Generalized Method of Moments Estimation of Spatial Autoregressive Processes

by

Lung-fei Lee*

Department of Economics

Ohio State University, Columbus, OH

August 2001

Abstract

The GMM is considered for the estimation of spatial autoregressive processes. This method has computational advantage over the conventional quasi-maximum likelihood method. The GMM estimators are shown to be consistent and asymptotically normal. Within certain classes of GMM estimators, best ones are derived. A best GMM estimator can have the same limiting distribution as the QML estimator. Some GMM estimators can be robust against unknown heteroskedasticity. The GMM can be extended to the estimation of higher-order spatial autoregressive processes without additional computational complexity.

Key Words: Spatial econometrics, spatial autoregressive processes, high-order spatial lags, GMM, quasi-maximum likelihood, efficiency, robustness

Classification: C13, C21, R15

Correspondence address:

Lung-fei Lee, Department of Economics, Ohio State University, Columbus, OH 43210-1172.

E-mail: lflee@econ.ohio-state.edu

* I am grateful to the National Science Foundation, Economics Program for financial support under the grant number SES-0111380 for my research.

1. Introduction

Spatial autoregressive (SAR) processes were introduced with the pioneer works of Whittle (1954) and Cliff and Ord (1973). There are various estimation problems of interest. In this paper, we develop computationally simpler methods than the conventional maximum likelihood (ML) method for the estimation of SAR processes. We propose a generalized method of moments (GMM) for the estimation of such processes.

In the existing econometrics literature, Kelejian and Prucha (1999a) have proposed a method of moments (MOM) for the estimation of the SAR process $Y_n = \rho W_{n,n} Y_n + \epsilon_n$ by exploring several moments of Y_n and $W_{n,n} Y_n$. Kelejian and Prucha (1999a) show that their parameter estimators of the model is consistent under some general regularity conditions. The asymptotic distribution of their estimator, however, has not been derived. Even though there are Monte Carlo evidences on possible efficiency of their estimators relative to the ML or the quasi-maximum likelihood (QML) estimator (under normal distribution specification), asymptotic relative efficiency properties are not available. The advantage of the MOM is the simpler computation with the MOM estimator than the ML or the QML estimator.

In this paper, we suggest a general GMM estimation framework, which is relatively computationally simpler than the QML and may have certain asymptotic relative efficiency or robust properties. The GMM estimation method introduced by Hansen (1982) has broad applications in macroeconometrics, financial econometrics and various economic fields. Hansen's GMM method goes beyond the nonlinear two-stage least squares (2SLS) method of Amemiya (1974) as it incorporates nonlinear moment conditions beyond moment conditions generated by orthogonality of instrumental variables (IV) and disturbances in a model. The GMM method has been noted for its possible use with the estimation of spatial models in the presence of exogenous variables, see, e.g., Anselin (1988, 1990), Land and Deane (1992), Kelejian and Robinson (1993), Kelejian and Prucha (1997, 1998), Lee (1999), among others. Those GMM methods are 2SLS methods as their moment conditions are based on exogenous variables in the model. For SAR processes, there are no relevant exogenous variables in the process and the 2SLS method is not applicable. However, in this paper, we notice that nonlinear moment conditions are available and they can be used for estimation in the GMM framework. The MOM in Kelejian and Prucha (1999a) is relevant but their moment equations (after modification) are only some related components in our estimation framework. Our GMM estimators can be shown to be consistent and asymptotically normal. Within certain classes of GMM estimators, the best selection of moment equations can be derived and the corresponding best GMM estimators are available. The best GMM estimator may have the same limiting distribution of the QML estimate under any distribution (satisfying certain general regularity conditions) for the disturbances. The GMM estimation framework can be easily extended to the estimation of high-order SAR processes.

This paper is organized as follows. In Section 2, the first-order SAR process is considered. The GMM estimation framework is introduced. Identification issues are discussed. Asymptotic properties of consistency, asymptotic normality and efficiency are established. GMM estimates with optimal weighting and best selection of moment equations are derived. In Section 3, the GMM estimation framework in Section 2 is generalized to the regression model where the disturbances form a SAR process. It is shown that the asymptotic properties of the GMM estimators are not affected when the disturbances are estimated (with the least squares residuals). Section 4 generalized the GMM framework to the estimation of general high-order SAR processes. Optimal and best GMM estimators are derived. Conclusions are drawn in Section 5. All the proofs of the propositions in the text are included in Appendix B. Appendix A collects some useful lemmas for the proofs.

2. Moment Conditions and Estimation of SAR Processes

A (first-order) SAR process is specified as

$$Y_n = \lambda W_n Y_n + \epsilon_n, \tag{2.1}$$

where Y_n is the n-dimensional vector of dependent variables, W_n is an $n \times n$ constant matrix of spatial weights with a zero diagonal, and the disturbances ϵ_{nj} of the vector $\epsilon_n = (\epsilon_{n1}, \cdots, \epsilon_{nn})'$ are independent and identically distributed (i.i.d.) with zero mean and variance σ^2. In order to distinguish the true parameters from other possible values of the parameters, λ_0 and σ_0^2 will denote, respectively, the true parameters of λ and σ^2. For any value λ, let $S_n(\lambda) = I_n - \lambda W_n$. At λ_0, denote $S_n = S_n(\lambda_0)$ for simplicity.

The SAR process is supposed to be an equilibrium model. Under the assumption that S_n is invertible, the equilibrium solution is

$$Y_n = S_n^{-1}\epsilon_n. \tag{2.2}$$

Because the endogeneity of Y_n and $W_n Y_n = G_n \epsilon_n$ where $G_n = W_n S_n^{-1}$ from (2.1) and (2.2), $W_n Y_n$ is generally correlated with ϵ_n. For the IV estimation of the model parameters of (2.1), valid IVs need to be constructed so that they are uncorrelated with ϵ_n but correlated with $W_n Y_n$. We suggest $P_n S_n(\lambda) Y_n$, where P_n is a $n \times n$ constant matrix with either a zero diagonal or, more generally, $tr(P_n) = 0$, as a possible IV function for the estimation of the model (2.1). The intuition behind this is that when P_n has a zero diagonal, the lth component of the IV vector $P_nS_nY_n(= P_n\epsilon_n)$ is $\sum_{j=1,j\neq l}^{n} p_{n,lj}\epsilon_{nj}$, where $p_{n,lj}$ is the (l,j)th entry of P_n, and ϵ_{nj} is the jth component of ϵ_n, which excludes ϵ_{nl} in its linear combination. So, each component of $P_nS_nY_n$ is uncorrelated with the corresponding component of ϵ_n. The selection of P_n with only $tr(P_n) = 0$ is more general because a P_n with a zero diagonal is a special case. The intuition for P_n with $tr(P_n) = 0$ is that while each component of $P_nS_nY_n$ may be correlated with the corresponding component of ϵ_n, the correlations may cancel each other. This is shown in the following proposition.

Proposition 2.1 *For any constant $n \times n$ matrix P_n with $tr(P_n) = 0$, $P_nS_nY_n$ is uncorrelated with ϵ_n.*

Proposition 2.1 provides a moment condition which can be useful for estimation. As $P_nS_n(\lambda)Y_n$ involves the unknown parameter λ, it can not be directly used as an IV in straightforward IV or 2SLS approaches. A possible way to use $P_nS_n(\lambda)Y_n$ for estimation is in the framework of GMM:

$$\min_{\lambda} g_n^2(\lambda), \tag{2.3}$$

where

$$g_n(\lambda) = Y_n'S_n'(\lambda)P_nS_n(\lambda)Y_n. \tag{2.4}$$

Alternatively, the GMM estimator $\hat{\lambda}_n$ may be solved from the quadratic equation $g_n(\hat{\lambda}_n) = 0$. Because of the quadratic expression of $g_n(\lambda)$ in (2.4), one may replace P_n by its symmetric counterpart $\frac{1}{2}P_n^s$ or, simply P_n^s, where $P_n^s = P_n + P_n'$ to arrive at the same moment equation and the same GMM estimate.[1]

Consider the identification problem of λ_0 in the GMM estimation framework with the moment function in (2.4). As the moment equation $E(g_n(\lambda)) = 0$ for a given P_n is a quadratic function of λ, it may have two distinct roots. Because $S_n(\lambda) = S_n + (\lambda_0 - \lambda)W_n$ and $S_n(\lambda)S_n^{-1} = I_n + (\lambda_0 - \lambda)G_n$,

$$\begin{aligned} E(g_n(\lambda)) &= E(\epsilon_n' S_n'^{-1} S_n'(\lambda)P_nS_n(\lambda)S_n^{-1}\epsilon_n) \\ &= \sigma_0^2 tr(S_n'^{-1}S_n'(\lambda)P_nS_n(\lambda)S_n^{-1}) \\ &= \sigma_0^2 tr[(I_n + (\lambda_0 - \lambda)G_n')P_n(I_n + (\lambda_0 - \lambda)G_n)] \\ &= \sigma_0^2[(\lambda_0 - \lambda)tr(P_n^sG_n) + (\lambda_0 - \lambda)^2 tr(G_n'P_nG_n)] \end{aligned} \tag{2.5}$$

[1] If a consistent estimate $\tilde{\lambda}_n$ of λ_0 is available, one might try to use the estimated $P_nS_n(\tilde{\lambda}_n)Y_n$ as an IV in a straightforward IV or 2SLS estimation approach. It can be shown that the resulted IV estimator can be consistent. However, the asymptotic distribution of such an IV estimator will depend on the asymptotic distribution of the initial consistent estimator $\tilde{\lambda}_n$. The resulted IV estimator may or may not have improved efficiency over the initial consistent estimator. This is so because the derivative of $P_nS_n(\lambda)Y_n$ with respect to λ is $-P_nW_nY_n$, which is correlated with ϵ_n. The latter does not provide an orthogonality condition which is needed in order to eliminate the influence of the asymptotic distribution of an initial estimate $\tilde{\lambda}_n$ on the resulted IV estimator.

