}{\partial \lambda^2} - P_n(\lambda_0) \right] + \frac{h_n}{n} P_n(\lambda_0)
\]
and
\[
\lim_{n \to \infty} \frac{h_n}{n} \left[ \frac{2}{n} tr^2(G_n) - tr(G_n G'_n) - tr(G_n^2) \right].
\]
Q.E.D.

**Proof of Proposition 4.** The asymptotic distribution of the quadratic random form follows from Lemma A.13. The matrix $C_n$ in $Q_n$ can be replaced by a symmetric matrix $A_n = \frac{1}{2} (C_n + C'_n)$ such that $Q_n = V_n' A_n V_n$. For this case, $E(Q_n) = 0$ because $tr(A_n) = 0$. The entries of $C_n$ are of order $O(\frac{1}{n})$ and so are those of $A_n$. Also $\frac{h_n}{n} \sigma^2_{Q_n} = O(1)$ and its limit is positive. So $A_n$ satisfies the conditions of Lemma A.13. Under Assumption 3', $\lim_{n \to \infty} \frac{h_n^{1+n}}{n} = 0$ for some small $\eta > 0$. As $v$ is normally distributed, its absolute moment of any order exists. By taking $\delta$ in Lemma A.13 large enough such that $\frac{2}{3} \leq \eta$, Assumption 3' implies that $\lim_{n \to \infty} \frac{h_n^{1+\delta}}{n} = 0$. Hence, the result follows from Lemma A.13. Q.E.D.

**Proof of Proposition 5.** When $v_i$'s are normally distributed, $\mu_4 = 3\sigma^4_0$ and the first term in the variance formula in (4.9) drops out.

For the case that $\lim_{n \to \infty} h_n = \infty$, the simplification depends on the order of relevant terms. As $G_{n,ii} = O(\frac{1}{n})$ by Lemma A.8, it implies that $C_{n,ii} = G_{n,ii} - \frac{tr(G_n)}{n} = O(\frac{1}{n})$ and, consequently, $\sum_{i=1}^n C_{n,ii}^2 = O(\frac{1}{n^2})$. Hence, $\frac{h_n}{n} \sum_{i=1}^n C_{n,ii}^2 = O(\frac{1}{h^2_n})$ goes to zero when $\lim_{n \to \infty} h_n = \infty$. Q.E.D.

**Proof of Theorem 5.** The result follows from Propositions 3 and 5. For the special cases, $\Omega = 0$. Q.E.D.

**Proof of Theorem 6.** As $\hat{\sigma}^2_n(\lambda_0) = \frac{1}{n} Y_n' S_n S_n Y_n = \frac{1}{n} V_n' V_n$, it follows that
\[
\sqrt{n}(\hat{\sigma}^2_n(\tilde{\lambda}_n) - \sigma^2_0) = \sqrt{n}(\hat{\sigma}^2_n(\tilde{\lambda}_n) - \hat{\sigma}^2_n(\lambda_0)) + \sqrt{n}(\hat{\sigma}^2_n(\lambda_0) - \sigma^2_0)
\]
and
\[
= ((\tilde{\lambda}_n + \lambda_0) \frac{1}{n} Y_n' W_n' W_n Y_n - 2\frac{1}{n} Y_n' W_n Y_n) \cdot \sqrt{n}(\tilde{\lambda}_n - \lambda_0) + \sqrt{n}(\frac{1}{n} V_n' V_n - \sigma^2_0)
\]
\[
= (\tilde{\lambda}_n + \lambda_0) \frac{\sqrt{n}}{n} Y_n' W_n' W_n Y_n - 2\frac{\sqrt{n}}{n} Y_n' W_n Y_n \cdot \sqrt{n}(\tilde{\lambda}_n - \lambda_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (v_i^2 - \sigma^2_0).
\]
From Propositions 3 and 5, 
\[ \sqrt{n} (\hat{\lambda}_n - \lambda_0) = \Sigma_{\lambda\lambda} \cdot \sqrt{\frac{h_n Q_n}{\sigma_0^2}} + o_P(1). \]
For any finite \( n \), we note that \( \sqrt{\frac{h_n Q_n}{\sigma_0^2}} \) and \( \frac{1}{n} \sum_{i=1}^{n} (v_i^2 - \sigma_0^2) \) are uncorrelated. This is so, because \( \text{tr}(C_n) = 0 \) implies that
\[
E[V_n^r C_n V_n(V_n^r V_n - n\sigma_0^2)] = E[\left( \sum_{i=1}^{n} \sum_{j=1}^{n} c_{n,ij} v_i v_j \right) (\sum_{k=1}^{n} v_k^2)] \\
= \sum_{i=1}^{n} c_{n,ii} E(v_i^4) + \sum_{i=1}^{n} \sum_{k \neq i} c_{n,ii} E(v_i^2 v_k^2) \\
= (\mu_4 + (n - 1)\sigma_0^4) \sum_{i=1}^{n} c_{n,ii} = 0.
\]