by using $tr(P_n) = 0$. The moment equation $E(g_n(\lambda)) = 0$ has two roots λ_1 and λ_2 with $\lambda_1 = \lambda_0$ and $\lambda_2 = \lambda_0 + \frac{tr(P_n^s G_n)}{tr(G_n' P_n G_n)}$ if $tr(G_n' P_n G_n) \neq 0$. Because $S_n^{-1} = I_n + \lambda_0 G_n$, the second root can be rewritten explicitly as $\lambda_2 = \frac{tr[(P_n' + S_n'^{-1} P_n) G_n]}{tr(G_n' P_n G_n)}$. From (2.5), there will be a unique root λ_0 if either $tr(G_n' P_n G_n) = 0$ and $tr(P_n^s G_n) \neq 0$, or $tr(G_n' P_n G_n) \neq 0$ and $tr(P_n^s G_n) = 0$. The condition $tr(P_n^s G_n) \neq 0$ is equivalent to the nonzero correlation of the IV $P_n^s S_n Y_n$ with $W_n Y_n$ because $E[(P_n^s S_n Y_n)' W_n Y_n] = E(\epsilon_n' P_n^s G_n \epsilon_n) = \sigma_0^2 tr(P_n^s G_n)$. In general, $tr(G_n' P_n G_n)$ would not necessarily be zero and there may be two distinct roots. In terms of the empirical moment equation $g_n(\lambda) = 0$, because $Y_n' S_n'(\lambda) P_n S_n(\lambda) Y_n = Y_n' P_n Y_n - \lambda Y_n' P_n^s W_n Y_n + \lambda^2 Y_n' W_n' P_n W_n Y_n$ is a quadratic function of λ, the explicit solutions of the empirical moment equation are

$$\hat{\lambda}_n = \{Y_n' P_n^s W_n Y_n \pm [(Y_n' P_n^s W_n Y_n)^2 - 4(Y_n' W_n' P_n W_n Y_n)(Y_n' P_n Y_n)]^{1/2}\}/(2Y_n' W_n' P_n W_n Y_n). \qquad (2.6)$$

In order to distinguish the consistent root from the inconsistent one in (2.6), extra information is necessary. The extra information needed is the sign of $tr(P_n^s G_n)$ as shown below in Proposition 2.2, i.e., the sign of the correlation of the IV $P_n^s S_n Y_n$ and $W_n Y_n$.

In order to justify rigorously possible asymptotic properties of the estimators, some regularity conditions in addition to the structure of the model in (2.1) will be assumed.

Assumption 1: *The ϵ_{nj}s are i.i.d.(0,σ^2) and its moments of order higher than the fourth exist.*

Assumption 2: *The weights matrices $\{W_n\}$ are uniformly bounded in both row and column sums. The elements of $W_n = (w_{n,ij})$ are of order $O(\frac{1}{h_n})$ uniformly in i and j.*

Assumption 3: *The matrices $\{S_n^{-1}\}$ are uniformly bounded in both row and column sums.*

The fourth or higher moments of ϵ_{nj} exist so that the variances of quadratic forms of ϵ_n in this model can be finite. The uniform boundedness assumptions on W_n and S_n^{-1} are originated in a series of papers by Kelejian and Prucha, see, e.g., Kelejian and Prucha (1998), in order to limit correlation across spatial units in a manageable degree. The uniform boundedness of matrices is equivalent to the boundedness of a sequence of norms of matrices. The sequence of square matrices $\{A_n\}$ is uniformly bounded in row sums (resp. column sums) if and only if the sequence $\{\| A_n \|\}$ where $\| \cdot \|$ is the row sums matrix norm (resp. colum sums matrix norm) is bounded (Horn and Johnson 1985). Any matrix norm $\| \cdot \|$ has the property that $\| A_n B_n \| \leq \| A_n \| \cdot \| B_n \|$. So, it holds immediately that the product of two matrices A_n and B_n which are uniformly bounded in row sums (resp. column sums) will be uniformly bounded in row sums (resp. columns sums).[2] The order $O(\frac{1}{h_n})$ of elements of W_n in Assumption 1 has been considered in Lee (1999b).

[2] These particular norms have some other useful properties that other matrix norms might not have. For

It provides explicit features on how the spatial weights matrix W_n shall expand as spatial units increase. The elements of S_n^{-1} in Assumption 3 do not have the order $O(\frac{1}{h_n})$.[3] However, Lemma A.1 implies that elements of $G_n = W_n S_n^{-1}$ have the uniform order $O(\frac{1}{h_n})$ because W_n has and S_n^{-1} is uniformly bounded in column sums. The constant matrices P_ns will be selected to have similar properties of W_n.

Assumption 4: *The constant matrices $\{P_n\}$ with either a zero diagonal or $tr(P_n) = 0$ are uniformly bounded in both row and column sums. The elements of $P_n = (p_{n,ij})$ are of order $O(\frac{1}{h_n})$ uniformly in i and j.*

The class consisting of matrix P_n that satisfies Assumption 4 and has $tr(P_n) = 0$ will be denoted as $\mathcal{P}_{1n}$. The class of matrix P_n that satisfies Assumption 4 but has a zero diagonal will be denoted by $\mathcal{P}_{2n}$. Because a matrix P_n with a zero diagonal has $tr(P_n) = 0$, $\mathcal{P}_{2n}$ is a subclass of $\mathcal{P}_{1n}$.

Assumption 5: *The $\{h_n\}$ can be a bounded or a divergent sequence with $\lim_{n\to\infty} \frac{h_n}{n} = 0$.*

The above assumption allows h_n to diverge to infinity but at a rate slower than the rate n. This assumption includes spatial models with spatial interactions for a unit with only a few of its (near) neighbors as well as interactions with a large number of neighbors. The latter includes spatial specifications in Case (1991, 1992). If h_n is divergent to infinity at the rate n, one can give an example that the GMM estimator may be inconsistent. The same phenomenon occurs for the ML or QMLE estimator (Lee 1999b).

Proposition 2.2 *Assuming that $\lim_{n\to\infty} \frac{h_n}{n} tr(P_n^s G_n) \neq 0$, if the sign of $tr(P_n^s G_n)$, where $P_n \in \mathcal{P}_{1n}$, were positive, the consistent root would be*

$$\hat{\lambda}_n = \{Y_n' P_n^s W_n Y_n - [(Y_n' P_n^s W_n Y_n)^2 - 4(Y_n' W_n' P_n W_n Y_n)(Y_n' P_n Y_n)]^{1/2}\}/(2Y_n' W_n' P_n W_n Y_n); \tag{2.7}$$

if $tr(P_n^s G_n)$ were negative, the consistent root would be

$$\hat{\lambda}_n = \{Y_n' P_n^s W_n Y_n + [(Y_n' P_n^s W_n Y_n)^2 - 4(Y_n' W_n' P_n W_n Y_n)(Y_n' P_n Y_n)]^{1/2}\}/(2Y_n' W_n' P_n W_n Y_n). \tag{2.8}$$

when $\lim_{n\to\infty} \frac{h_n}{n} tr(G_n' P_n G_n) \neq 0$. In the event that $\lim_{n\to\infty} \frac{h_n}{n} tr(G_n' P_n G_n) = 0$, $\hat{\lambda}_n = Y_n' P_n Y_n / Y_n' P_n^s W_n Y_n$ is the unique consistent root.

Unfortunately, because G_n involves the unknown parameter λ_0, one may not, in general, be able to determine the sign of $tr(P_n^s G_n)$. However, if an initial consistent estimate of λ_0 is available, the sign of $tr(P_n^s G_n)$ can

example, if x_n is a column vector with uniformly bounded elements, then $\{\| x_n \|\}$ is bounded with the row sum norm. This is not so with the Euclidian norm.

[3] For example, at $\lambda_0 = 0$, S_n^{-1} is the identity matrix I_n.

be estimated. More on this will be discussed later. In particular, there is an interest in selecting P_n^s closely related to G_n.

Instead of investigating each of the roots as in Proposition 2.2, the following proposition shows that the GMM estimator $\hat{\lambda}_n$ is locally consistent. In order to show that the objective function of the GMM can uniformly converge in probability to a well defined limiting function, we assume as usual for nonlinear estimation in the GMM framework that the parameter space of λ_0 is a compact set.

Assumption 6: *The parameter space Λ of λ is a compact set of the real line with λ_0 in its interior.*

In this assumption, the range of Λ does not need to be specific but the true parameter λ_0 has to satisfy Assumption 3. In the literature, it is quite common to specify the range of λ to be $(-1,1)$ when $W_{n,n}$ is row-normalized, as it guarantees that $S_n(\lambda)^{-1}$ exist whenever $\lambda \in (-1,1)$ and $S_n^{-1}(\lambda)$ can have a series expansion in terms of the powers of W_n (see, e.g., Anselin 1988). Our GMM estimation framework imposes less restrictive assumptions on the parameter space Λ.

Proposition 2.3 *Suppose that $\lim_{n\to\infty} \frac{h_n}{n} tr(P_n^s G_n) \neq 0$, then the GMM estimator $\hat{\lambda}_n$ derived from $\min_{\lambda \in \Lambda_0} g_n^2(\lambda)$ for some small neighborhood Λ_0 of λ_0, is a consistent estimator of λ_0.*

The asymptotic distribution of $\hat{\lambda}_n$ can be derived from a Taylor expansion of $g_n(\hat{\lambda}_n)$ at λ_0. The first and second order derivatives of $g_n(\lambda)$ are

$$\frac{dg_n(\lambda)}{d\lambda} = -Y_n' S_n'(\lambda) P_n^s W_n Y_n, \qquad \frac{d^2 g_n(\lambda)}{d\lambda^2} = Y_n' W_n' P_n^s W_n Y_n. \tag{2.9}$$

The asymptotic distribution of the consistent root $\hat{\lambda}_n$ is in the following proposition. In order for the central limit theorem of a quadratic form in the Appendix to be applicable, Assumption 5) needs to be strengthened.

Assumption 5′: $\lim_{n\to\infty} \frac{h_n^{1+\frac{2}{\delta}}}{n} = 0$ *where $\delta > 0$ such that $E(|\epsilon|^{4+2\delta})$ exists.*

In the event that ϵ has moments of any finite order, δ can be taken to be arbitrarily large. For those cases, Assumptions 5′ is only slightly stronger than Assumption 5.

Proposition 2.4 *Let $P_n \in \mathcal{P}_{1n}$. The consistent root $\hat{\lambda}_n$ from $\min_{\lambda\in\Lambda} g_n^2(\lambda)$ has the asymptotic distribution that*

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \xrightarrow{D} N\left(0, \lim_{n\to\infty}\left[(\kappa_4 - 3)\frac{\sum_{i=1}^n p_{n,ii}^2}{\frac{h_n}{n} tr^2(P_n^s G_n)} + \frac{tr(P_n P_n^s)}{\frac{h_n}{n} tr^2(P_n^s G_n)}\right]\right), \tag{2.10}$$

where $\kappa_4 = \frac{\mu_4}{\sigma_0^4}$ is the kurtosis of ϵ_{ni}.

If (i) $\epsilon_n \sim N(0, \sigma_0^2 I_n)$ or (ii) $P_n \in \mathcal{P}_{2n}$ or (iii) $\lim_{n\to\infty} h_n = \infty$, then

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \xrightarrow{D} N\left(0, \lim_{n\to\infty} \frac{tr(P_n P_n^s)}{\frac{h_n}{n} tr^2(P_n^s G_n)}\right). \tag{2.11}$$

The rate of convergence of the GMM estimator $\hat{\lambda}_n$ depends on h_n. If $\{h_n\}$ is a bounded sequence, it converges in probability to λ_0 at the usual $\sqrt{n}$-rate. When $\{h_n\}$ is a divergent sequence, its rate of convergence can be lower than the $\sqrt{n}$-rate. These rates of convergence match those of the QMLE in Lee (1999b).

The literature on GMM estimation is silent on the problem of possible multiple roots of moment equations. It assumes that the moment equations have a unique root. Because of nonlinearity in the objective function of the GMM method, it is in general difficult to analytically check whether the moment equations have a unique root or not. For our model, because the nonlinearity is only quadratic, it is relatively easier to reveal the multiple roots issue. To overcome this difficulty, a possible strategy is to employ a few more functionally independent moment equations. For our problem, even though each moment equation might have two distinct roots, the common solution set of distinct moment equations may be a singleton. Suppose that P_{1n} and P_{2n} are two distinct $n \times n$ constant matrices from $\mathcal{P}_{1n}$. The two corresponding moment equations will be $E(Y_n' S_n'(\lambda) P_{1n} S_n(\lambda) Y_n) = 0$ and $E(Y_n' S_n'(\lambda) P_{2n} S_n(\lambda) Y_n) = 0$. The inconsistent root of the first moment equation has the value $\lambda_0 + tr(P_{1n}^s G_n)/tr(G_n' P_{1n} G_n)$ and the one of the second equation has $\lambda_0 + tr(P_{2n}^s G_n)/tr(G_n' P_{2n} G_n)$. Thus, if $\frac{tr(P_{1n}^s G_n)}{tr(G_n' P_{1n} G_n)} \neq \frac{tr(P_{2n}^s G_n)}{tr(G_n' P_{2n} G_n)}$, the common root of the two moment equations will be the unique λ_0. Identification of the SAR process in GMM estimation framework can thus be achieved when distinctive moment conditions are employed. In practice, specific IV matrices from $\mathcal{P}_{1n}$ or $\mathcal{P}_{2n}$ can be constructed from the spatial weights matrix W_n, for examples, W_n itself, $\left(W_n' W_n - \frac{tr(W_n' W_n)}{n} I_n\right)$, and $\left(W_n^2 - \frac{tr(W_n^2)}{n} I_n\right)$, etc. The selection of W_n and $\left(W_n' W_n - \frac{tr(W_n' W_n)}{n} I_n\right)$ is related to the MOM in Kelejian and Prucha (1999a) as discussed in a subsequent paragraph.

Suppose that $P_{1n}, \cdots, P_{mn}$ are m distinct constant square matrices of dimension n from $\mathcal{P}_{1n}$. The set of IV functions can be $P_{jn} S_n(\lambda) Y_n$, $j = 1, \cdots, m$. With these IV functions,

$$g_n(\lambda) = (Y_n' S_n'(\lambda) P_{1n} S_n(\lambda) Y_n, \cdots, Y_n' S_n'(\lambda) P_{mn} S_n(\lambda) Y_n)'. \tag{2.12}$$

As in the general GMM framework, these moment equations can be combined into a smaller set of equations by a constant matrix a_n and a GMM estimator can be derived from the minimization problem: $\min_\lambda g_n'(\lambda) a_n' a_n g_n(\lambda)$. The asymptotic distribution of the GMM estimator $\hat{\lambda}_n$ can be derived from the Taylor expansion

$$\hat{\lambda}_n - \lambda_0 = -\left(\frac{\partial g_n'(\hat{\lambda}_n)}{\partial \lambda} a_n' a_n \frac{\partial g_n(\bar{\lambda}_n)}{\partial \lambda} \right)^{-1} \frac{\partial g_n'(\hat{\lambda}_n)}{\partial \lambda} a_n' a_n g_n(\lambda_0). \tag{2.13}$$

The asymptotic distribution of $\hat{\lambda}_n$ will involve the variance of $g_n(\lambda_0) = (\epsilon_n' P_{1n}\epsilon_n, \cdots, \epsilon_n' P_{mn}\epsilon_n)'$. For any square matrix A of dimension n, let $vec_D(A) = (a_{11}, \cdots, a_{nn})'$ denote the vector formed by the diagonal elements of A. Furthermore, let $\text{Diag}(A) = diag(a_{11}, \cdots, a_{nn})$ be the $n \times n$ diagonal matrix associated with the diagonal elements of A.

Lemma 2.1 *For any two square matrices A and B of dimension n,*

$$E(\epsilon_n' A\epsilon_n \cdot \epsilon_n' B\epsilon_n) = (\mu_4 - 3\sigma_0^4) vec_D'(A) vec_D(B) + \sigma_0^4[tr(A)tr(B) + tr(AB^s)],$$

where $\mu_4 = E(\epsilon_{ni}^4)$ is the fourth moment of ϵ_{ni}.

If (i) both A and B are matrices with zero diagonals, or (ii) $tr(A) = tr(B) = 0$ and $\epsilon_n \sim N(0, \sigma_0^2 I_n)$, then $E(\epsilon_n' A\epsilon_n \cdot \epsilon_n' B\epsilon_n) = \sigma_0^4 tr(AB^s) = \sigma_0^4 tr(BA^s)$.

The variance matrix of $g_n(\lambda_0)$ can be derived with the results in Lemma 2.1. If (i) the matrices P_{jn}, $j = 1, \cdots, m$, are from $\mathcal{P}_{2n}$, or (ii) $\epsilon_n \sim N(0, \sigma_0^2 I_n)$ with $P_{jn} \in \mathcal{P}_{1n}$ for all $j = 1, \cdots, m$, then $\text{var}(g_n(\lambda_0)) = \sigma_0^4 V_n$, where

$$V_n = \begin{pmatrix} tr(P_{1n}P_{1n}^s) & \cdots & tr(P_{1n}P_{mn}^s) \\ \vdots & \vdots & \vdots \\ tr(P_{mn}P_{1n}^s) & \cdots & tr(P_{mn}P_{mn}^s) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} tr(P_{1n}^s P_{1n}^s) & \cdots & tr(P_{1n}^s P_{mn}^s) \\ \vdots & \vdots & \vdots \\ tr(P_{mn}^s P_{1n}^s) & \cdots & tr(P_{mn}^s P_{mn}^s) \end{pmatrix}. \quad (2.14)$$

The second expression in (2.14) follows from the identity that $tr(P_{1n}P_{2n}^s) = \frac{1}{2} tr(P_{1n}^s P_{2n}^s)$. In general, $\text{var}(g_n(\lambda_0)) = \sigma_0^4 \Omega_n$ where

$$\Omega_n = (\kappa_4 - 3)[vec_D(P_{1n}), \cdots, vec_D(P_{mn})]'[vec_D(P_{1n}), \cdots, vec_D(P_{mn})] + V_n. \quad (2.15)$$

For any two conformable matrices A and B, it is obvious that $tr(AB) = vec'(A')vec(B)$. The V_n in (2.14) can be rewritten as $V_n = \frac{1}{2}(vec(P_{1n}^s), \cdots, vec(P_{mn}^s))'(vec(P_{1n}^s), \cdots, vec(P_{mn}^s))$. The V_n is nonsingular as long as P_ns are chosen so that $vec(P_{jn}^s)$ for $j = 1, \cdots, m$ are linearly independent. This is so also for Ω_n.[4]

Proposition 2.5 *Suppose P_{jn}, $j = 1, \cdots, m$, are from $\mathcal{P}_{1n}$ so that $a_0 \lim_{n\to\infty} \frac{h_n}{n} E(g_n(\lambda)) = 0$ has a unique root at λ_0 in Λ, where a_n converges to a_0. Then, the GMM estimator $\hat{\lambda}_n$ derived from the minimization $\min_{\lambda \in \Lambda} g_n'(\lambda) a_n' a_n g_n(\lambda)$ is a consistent estimator of λ_0, and $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \xrightarrow{D} N(0, \Sigma)$, where*

$$\Sigma = \lim_{n\to\infty} [(\frac{h_n}{n} d_n)' a_0' a_0 (\frac{h_n}{n} d_n)]^{-1} (\frac{h_n}{n} d_n)' a_0' a_0 (\frac{h_n}{n} \Omega_n) a_0' a_0 (\frac{h_n}{n} d_n) [(\frac{h_n}{n} d_n)' a_0' a_0 (\frac{h_n}{n} d_n)]^{-1}, \quad (2.16)$$

with $d_n = (tr(P_{1n}^s G_n), \cdots, tr(P_{mn}^s G_n))'$ under the assumption that $a_0 \lim_{n\to\infty} \frac{h_n}{n} d_n \neq 0$.