\[
E(h_n^2 Y_n^r W_n Y_n) = \frac{h_n^2}{n} \sigma_0^2 \text{tr}(S_n^{-1} G_n) \text{ and } E(h_n^2 Y_n^r W_n^r Y_n W_n Y_n) = \frac{h_n^2}{n} \sigma_0^2 \text{tr}(G'_n G_n). \]
Lemma A.12 implies that 
\[
\text{var}(\sqrt{n} h_n^2 Y_n^r W_n Y_n) = O(\frac{1}{n}) \text{ and } \text{var}(\sqrt{n} h_n^2 Y_n^r W_n^r Y_n W_n Y_n) = O(\frac{1}{n}).
\]
It follows by the Chebyshev inequality that
\[
\sqrt{n} h_n^2 Y_n^r W_n Y_n - \frac{h_n^2}{n} \sigma_0^2 \text{tr}(S_n^{-1} G_n) = o_P(1) \text{ and } \sqrt{n} h_n^2 Y_n^r W_n^r Y_n W_n Y_n - \frac{h_n^2}{n} \sigma_0^2 \text{tr}(G'_n G_n) = o_P(1).
\]
As \( S_n^{-1} - \lambda_0 G_n = I_n, \text{tr}(S_n^{-1} G_n) - \lambda_0 \text{tr}(G_n G_n) = \text{tr}(G_n) \). These imply that
\[
\sqrt{n} (\hat{\sigma}_n^2 (\hat{\lambda}_n) - \sigma_0^2) = -2\sigma_0^2 \frac{h_n^2}{n} \text{tr}(G_n) \Sigma_{\lambda\lambda} \cdot \sqrt{\frac{h_n Q_n}{\sigma_0^2}} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (v_i^2 - \sigma_0^2) + o_P(1).
\]
As \( \sqrt{n} (\hat{\lambda}_n - \lambda_0) = \Sigma_{\lambda\lambda} \cdot \sqrt{\frac{h_n Q_n}{\sigma_0^2}} + o_P(1) \), the results follow from the Martingale central limit theorem as in the proof of Proposition 4. For the case that \( \lim_{n \to \infty} h_n = \infty, \psi = 0 \) as \( \sqrt{\frac{h_n}{n}} \text{tr}(G_n) = O(\sqrt{\frac{1}{h_n}}) \to 0. \)
Q.E.D.
Appendix C: The Method of Moments

The purpose of the Appendix is to illustrate that initial consistent estimates can be easily obtained. These consistent estimates can be used as initial estimates for the QML method or for the construction of a second-round estimator.

The method of moments can be applied to the spatial autoregressive process (2.1). We investigate a method of moments estimator related to that of Kelejian and Prucha (1999a) under our setting. The moments considered are based on functions of second moments of $V_n$, namely,

$$E(V_n^2V_n) = n\sigma_0^2, \quad E[(W_{n,n}V_n)'(W_{n,n}V_n)] = \sigma_0^2 \text{tr}(W_{n,n}'W_{n,n}), \quad E[(W_{n,n}V_n)'V_n] = \sigma_0^2 \text{tr}(W_{n,n}) = 0.$$