[4] If Ω_n were singular, there would exist a nonzero vector of constants $\alpha = (\alpha_1, \cdots, \alpha_m)'$ such that $\alpha' g_n(\lambda_0) = 0$ with probability one. That is, $\epsilon_n'(\sum_{j=1}^m \alpha_j P_{jn})\epsilon_n = 0$ almost everywhere for ϵ_n. This would be possible if and only if $\sum_{j=1}^m \alpha_j P_{jn} = 0$, i.e., $vec(P_{jn})$s would be linearly dependent.

From the limiting distribution of $\hat{\lambda}_n$ in Proposition 2.5, the optimal choice of the weighting matrix $a_n' a_n$ is, as usual, the inverse of a matrix proportional to the variance matrix of $g_n(\lambda_0)$. In general, the optimal weighting matrix for $g_n(\lambda_0)$ is the inverse of Ω_n. The matrix V_n^{-1} can be the optimal weighting matrix in special circumstances including that $\epsilon_n \sim N(0, \sigma_0^2 I_n)$ or P_ns are from $\mathcal{P}_{2n}$. A less apparent case is the spatial process with $\lim_{n\to\infty} h_n = \infty$.

Proposition 2.6 *Suppose that the limit of* $\frac{h_n}{n}\Omega_n$ *exists and is a nonsingular matrix, and* $\frac{h_n}{n}(\hat{\Omega}_n - \Omega_n) = o_P(1)$*, then the optimal GMM estimator* $\hat{\lambda}_{v,n}$ *derived from* $\min_{\lambda\in\Lambda} g_n'(\lambda)\hat{\Omega}_n^{-1} g_n(\lambda)$ *based on* $g_n(\lambda)$ *with* P_n*s from* $\mathcal{P}_{1n}$ *has the asymptotic distribution:*

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{v,n} - \lambda_0) \xrightarrow{D} N(0, \sigma_0^4 \Sigma_v), \tag{2.17}$$

where $\Sigma_v = (\lim_{n\to\infty} \frac{h_n}{n}\Sigma_{vn})^{-1}$ *with* $\Sigma_{vn} = \big(tr(P_{1n}^s G_n), \cdots, tr(P_{mn}^s G_n)\big)\Omega_n^{-1}\big(tr(P_{1n}^s G_n), \cdots, tr(P_{mn}^s G_n)\big)'$, *assuming that the limit of* $\frac{h_n}{n}\Sigma_{vn}$ *exists and is nonzero. Furthermore,*

$$g_n'(\hat{\lambda}_{v,n})\hat{\Omega}_n^{-1} g_n(\hat{\lambda}_{v,n}) \xrightarrow{D} \sigma_0^4 \chi^2(m-1). \tag{2.18}$$

For the special cases that (i) P_{jn}*,* $j = 1, \cdots, m$*, are from* $\mathcal{P}_{2n}$*, or (ii)* $\epsilon_n \sim N(0, \sigma_0^2 I_n)$*, or (iii)* $\lim_{n\to\infty} h_n = \infty$*, then* V_n *can be used as the weighting matrix in place of* $\hat{\Omega}_n$ *for the optimal GMM estimation.*

A nice feature of V_n in (2.14) is that it does not involve any unknown parameter of the model and the computation of the GMM with V_n^{-1} as its weighting matrix does not require a two step procedure as in a typical optimal GMM estimation. On the other hand, in order to use the general weighting matrix Ω_n, the moment parameters μ_4 and σ_0^2 need to be estimated. The λ_0 can be estimated by the GMM without weighting in an initial step. The initial GMM estimate can be used to estimate the disturbances of ϵ_n. The moments σ_0^2 and μ_4 can then be estimated by corresponding empirical moments using the estimated residuals. Alternatively, one may select P_ns from $\mathcal{P}_{2n}$ and use the corresponding optimal GMM with V_n^{-1} as the weighting matrix to obtain an initial estimate of λ_0. The estimated $\hat{\Omega}_n$ can be used as the optimal feasible weighting matrix. The use of V_n does not require these steps. The corresponding V_n gives the optimal weighting matrix for the moments $g_n(\lambda)$ when the P_{jn}s are selected from $\mathcal{P}_{2n}$, a subclass of $\mathcal{P}_{1n}$. If P_{jn}s are selected from the broader class $\mathcal{P}_{1n}$, the corresponding V_n could provide the optimal weighting matrix only when ϵ_n has the moment restriction $\mu_4 = 3\sigma_0^4$, which includes the normal distributional case, or when $\lim_{n\to\infty} h_n = \infty$. With the optimal weighting matrix, the minimized objective function in (2.18) is asymptotically χ^2 distributed with $(m-1)$ degree of freedom, which provides a goodness-of-fit diagnostic test for the spatial model when $m > 1$.

The computation of the GMM estimators in Propositions 2.5 and 2.6 is essentially that of the nonlinear least squares (NLS). Consider the GMM estimation in Proposition 2.5. From (2.12),

$$a_n g_n(\lambda) = \sum_{l=1}^m a_{nl} Y_n' S_n'(\lambda) P_{ln} S_n(\lambda) Y_n = \sum_{l=1}^m a_{nl} Y_n' P_{ln} Y_n - \sum_{l=1}^m a_{nl} Y_n' W_n' P_{ln}^s Y_n \cdot \lambda + \sum_{l=1}^m a_{nl} Y_n' W_n' P_{ln} W_n Y_n \cdot \lambda^2.$$

The GMM estimation is equivalent to the NLS estimation of the following nonlinear-in-parameters regression equation:

$$\sum_{l=1}^m a_{nl} Y_n' P_{ln} Y_n = \sum_{l=1}^m a_{nl} Y_n' W_n' P_{ln}^s Y_n \cdot \lambda - \sum_{l=1}^m a_{nl} Y_n' W_n' P_{ln} W_n Y_n \cdot \lambda^2 + \xi_k,$$

where ξ_k is the k-dimensional vector of equation residuals (disturbances). For the optimum GMM estimation in Proposition 2.6, as

$$g_n(\lambda) = \begin{pmatrix} Y_n' S_n'(\lambda) P_{1n} S_n(\lambda) Y_n \\ \vdots \\ Y_n' S_n'(\lambda) P_{mn} S_n(\lambda) Y_n \end{pmatrix} = \begin{pmatrix} Y_n' P_{1n} Y_n \\ \vdots \\ Y_n' P_{mn} Y_n \end{pmatrix} - \begin{pmatrix} Y_n' W_n' P_{1n}^s Y_n \\ \vdots \\ Y_n' W_n' P_{mn}^s Y_n \end{pmatrix} \lambda + \begin{pmatrix} Y_n' W_n' P_{1n} W_n Y_n \\ \vdots \\ Y_n' W_n' P_{mn} W_n Y_n \end{pmatrix} \lambda^2,$$

it is equivalent to the generalized nonlinear least squares (GNLS) estimation of the nonlinear-in-parameter equation:

$$\begin{pmatrix} Y_n' P_{1n} Y_n \\ \vdots \\ Y_n' P_{mn} Y_n \end{pmatrix} = \begin{pmatrix} Y_n' W_n' P_{1n}^s Y_n \\ \vdots \\ Y_n' W_n' P_{mn}^s Y_n \end{pmatrix} \lambda - \begin{pmatrix} Y_n' W_n' P_{1n} W_n Y_n \\ \vdots \\ Y_n' W_n' P_{mn} W_n Y_n \end{pmatrix} \lambda^2 + \xi_m,$$

where ξ_m is a m-dimensional vector of residual with variance matrix Ω_n.[5]

The selection of P_ns for IV functions requires them to be matrices either from $\mathcal{P}_{1n}$ or $\mathcal{P}_{2n}$, and be correlated with G_n in that $tr(P_n^s G_n) \neq 0$. Other than those, the selection of P_ns can be arbitrary. As the asymptotic variance of the GMM estimator $\hat{\lambda}_n$ depends on the selected P_ns. The possible best selection of P_n is an interesting issue. Intuitively, one should choose a P_n so that its correlation with G_n be maximized. One can not use G_n directly for P_n because G_n may neither have a zero diagonal nor a zero trace. Instead of G_n, possible candidates may be $(G_n - \frac{tr(G_n)}{n} I_n)$ or $(G_n - Diag(G_n))$, which are modified from G_n so that they are in either $\mathcal{P}_{1n}$ or $\mathcal{P}_{2n}$. The following proposition shows that $(G_n - \frac{tr(G_n)}{n} I_n)$ and $(G_n - \text{Diag}(G_n))$ are also relevant in the evaluation of the trace of the product of P_n^s and G_n.

Lemma 2.2 *Suppose that A and B are two $n \times n$ matrices.*

[5] By casting the GMM estimation in a NLS framework, one may also address the identification of λ_0 via those nonlinear (moment) equations. The identification of parameters in the MOM in Kelejian and Prucha (1999a) is addressed via a least square estimation.

(i) If $tr(A)=0$*, then* $tr(AB)=tr[A(B-\frac{tr(B)}{n}I_n)]$.

(ii) If $\mathrm{Diag}(A)=0$*, then* $tr(AB)=tr[A(B-Diag(B))]$.

This lemma states that when A is a square matrix with $tr(A)=0$, a conformable matrix B in the product AB can be replaced by $(B-\frac{tr(B)}{n}I_n)$ without changing the value $tr(AB)$. Similarly, if A is a square matrix with a zero diagonal, B can be replaced by $(B-Diag(B))$ without changing the value $tr(AB)$. This lemma implies that if $P_n\in\mathcal{P}_{1n}$, $tr(P_n^sG_n)=tr(P_n^s(G_n-\frac{tr(G_n)}{n}I_n))$, and if $P_n\in\mathcal{P}_{2n}$, $tr(P_n^sG_n)=tr(P_n^s(G_n-Diag(G_n)))$. The following proposition confirms that $(G_n-\frac{tr(G_n)}{n}I_n)$ is the optimal matrix within the class of matrices $\mathcal{P}_{1n}$, and $(G_n-Diag(G))$ is optimal in $\mathcal{P}_{2n}$. They are optimal in the sense of maximizing the scalar correlation coefficient of $P_n^sS_nY_n$ and W_nY_n within their relevant classes. As $P_n^sS_nY_n=P_n^s\epsilon_n$ and $W_nY_n=G_n\epsilon_n$, they have zero means and their scalar variances are, respectively, $E(\epsilon_n'P_n^{2s}\epsilon_n)=\sigma_0^2tr(P_n^{2s})$ and $E(\epsilon_n'G_n'G_n\epsilon_n)=\sigma_0^2tr(G_n'G_n)$. Hence, the squared scalar correlation coefficient of $P_n^sS_nY_n$ and W_nY_n is $r_n^2=\frac{tr^2(P_n^sG_n)}{tr(P_n^{2s})tr(G_n'G_n)}$.

Proposition 2.7 *(i) In the class of constant matrices* $\mathcal{P}_{1n}$*,*

$$\max_{P_n\in\mathcal{P}_{1n}}\frac{tr^2(P_n^sG_n)}{tr(P_n^{s2})}=\frac{1}{4}\max_{P_n\in\mathcal{P}_{1n}}\frac{tr^2(P_n^sG_n^s)}{tr(P_n^{s2})}=\frac{1}{4}\frac{tr^2[(G_n-\frac{tr(G_n)}{n}I_n)^sG_n^s]}{tr[(G_n-\frac{tr(G_n)}{n}I_n)^{s2}]}=\frac{1}{4}tr[(G_n-\frac{tr(G_n)}{n}I_n)^{s2}],$$

and, (ii) in the class of constant matrices $\mathcal{P}_{2n}$*,*

$$\max_{P_n\in\mathcal{P}_{2n}}\frac{tr^2(P_n^sG_n)}{tr(P_n^{2s})}=\frac{1}{4}\max_{P_n\in\mathcal{P}_{2n}}\frac{tr^2(P_n^sG_n^s)}{tr(P_n^{2s})}=\frac{1}{4}\frac{tr^2[(G_n-Diag(G_n))^sG_n^s]}{tr[(G_n-Diag(G_n))^{s2}]}=\frac{1}{4}tr[(G_n-Diag(G_n))^{s2}].$$

The intuition that selecting P_n to maximize the correlation of $P_nS_nY_n$ with W_nY_n within the relevant class may provide best IV estimate is confirmed in the following proposition. Let $\mathcal{M}_{1n}=\{\hat\lambda_{v,n}\}$ (resp., $\mathcal{M}_{2n}$) be the class of optimal GMM estimators derived from $\min_{\lambda\in\Lambda}g_n'(\lambda)\Omega_n^{-1}g_n(\lambda)$ (resp., $\min_{\lambda\in\Lambda}g_n'(\lambda)V_n^{-1}g_n(\lambda)$), where $g_n(\lambda)$ is a vector of moments functions with P_ns from $\mathcal{P}_{1n}$ (resp., $\mathcal{P}_{2n}$).

Proposition 2.8 *Within the class of optimal GMM estimators* $\mathcal{M}_{2n}$*, the best estimator is the consistent root* $\hat\lambda_{2b,n}$ *derived from* $\min_{\lambda\in\Lambda}[Y_n'S_n'(\lambda)(G_n-Diag(G_n))S_n(\lambda)Y_n]^2$ *in the sense that* $\sqrt{\frac{n}{h_n}}(\hat\lambda_{2b,n}-\lambda_0)\xrightarrow{D}N(0,\Sigma_{2b})$ *with* $\Sigma_{2b}\le\Sigma_v$*, where* Σ_v *is the limiting variance matrix of* $\sqrt{\frac{n}{h_n}}(\hat\lambda_{v,n}-\lambda_0)$ *in Proposition* 2.6 *and*

$$\Sigma_{2b}=(\lim_{n\to\infty}\frac{h_n}{n}tr[(G_n-Diag(G_n))^sG_n])^{-1}. \tag{2.19}$$

In the event that $\epsilon_n\sim N(0,\sigma_0^2I_n)$ *or for the case that* $\lim_{n\to\infty}h_n=\infty$*, within the broader class of estimators* $\mathcal{M}_{1n}$*, the consistent root* $\hat\lambda_{1b,n}$ *derived from* $\min_{\lambda\in\Lambda}[Y_n'S_n'(\lambda)(G_n-\frac{tr(G_n)}{n}I_n)S_n(\lambda)Y_n]^2$ *is the best GMM estimator with* $\sqrt{\frac{n}{h_n}}(\hat\lambda_{1b,n}-\lambda_0)\xrightarrow{D}N(0,\Sigma_{1b})$ *where*

$$\Sigma_{1b}=(\lim_{n\to\infty}\frac{h_n}{n}tr[(G_n-\frac{tr(G_n)}{n}I_n)^sG_n])^{-1}. \tag{2.20}$$

The estimate $\hat{\lambda}_{2b,n}$ is optimal within the class $\mathcal{P}_{2n}$ regardless of the distribution of ϵ_n. For the special cases that ϵ_n is normally distributed or $\{h_n\}$ is a divergent sequence, $\hat{\lambda}_{1b,n}$ is optimal within the broader class of $\mathcal{P}_{1n}$ and may relatively be more efficient than $\hat{\lambda}_{2b,n}$. The relative efficiency of $\hat{\lambda}_{1b,n}$ over $\hat{\lambda}_{2b,n}$ can be quantified by comparing their precision matrices:

$$
\begin{aligned}
&\frac{h_n}{n}\{tr[(G_n - \frac{tr(G_n)}{n}I_n)^s G_n] - tr[(G_n - Diag(G_n))^s G_n]\} \\
&= 2\frac{h_n}{n} tr[(Diag(G_n) - \frac{tr(G_n)}{n}I_n)G_n] = 2\frac{h_n}{n} tr[(Diag(G_n) - \frac{tr(G_n)}{n}I_n)Diag(G_n)] \\
&= 2\frac{h_n}{n}\sum_{j=1}^{n}(G_{n,jj} - \frac{\sum_{i=1}^{n} G_{n,ii}}{n})^2.
\end{aligned}
$$

From this, for $\{h_n\}$ being a bounded sequence, $\hat{\lambda}_{1b,n}$ is more precise as it takes into account the variance of the diagonal elements of G_n. The difference of the precision matrices is two times of the empirical variance of the diagonal elements of G_n.[6] The empirical variance is zero only for cases where the diagonal elements of G_n are identical. For the case that $\lim_{n\to\infty} h_n = \infty$, the difference shall vanish as n goes to infinity because $G_{n,ii} = O(\frac{1}{h_n})$ implies that $\frac{h_n}{n}\sum_{j=1}^{n}(G_{n,jj} - \frac{\sum_{i=1}^{n} G_{n,ii}}{n})^2 = \frac{h_n}{n}O(\frac{n}{h_n^2}) = O(\frac{1}{h_n}) = o(1)$. That is, when $\lim_{n\to\infty} h_n = \infty$, $\hat{\beta}_{1b,n}$ and $\hat{\beta}_{2b,n}$ have the same limiting distribution.

However, estimators with P_n from $\mathcal{P}_{2n}$, which includes $\hat{\lambda}_{2n}$, may have some robust properties than those from $\mathcal{P}_{1n}$. The consistency of GMM estimator with P_n from $\mathcal{P}_{1n}$ or $\mathcal{P}_{2n}$ is based on the fundamental moment property that $E(\epsilon_n' P_n \epsilon_n) = 0$ as in proof of Proposition 2.1. This is valid because ϵ_{ni}'s are i.i.d. with zero mean and a common variance. If ϵ_{ni}'s had heteroskedastic variances, $E(\epsilon_n \epsilon_n')$ would be a diagonal matrix not proportional to an identity matrix. In such a case, when P_n is from $\mathcal{P}_{1n}$, $E(\epsilon_n' P_n \epsilon_n) = tr[P_n E(\epsilon_n \epsilon_n')]$ would not necessarily be zero. However, when P_n has a zero diagonal, $tr[P_n E(\epsilon_n \epsilon_n')] = tr[Diag(P_n)E(\epsilon_n \epsilon_n')] = 0$ because $E(\epsilon_n \epsilon_n')$ is a diagonal matrix and $Diag(P_n) = 0$. So it is possible that $\hat{\lambda}_{2b,n}$ may be consistent against unknown heteroskedastic disturbances in the model.[7]

When ϵ_n is not normally distributed or $\{h_n\}$ is a bounded sequence, the GMM estimator $\hat{\lambda}_{1b,n}$ in the preceding proposition may not have any optimal property. However, it is interesting to note that it has the same limiting distribution as the QMLE for the model (2.1).

Proposition 2.9 *The consistent root* $\hat{\lambda}_{1b,n}$ *derived from* $\min_{\lambda\in\Lambda}[Y_n' S_n'(\lambda)(G_n - \frac{tr(G_n)}{n}I_n)S_n(\lambda)Y_n]^2$ *has*

[6] The h_n shall be normalized to one in this case.

[7] A rigorous analysis of the robustness property of this estimator and its associated robust (White's) test statistics is beyond the scope of this paper but it shall be investigated in a separate paper.

the limiting distribution

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{1b,n} - \lambda_0) \xrightarrow{D} N\left(0, \lim_{n\to\infty}\left[(\kappa_4 - 3)\frac{\sum_{i=1}^n (G_{n,ii} - \frac{tr(G_n)}{n})^2}{\frac{h_n}{n} tr^2[(G_n - \frac{tr(G_n)}{n} I_n)^s G_n]} + \frac{1}{\frac{h_n}{n} tr[(G_n - \frac{tr(G_n)}{n} I_n)^s G_n]}\right]\right), \tag{2.21}$$

which is the same as the QMLE $\hat{\lambda}_{QM,n}$ *of* λ_0 *derived from* $\max_{\theta\in\Theta} \ln L_n(\theta)$ *where* $\theta = (\lambda, \sigma^2)$ *and*

$$L_n(\theta) = \frac{|S_n(\lambda)|}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp(-\frac{1}{2\sigma^2} Y_n' S_n'(\lambda) S_n(\lambda) Y_n)$$

based on the normal distributional specification of ϵ_n *in* (2.1).[8]

The best estimate in $\mathcal{M}_{1n}$ (resp., $\mathcal{M}_{2n}$) associated with $(G_n - \frac{tr(G_n)}{n} I_n)$ (resp., $(G_n - Diag(G_n))$) involves the unknown λ_0 in G_n as $G_n = W_n S_n^{-1}$. The unknown λ_0 can be estimated with some P_ns from $\mathcal{P}_{1n}$ or $\mathcal{P}_{2n}$ within the GMM framework. With an initial consistent estimate $\hat{\lambda}_n$, G_n can be estimated by $\hat{G}_n = W_n S_n^{-1}(\hat{\lambda}_n)$. The additional computation for using this feasible best matrix is to obtain the inverse of $S_n(\hat{\lambda}_n)$, which, however, needs to be inverted only once.[9] The following proposition shows that the feasible GMM estimator with G_n replaced by $\hat{G}_n$ in the IV functions has the same limiting distribution as the corresponding best GMM estimator in Proposition 2.8.