Equivalently, these moment equations are

$$E\left(\frac{1}{n}[Y_n - (W_{n,n}Y_n)\lambda_0]'[Y_n - (W_{n,n}Y_n)\lambda_0]\right)$$

$$= E\left(\frac{1}{n}Y_n^2Y_n\lambda_0\right) - 2E\left(\frac{1}{n}Y_n'W_{n,n}Y_n\lambda_0\right) + E\left(\frac{1}{n}(W_{n,n}Y_n)'(W_{n,n}Y_n)\right)\lambda_0^2 = \sigma_0^2,$$  \hspace{1cm} (C.1)

$$E(Y_n'W_{n,n}'W_{n,n}Y_n) - 2E(Y_n'W_{n,n}'W_{n,n}Y_n)\lambda_0 + E(Y_n'W_{n,n}'W_{n,n}Y_n)\lambda_0^2 = \sigma_0^2 \text{tr}(W_{n,n}'W_{n,n}),$$  \hspace{1cm} (C.2)

and

$$E(Y_n'W_{n,n}'W_{n,n}Y_n) - [E(Y_n'W_{n,n}'Y_n) + E(Y_n'W_{n,n}'W_{n,n}Y_n)]\lambda_0 + E(Y_n'W_{n,n}'W_{n,n}Y_n)\lambda_0^2 = 0.$$  \hspace{1cm} (C.3)

By concentrating these moments with respect to $\sigma^2$, (C.1) and (C.2) imply that

$$E(Y_n'W_{n,n}'W_{n,n}Y_n) - \text{tr}(W_{n,n}'W_{n,n})E\left(\frac{1}{n}Y_n^2Y_n\lambda_0\right) - 2[E(Y_n'W_{n,n}'W_{n,n}Y_n) - \text{tr}(W_{n,n}'W_{n,n})E\left(\frac{1}{n}Y_n^2W_{n,n}'W_{n,n}Y_n\right)]\lambda_0$$

$$+ [E(Y_n'W_{n,n}'W_{n,n}Y_n) - \text{tr}(W_{n,n}'W_{n,n})E\left(\frac{1}{n}Y_n^2W_{n,n}'W_{n,n}Y_n\right)]\lambda_0^2 = 0.$$  \hspace{1cm} (C.4)

A method of moments estimator of $\lambda$ can be derived from the moment equations (C.3) and (C.4). Denote

$$b_{n1} = Y_n'W_{n,n}'W_{n,n}Y_n - \text{tr}(W_{n,n}'W_{n,n}) \cdot \frac{1}{n}Y_n^2Y_n, \quad b_{n2} = Y_n'W_{n,n}'W_{n,n}Y_n,$$

$$a_{n11} = 2[Y_n'W_{n,n}'W_{n,n}Y_n - \text{tr}(W_{n,n}'W_{n,n}) \cdot \frac{1}{n}Y_n^2W_{n,n}'W_{n,n}Y_n],$$

$$a_{n12} = -Y_n'W_{n,n}'W_{n,n}Y_n + \text{tr}(W_{n,n}'W_{n,n}) \cdot \frac{1}{n}Y_n^2W_{n,n}'W_{n,n}Y_n,$$

$$a_{n21} = Y_n'W_{n,n}'W_{n,n}Y_n + Y_n'W_{n,n}'W_{n,n}Y_n, \quad a_{n22} = -Y_n'W_{n,n}'W_{n,n}Y_n.$$  \hspace{1cm} Let $A_n = \begin{pmatrix} a_{n11} & a_{n12} \\ a_{n21} & a_{n22} \end{pmatrix}$ and $b_n = \begin{pmatrix} b_{n1} \\ b_{n2} \end{pmatrix}$. The estimation of $\lambda$ based on (C.3) and (C.4) can be done with or without imposing the nonlinear
constraints on the coefficients of these moment equations. For the estimation without imposing constraints, a moment estimator \( \hat{\lambda}_{n1} \) of \( \lambda \) can be solved from the linear system: \( A_n(\hat{\lambda}_{n1}, \hat{\mu}_n)' = b_n \). That is,

\[
\hat{\lambda}_{n1} = (1, 0) A_n^{-1} b_n.
\]