To simplify the following presentations, for any $n \times n$ matrix A_n, we shall denote the adjusted matrix $(A_n - \frac{tr(A_n)}{n} I_n)$ or the matrix $(A_n - Diag(A_n))$ by A_n^d.

Proposition 2.10 *Suppose* $\hat{\lambda}_n$ *is a* $\sqrt{\frac{n}{h_n}}$*-consistent estimate of* λ_0*, and* $\hat{G}_n = W_n S_n^{-1}(\hat{\lambda}_n)$.

Then, $\min_{\lambda\in\Lambda} [Y_n' S_n'(\lambda) \hat{G}_n^d S_n(\lambda) Y_n]^2$ *has a consistent root* $\tilde{\lambda}_{b,n}$ *which has the same limiting distribution of* $\hat{\lambda}_{b,n}$ *derived from* $\min_{\lambda\in\Lambda} [Y_n' S_n'(\lambda) G_n^d S_n(\lambda) Y_n]^2$.

The moment equation $Y_n' S_n'(\lambda) \hat{G}_n^d S_n(\lambda) Y_n = 0$ may have two roots. However, the consistent root can be easily identified because the sign of $tr((G_n^s)^d G_n)$ can be determined. This is so, because $tr((G_n^s)^d G_n) = \frac{1}{2} tr((G_n^s)^d (G_n^s)^d) > 0$ whenever $(G_n^s)^d \neq 0$.

Proposition 2.11 *Under the assumption that* $\lim_{n\to\infty} \frac{h_n}{n} tr[(G_n^s)^d (G_n^s)^d] \neq 0$*, the consistent root for the moment equation* $Y_n' S_n'(\lambda) \hat{G}_n^d S_n(\lambda) Y_n = 0$ *is*

$$\tilde{\lambda}_n = \{Y_n'(\hat{G}_n^s)^d W_n Y_n - \left[(Y_n'(\hat{G}_n^s)^d W_n Y_n)^2 - 4Y_n' W_n' (\hat{G}_n)^d W_n Y_n \cdot Y_n' (\hat{G}_n)^d Y_n\right]^{\frac{1}{2}}\}/(2Y_n' W_n' (\hat{G}_n)^d W_n Y_n). \tag{2.22}$$

---

[8] The concentrated log likelihood function of λ is $\ln L_n(\lambda) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\hat{\sigma}_n^2(\lambda) + \ln|S_n(\lambda)|$, where $\hat{\sigma}_n^2(\lambda) = \frac{1}{n} Y_n' S_n'(\lambda) S_n(\lambda) Y_n$. Its associated first order derivative is $\frac{\partial \ln L_n(\lambda)}{\partial\lambda} = \frac{1}{\hat{\sigma}_n^2(\lambda)} Y_n' W_{n,n}' S_n(\lambda) Y_n - tr(W_{n,n} S_n^{-1}(\lambda))$ (see, Lee 1999b). Our moment equation in the GMM framework is not the likelihood equation $\frac{\partial \ln L_n(\lambda)}{\partial\lambda} = 0$.

[9] The Cholesky decomposition, for example, can be an attractive general method. If W_n is a sparse matrix, proper sparse matrix inverse subroutines may be valuable (Page and Barry 1997).

The best estimators in (2.22) are of interest as they have close form expressions. In a finite sample, when the numerical value of $(Y_n'(\hat{G}_n^s)^d W_n Y_n)^2 - 4Y_n' W_n'(\hat{G}_n)^d W_n Y_n \cdot Y_n'(\hat{G}_n)^d Y_n$ is positive, the best estimates are immediate available. In the event that the numerical value under the square root operator in (2.22) is negative, the estimator in (2.22) would take a complex value. In that case, one shall resort to the GMM minimization and the estimate (minimizer) shall take the value $Y_n'(\hat{G}_n^s)^d W_n Y_n / Y_n' W_n'(\hat{G}_n^s)^d W_n Y_n$, which is equivalent to the root in (2.22) by setting the negative value under the square root operator to zero. In any case, the computational burden of the best estimators will mainly be in the computation of an initial consistent estimate by the GMM or NLS and the evaluation of $\hat{S}_n^{-1}$ in $\hat{G}_n$.

Our GMM estimation has focused on the estimation of the spatial parameter λ. With a consistent estimate $\hat{\lambda}_n$ available, the parameter σ^2 can be estimated as $\hat{\sigma}_n^2 = \hat{\epsilon}_n' \hat{\epsilon}_n / n$ where $\hat{\epsilon}_n = S_n(\hat{\lambda}_n) Y_n$ is the estimated residual. By expansion,

$$\hat{\sigma}_n^2 - \sigma_0^2 = (\frac{1}{n}\epsilon_n'\epsilon_n - \sigma_0^2) - (\hat{\lambda}_n - \lambda_0)\frac{\epsilon_n' G_n^s \epsilon_n}{n} + (\hat{\lambda}_n - \lambda_0)^2 \frac{\epsilon_n' G_n' G_n \epsilon_n}{n}. \tag{2.23}$$

The terms $\frac{\epsilon_n' G_n^s \epsilon_n}{n}$ and $\frac{\epsilon_n' G_n' G_n \epsilon_n}{n}$ are of order $O(\frac{1}{h_n})$ by Lemma A.3, and $(\frac{1}{n}\epsilon_n'\epsilon_n - \sigma_0^2) = o_P(1)$ by the law of large numbers. Hence, $\hat{\sigma}_n^2$ is a consistent estimator of σ_0^2. The asymptotic distribution of $\hat{\sigma}_n^2$ may depend on the asymptotic distribution of $\hat{\lambda}_n$ except for the case that $\lim_{n\to\infty} h_n = \infty$. This is so as follows. In general,

$$\begin{aligned}
\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2) &= \frac{1}{\sqrt{n}}(\epsilon_n'\epsilon_n - n\sigma_0^2) - \sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0)\frac{\sqrt{h_n}}{n}\epsilon_n' G_n^s \epsilon_n + (\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0))^2 \cdot \frac{h_n}{\sqrt{n}} \frac{\epsilon_n' G_n' G_n \epsilon_n}{n} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n (\epsilon_{ni}^2 - \sigma_0^2) - \sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \cdot \frac{\sqrt{h_n}}{n}\epsilon_n' G_n^s \epsilon_n + o_P(1),
\end{aligned} \tag{2.24}$$

because $\frac{h_n}{n} = o(1)$ and $\frac{1}{n}\epsilon_n' G_n' G_n \epsilon_n = O(\frac{1}{h_n})$, when $\hat{\lambda}_n$ is $\sqrt{\frac{n}{h_n}}$-consistent. When $\lim_{n\to\infty} h_n = \infty$, the term $\frac{\sqrt{h_n}}{n}\epsilon_n' G_n^s \epsilon_n = O_P(\frac{1}{\sqrt{h_n}}) = o_P(1)$ and, in this case, $\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\epsilon_{ni}^2 - \sigma_0^2) + o_P(1) \xrightarrow{D} N(0, (\mu_4 - \sigma_0^4))$. When $\{h_n\}$ is a bounded sequence, the asymptotic distribution of $\hat{\sigma}_n^2$ will depend on the asymptotic distribution of $\hat{\lambda}_n$ evidenced from (2.24). As the best GMM estimate $\hat{\lambda}_n$ from $\mathcal{P}_{1n}$ has the same asymptotic distribution as the QMLE under the normal distributional specification and the QMLE of σ_0^2 is also equal to the estimated residuals' second moments, $\hat{\sigma}_n^2$ with the best GMM estimate $\hat{\lambda}_n$ will have the same asymptotic distribution of the QMLE of σ_0^2.

Kelejian and Prucha (1999a) have suggested an MOM for the estimation of the SAR model (2.1). The moments used in their estimation are based on the moment properties that $E(\epsilon_n'\epsilon_n) = n\sigma_0^2$, $E(\epsilon_n' W_n' W_n \epsilon_n) =$

$\sigma_0^2 tr(W_n'W_n)$ and $E(\epsilon_n' W_n \epsilon_n) = 0$. The corresponding vector of empirical moments is

$$g_n(\theta) = (Y_n'S_n'(\lambda)S_n(\lambda)Y_n - n\sigma^2, Y_n'S_n'(\lambda)W_n'W_nS_n(\lambda)Y_n - \sigma^2 tr(W_n'W_n), Y_n'S_n'(\lambda)W_nS_n(\lambda)Y_n)'. \quad (2.25)$$

They suggest the estimation of θ by the unweighted MOM: $\min_{\theta\in\Theta} g_n'(\theta)g_n(\theta)$. Kelejian and Prucha (1999a) show that the resulted MOM estimator is consistent. Comparing their MOM approach with our GMM approach, there are some similarities but they are different. The third moment in (2.25) captures the correlation of $W_n\epsilon_n$ and ϵ_n. This moment equation corresponds to a moment equation in our GMM framework by taking W_n for the IV function $W_nS_n(\lambda)Y_n$ as $W_n \in \mathcal{P}_{2n}$. By selecting $(W_n'W_n - \frac{tr(W_n'W_n)}{n}I_n)$ for a moment function in our GMM framework would have some similarities with the second moment equation in (2.25) of Kelejian and Prucha (1999a). The moment function $Y_n'S_n'(\lambda)(W_n'W_n - \frac{tr(W_n'W_n)}{n}I_n)S_n(\lambda)Y_n$ can be written as

$$\begin{aligned} & Y_n'S_n'(\lambda)(W_n'W_n - \frac{tr(W_n'W_n)}{n}I_n)S_n(\lambda)Y_n \\ &= [Y_n'S_n'(\lambda)W_n'W_nS_n(\lambda)Y_n - \sigma^2 tr(W_n'W_n)] - \frac{tr(W_n'W_n)}{n}[Y_n'S_n'(\lambda)S_n(\lambda)Y_n - n\sigma^2], \end{aligned} \quad (2.26)$$

which is a linear combination of the first and second moments in (2.25). The MOM in Kelejian and Prucha (1999a) will jointly estimate λ and ρ. In our GMM framework, the estimation focuses solely on the estimation of λ. The linear combination in (2.26) eliminates the estimation of σ^2 and focuses on the estimation of λ. The σ^2 can be estimated after the estimation of λ via the estimated residuals of ϵ_n. Alternatively, given a value λ, σ^2 can be estimated from the first moment equation implied by (2.25), i.e., $\sigma^2 = \frac{1}{n}Y_n'S_n'(\lambda)S_n(\lambda)Y_n$. Substitute this solution for σ^2 into the second moment in (2.25), one will arrive at the moment function $Y'S_n'(\lambda)W_n'W_nS_n(\lambda)Y_n - \frac{1}{n}Y_n'S_n'(\lambda)S_n(\lambda)Y_n tr(W_n'W_n) = Y_n'S_n'(\lambda)(W_n'W_n - \frac{tr(W_n'W_n)}{n}I_n)S_n(\lambda)Y_n$ in our GMM framework. Such a sequential estimation strategy will slightly simplify the computation as it involves one less parameter in the nonlinear optimization.

Ord (1975) has indicated the possible use of the third moment equation alone in (2.25) to solve for an estimate of λ. He points out that the relative inefficiency of that moment estimate relative to the MLE increases as λ increases and he favors the ML method for estimation. The MOM estimation in Kelejian and Prucha (1999a) uses the additional first two moments in (2.25). Kelejian and Prucha (1999a) have considered the consistency but not the asymptotic distribution of their MOM estimators. Their Monte Carlo experiment has shown efficiency close to those of the QMLEs under a variety of distributions. This is a much better improvement than that of Ord (1975). However, theoretically, it is unlikely that their moment method has any efficiency property because the selection of their moment equations has not incorporated any efficiency consideration and their suggested MOM does not incorporate proper weighting across their moment equations. In our GMM framework, our best GMM estimators of λ_0 and σ_0^2 can be asymptotically efficient as the QMLE for any distribution (satisfying relevant regularity moment conditions).

3. GMM Estimation of the Regression Model with SAR Disturbances

The regression model with SAR disturbances is specified as

$$Y_n = X_n\beta + u_n, \qquad u_n = \lambda W_n u_n + \epsilon_n, \tag{3.1}$$

where ϵ_n has zero mean and variance $\sigma_0^2 I_n$, and W_n is a spatial weights matrix. The exogenous variables are assumed to satisfy the conventional property:

Assumption 7: *The elements of X_n are uniformly bounded, and $\lim_{n\to\infty}\frac{X_n'X_n}{n}$ exists and is nonsingular.*

This model implies that

$$S_n(\lambda)Y_n = S_n(\lambda)X_n\beta + \epsilon_n, \tag{3.2}$$

which is in the Durbin spatial lag form (Anselin 1988).

The regression model is a generalized linear model with variance $S_n^{-1}S_n^{'-1}$ for the disturbance vector u_n. Let $u_n(\beta) = Y_n - X_n\beta$. A possible estimator of β is the generalized least squares estimator (GLSE) with a consistently estimated variance matrix. In order to estimate the variance matrix $S_n^{-1}S_n^{'-1}$, one needs to estimate the unknown parameter λ in the SAR disturbance process.

Let $\hat{\beta}_{L,n} = (X_n'X_n)^{-1}X_n'Y_n$ be the ordinary least square estimator (OLSE). The disturbance u_n can then be estimated by the estimated residual $u_n^* = Y_n - X_n\hat{\beta}_{L,n}$. The estimated residual is related to ϵ_n as $u_n^* = Q_n\epsilon_n$ where $Q_n = (I_n - X_n(X_n'X_n)^{-1}X_n')S_n^{-1}$. We suggest the estimation of λ_0 by the GMM method: $\min_{\lambda\in\Lambda} g_n'(\lambda)a_n'a_n g_n(\lambda)$ with

$$g_n(\lambda) = (u_n^{*'}S_n'(\lambda)P_{1n}S_n(\lambda)u_n^*, \cdots, u_n^{*'}S_n(\lambda)P_{mn}S_n(\lambda)u_n^*)'. \tag{3.3}$$

The following proposition shows that the GMM estimator $\hat{\lambda}_n$ is $\sqrt{\frac{n}{h_n}}$-consistent and it has the limiting distribution of the corresponding GMM estimator of the SAR process for u_n as if u_n is observable.

Proposition 3.1 *Suppose P_{jn}, $j = 1, \cdots, m$, are selected from $\mathcal{P}_{1n}$ so that $a_0\lim_{n\to\infty}\frac{h_n}{n}E(g_n(\lambda)) = 0$ has a unique root at λ_0 in Λ, where a_n converges to a_0 and $g_n(\lambda)$ is in (3.3). Then, the GMM estimator $\hat{\lambda}_n$ derived from $\min_{\lambda\in\Lambda} g_n'(\lambda)a_n'a_n g_n(\lambda)$ is a consistent estimator of λ_0, and $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \xrightarrow{D} N(0,\Sigma)$, where*

$$\Sigma = \lim_{n\to\infty}\left[(\frac{h_n}{n}d_n)'a_0'a_0(\frac{h_n}{n}d_n)\right]^{-1}(\frac{h_n}{n}d_n)'a_0'a_0(\frac{h_n}{n}\Omega_n)a_0'a_0(\frac{h_n}{n}d_n)\left[(\frac{h_n}{n}d_n)'a_0'a_0(\frac{h_n}{n}d_n)\right]^{-1} \tag{3.4}$$

with $d_n = (tr(P^s_{1n}G_n), \cdots, tr(P^s_{mn}G_n))'$*, under the assumption that* $a_0 \lim_{n\to\infty} \frac{h_n}{n} d_n \neq 0$*.*

With any consistent estimator $\hat{\lambda}_n$ of λ_0, the following proposition shows that the feasible GLSE is asymptotically equivalent to the exact GLSE.

Proposition 3.2 *Let* $\hat{\lambda}_n$ *be a consistent estimator of* λ_0 *and* $\hat{S}_n = I_n - \hat{\lambda}_n W_n$*. The feasible GLSE* $\hat{\beta}_{G,n}$*, where* $\hat{\beta}_{G,n} = (X'_n \hat{S}'_n \hat{S}_n X_n)^{-1} X'_n \hat{S}'_n \hat{S}_n Y_n$*, has the asymptotic distribution that*

$$\sqrt{n}(\hat{\beta}_{G,n} - \beta_0) \xrightarrow{D} N(0, \sigma_0^2 (\lim_{n\to\infty} \frac{1}{n} X'_n S'_n S_n X_n)^{-1}), \tag{3.5}$$

assuming that the limit of $\frac{1}{n} X'_n S'_n S_n X_n$ *exists and is a nonsingular matrix.*

As usual, the asymptotic distribution of the GLSE does not require any specific rate on the consistent estimate $\hat{\lambda}_n$ in the estimation of the weighting matrix. So even though $\hat{\lambda}_n$ may converge at a rate lower than the $\sqrt{n}$-rate, the rate of convergence of $\hat{\beta}_{G,n}$ and its limiting distribution will not be affected.

The following proposition summarizes the main results of the best GMM estimates of λ_0 for the SAR disturbance process. It will be useful if λ_0 in addition to the regression coefficients β is also the interest of the model. This result shows that the same asymptotic efficient properties of estimates for the SAR process hold for the SAR disturbance process in the GMM framework when the unobservable disturbance u_n is replaced by its least squares estimated residuals.

Proposition 3.3 *Suppose that* $\hat{\lambda}_n$ *is a* $\sqrt{\frac{n}{h_n}}$*-consistent estimate of* λ_0 *and* $\hat{G}_n = W_n S_n^{-1}(\hat{\lambda}_n)$*.*

Within the class of optimal GMM estimators $\mathcal{M}_{2n}$*, the best estimator is the consistent root* $\hat{\lambda}_{2b,n}$ *derived from* $\min_{\lambda\in\Lambda}[u_n^{*'} S'_n(\lambda)(\hat{G}_n - Diag(\hat{G}_n))S_n(\lambda)u_n^*]^2$ *in the sense that* $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{2b,n} - \lambda_0) \xrightarrow{D} N(0, \Sigma_{2b})$ *with* $\Sigma_{2b} \leq \Sigma$*, where* Σ *is the limiting variance matrix of* $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0)$ *in Proposition (3.1) and*

$$\Sigma_{2b} = (\lim_{n\to\infty} \frac{h_n}{n} tr[(G_n - Diag(G_n))^s G_n])^{-1}. \tag{3.6}$$

The consistent root $\hat{\lambda}_{2b,n}$ *is*

$$\begin{aligned} \hat{\lambda}_{2b,n} = \{&u_n^{*'}(\hat{G}_n - Diag(\hat{G}_n))^s W_n u_n^* - [(u_n^{*'}(\hat{G}_n - Diag(\hat{G}_n))^s W_n u_n^*)^2 \\ &- 4u_n^{*'} W'_n(\hat{G}_n - Diag(\hat{G}_n))W_n u_n^* \cdot u_n^{*'}(\hat{G}_n - Diag(\hat{G}_n))u_n^*]^{\frac{1}{2}}\}/(2u_n^{*'} W'_n(\hat{G}_n - Diag(\hat{G}_n))W_n u_n^*). \end{aligned} \tag{3.7}$$