It follows that \( (\hat{\lambda}_{n1}, \hat{\mu}_n)' - (\lambda_0, \lambda_0^2)' = A_n^{-1} c_n \) where \( c_n = b_n - A_n(\lambda_0, \lambda_0^2)' \). From Lemma A.12, all the entries of \( A_n \) and \( b_n \) are of order \( O_P(n^{-1}) \). The moment equations (C.3) and (C.4) imply that \( E(c_n) = 0 \). The variance of \( c_n \) is of order \( O(n^{-1}) \) from Lemma A.12. Therefore, \( \frac{h_n}{n} c_n = o_P(1) \). Lemma A.12 implies also that \( \frac{h_n}{n} A_n = O_P(1) \). Under the identification condition that \( \lim_{n \to \infty} \frac{h_n}{n} = \Gamma \) exists and \( \Gamma \) is nonsingular, \( \hat{\lambda}_{n1} = \lambda_0 + (1, 0)(\frac{h_n}{n} A_n)^{-1} \frac{h_n}{n} c_n = \lambda_0 + o_P(1) \), i.e., \( \hat{\lambda}_{n1} \) is consistent. Let \( c_{n1} \) and \( c_{n2} \) be, respectively, the first and second entries of \( c_n \). As \( Y_n = S_n^{-1} V_n \),

\[
c_{n1} = V_n s_n^{-1}(W_{n,n} W_{n,n} - tr(W_{n,n} W_{n,n}) I_n - 2[W_{n,n} W_{n,n} - tr(W_{n,n} W_{n,n}) W_{n,n}]\lambda_0
\]

\[+ [W_{n,n} W_{n,n} - tr(W_{n,n} W_{n,n}) W_{n,n}]\lambda_0^2] S_n^{-1} V_n,
\]

and

\[
c_{n2} = V_n s_n^{-1}(W_{n,n} - 2(W_{n,n} W_{n,n})\lambda_0 + W_{n,n} W_{n,n}\lambda_0^2] S_n^{-1} V_n.
\]

Let \( \Sigma_n \) be the variance matrix of \( c_n \). The central limit theorem for the quadratic form (Lemma A.13) is applicable to \( c_n \). One has \( \sqrt{\frac{h_n}{n} c_n} \xrightarrow{D} N(0, \lim_{n \to \infty} \frac{h_n}{n} \Sigma_n) \). It follows that

\[
\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{n1} - \lambda_0) = (1, 0)(\frac{h_n}{n} A_n)^{-1} \sqrt{\frac{h_n}{n} c_n} \xrightarrow{D} N(0, (1, 0)\Gamma^{-1}(\lim_{n \to \infty} \frac{h_n}{n} \Sigma_n)\Gamma^{-1}).
\]

Alternatively, \( \lambda \) can be estimated by the minimization: \( \min_{\lambda} \frac{1}{2} (A_n \theta - b_n)'(A_n \theta - b_n) \) where \( \theta = (\lambda, \lambda^2)' \). Let \( \hat{\lambda}_{n2} \) be the estimator of this method. It follows from the mean-value theorem that

\[
\hat{\lambda}_{n2} - \lambda_0 = [(0, 2) A_n'(A_n \bar{\theta} - b_n) + (1, 2\hat{\lambda}) A_n' A_n(1, 2\hat{\lambda})']^{-1} (1, 2\lambda_0) A_n' c_n,
\]

where \( \hat{\lambda} \) lies between \( \hat{\lambda}_{n2} \) and \( \lambda_0 \) and \( \bar{\theta} = (\bar{\lambda}, \bar{\lambda}^2)' \). The consistency of \( \hat{\lambda}_{n2} \) can be easily shown as \( \Gamma \) is invertible. Furthermore,

\[
\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{n2} - \lambda_0) \xrightarrow{D} N(0, [(1, 2\lambda_0)\Gamma(1, 2\lambda_0)'(1, 2\lambda_0)\Gamma'(\lim_{n \to \infty} \frac{h_n}{n} \Sigma_n)\Gamma(1, 2\lambda_0)'(1, 2\lambda_0)\Gamma'(1, 2\lambda_0)]^{-1}).
\]

\[ 
36
\]
References


*Geographical Analysis* 9, 175-184.