In the event that $\epsilon_n \sim N(0, \sigma_0^2 I_n)$ *or for the case that* $\lim_{n\to\infty} h_n = \infty$*, within the broader class of estimators* $\mathcal{M}_{1n}$*, the consistent root* $\hat{\lambda}_{1b,n}$ *derived from* $\min_{\lambda\in\Lambda}[u_n^{*'} S'_n(\lambda)(\hat{G}_n - \frac{tr(\hat{G}_n)}{n} I_n)S_n(\lambda)u_n^*]^2$ *is the best GMM estimator with* $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{1b,n} - \lambda_0) \xrightarrow{D} N(0, \Sigma_{1b})$*, where*

$$\Sigma_{1b} = (\lim_{n\to\infty} \frac{h_n}{n} tr[(G_n - \frac{tr(G_n)}{n} I_n)^s G_n])^{-1}. \tag{3.8}$$

The consistent root $\hat{\lambda}_{1b,n}$ *is*

$$\hat{\lambda}_{1b,n} = \{u_n^{*'}(\hat{G}_n - \frac{tr(\hat{G}_n)}{n}I_n)^s W_n u_n^* - [(u_n^{*'}(\hat{G}_n - \frac{tr(\hat{G}_n)}{n}I_n)^s W_n u_n^*)^2 - 4u_n^{*'}W_n'(\hat{G}_n - \frac{tr(\hat{G}_n)}{n}I_n)W_n u_n^* \cdot u_n^{*'}(\hat{G}_n - \frac{tr(\hat{G}_n)}{n}I_n)u_n^*]^{\frac{1}{2}}\}/(2u_n^{*'}W_n'(\hat{G}_n - \frac{tr(\hat{G}_n)}{n}I_n)W_n u_n^*). \tag{3.9}$$

When ϵ_n is $N(0, \sigma_0^2 I_n)$, the asymptotic variance matrix of the MLEs of β, σ^2, and λ is known to be block diagonal:

$$AsyVar(\beta, \sigma^2, \lambda) = \begin{pmatrix} \frac{1}{\sigma_0^2}(S_n X_n)'(S_n X_n) & 0 & 0 \\ 0 & \frac{n}{2\sigma_0^4} & \frac{tr(G_n)}{\sigma_0^2} \\ 0 & \frac{tr(G_n)}{\sigma_0^2} & tr(G_n^2) + tr(G_n' G_n) \end{pmatrix}^{-1} \tag{3.10}$$

(see, e.g., p.258 of Anselin and Bera (1998)). Comparing the asymptotic variances and covariances of the feasible GLSE in (3.5) and the best $\hat{\lambda}_{1b,n}$ in (3.8) with those of the MLEs of β and λ in (3.10), their asymptotic distributions are the same.[10]

4. GMM Estimation of High Order SAR Processes

Without loss of generality, consider the estimation of a SAR process with p spatial lags:

$$Y_n = (\lambda_1 W_{1n} + \lambda_2 W_{2n} + \cdots + \lambda_p W_{pn})Y_n + \epsilon_n, \tag{4.1}$$

where ϵ_{nj}s are i.i.d.$(0, \sigma_0^2)$ and W_{ln}, $l = 1, \cdots, p$ are p different spatial weights matrices.[11] For this model, denote $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_p)'$, $S_n(\lambda) = I_n - \sum_{j=1}^p \lambda_j W_{jn}$, $S_n = S_n(\lambda_0)$, and $G_{jn} = W_{jn} S_n^{-1}$ for $j = 1, \cdots, p$. Assumption 2 will be strengthened to Assumption 2′ below to incorporate all the spatial weights matrices in this model. Assumption 3 is assumed for the newly defined S_n matrix of this model. The parameter space Λ in Assumption 6 will refer to the parameter vector λ and is a compact subset of the p-dimensional Euclidian space.

Assumption 2′: *The weights matrices* $\{W_{ln}\}$*,* $l = 1, \cdots, p$*, are uniformly bounded in both row and column sums. The elements of* W_{ln} *are of order* $O(\frac{1}{h_n})$ *uniformly in* i *and* j *for each* $l = 1, \cdots, p$*.*

For GMM estimation of the model, we suggest the IVs functions $P_{jn} S_n(\lambda) Y_n$, $j = 1, \cdots, m$, where P_{jn}s are constant matrices from either $\mathcal{P}_{1n}$ or $\mathcal{P}_{2n}$. The empirical moment functions are

$$g_n(\lambda) = \begin{pmatrix} (P_{1n} S_n(\lambda) Y_n)' \\ \vdots \\ (P_{mn} S_n(\lambda) Y_n)' \end{pmatrix} S_n(\lambda) Y_n = \begin{pmatrix} Y_n' S_n'(\lambda) P_{1n} S_n(\lambda) Y_n \\ \vdots \\ Y_n' S_n'(\lambda) P_{mn} S_n(\lambda) Y_n \end{pmatrix}. \tag{4.2}$$

---

[10] Note that the explicit expression of the asymptotic variance of the MLE of λ_0 from (3.10) is the inverse of $tr(G_n^s G_n) - \frac{tr(G_n)}{\sigma_0^2}(\frac{n}{2\sigma_0^4})^{-1}\frac{tr(G_n)}{\sigma_0^2} = tr(G_n^s G_n) - 2\frac{tr^2(G_n)}{n}$ by the inverse formula for a partitioned matrix. On the other hand, $tr((G_n - \frac{tr(G_n)}{n}I_n)^s G_n) = tr((G_n^s - 2\frac{tr(G_n)}{n}I_n)G_n) = tr(G_n^s G_n) - 2\frac{tr^2(G_n)}{n}$.

[11] If some of the spatial weights matrices are identical, there is a trivial underidentification problem (Anselin 1988). Some justifications on the specification of this model can be found in Anselin (1988).

At λ_0, $g_n(\lambda_0) = (\epsilon_n' P_{1n}\epsilon_n, \cdots, \epsilon_n' P_{mn}\epsilon_n)'$ and, hence, $E(g_n(\lambda_0)) = 0$. It follows that

$$E(g_n(\lambda)) = \begin{pmatrix} \sigma_0^2 tr(S_n^{'-1} S_n'(\lambda) P_{1n} S_n(\lambda) S_n^{-1}) \\ \vdots \\ \sigma_0^2 tr(S_n^{'-1} S_n'(\lambda) P_{mn} S_n(\lambda) S_n^{-1}) \end{pmatrix}. \quad (4.3)$$

Each of the moment equations is a second degree algebraic equation in p variables λ_k, $k = 1, \cdots, p$ and its solution set is complex and not illuminating. Identification conditions, however, can be derived by investigating some characteristics of the moment equations $E(g_n(\lambda)) = 0$ of (4.3). As $S_n(\lambda) = S_n + \sum_{k=1}^p (\lambda_{k0} - \lambda_k) W_{kn}$, $S_n(\lambda) S_n^{-1} = I_n + \sum_{k=1}^p (\lambda_{k0} - \lambda_k) G_{kn}$. Let $q_{n,k}(j) = tr(P_{jn}^s G_{kn})$ and $q_{n,kl}(j) = tr(G_{kn}' P_{jn} G_{ln})$ for $k, l = 1, \cdots, p$, and $j = 1, \cdots, m$. It follows that

$$tr(S_n^{'-1} S_n'(\lambda) P_{jn} S_n(\lambda) S_n^{-1}) = \sum_{k=1}^p q_{n,k}(j)(\lambda_{k0} - \lambda_k) + \sum_{k=1}^p \sum_{l=1}^p q_{n,kl}(j)(\lambda_{k0} - \lambda_k)(\lambda_{l0} - \lambda_l),$$

for $j = 1, \cdots, m$. It is apparent that λ_0 is a common solution of these m moment equations. Let $q_{n,k}$ be the m-dimensional vector with $q_{n,k}(j)$ being its jth element. Similarly, $q_{n,kl}$, etc., are defined. Identification conditions for λ_0 can be stated in terms of those q_n vectors. The necessary and sufficient condition for the m-moment equations to have a unique solution vector at λ_0 is that the vectors q_ns do not have a linear combination with some nonzero nonlinear coefficients in the form that

$$\sum_{k=1}^p q_{n,k}\delta_k + \sum_{k=1}^p \sum_{l=1}^p q_{n,kl}\delta_k\delta_l = 0. \quad (4.4)$$

This condition is a necessary and sufficient condition for identification of the p-order SAR process in the GMM framework. A sufficient condition is that the q_ns are linearly independent. As the q_ns together will form a matrix of dimension $m \times [p(p+1)]$, in order to have the sufficient identification condition satisfied, the number of P_ns has to be at least as $p(p+1)$. Weaker sufficient conditions are available. If there were a solution of λ_1 not equal to λ_{10}, the moment equation (4.4) would have $\delta_1 \neq 0$. This would imply that each of $q_{n,1}$ and $q_{n,11}$ would be linearly dependent on all the other $[p(p+1) - 1]$ vectors. So it is sufficient to identify λ_{10} if either $q_{n,1}$ or $q_{n,11}$ are linearly independent of the other $[p(p+1) - 1]$ vectors. Once the identification of λ_{10} is achieved, the moment equations in (4.4) will be reduced to the moment equations for a $(p-1)$-order SAR process. A set of weaker identification conditions can thus be recursively derived.

The variance matrix of $g_n(\lambda_0)$ is $\text{var}(g_n(\lambda_0)) = \sigma_0^4 \Omega_n$, where Ω_n has the general expression in (2.15) if P_{jn}'s are from $\mathcal{P}_{1n}$, and equals V_n in (2.14) if P_{jn}'s are from $\mathcal{P}_{2n}$. The derivatives of $g_n(\lambda)$ with respect to λ form the matrix

$$\frac{\partial g_n(\lambda)}{\partial \lambda'} = -\begin{pmatrix} (W_{1n}Y_n)' P_{1n}^s S_n(\lambda) Y_n, & \cdots, & (W_{pn}Y_n)' P_{1n}^s S_n(\lambda) Y_n \\ \vdots & & \vdots \\ (W_{1n}Y_n)' P_{mn}^s S_n(\lambda) Y_n, & \cdots, & (W_{pn}Y_n)' P_{mn}^s S_n(\lambda) Y_n \end{pmatrix}.$$

It follows that $\frac{\partial E(g_n(\lambda_0))}{\partial \lambda'} = -\sigma_0^2 D_n$ where

$$D_n = \begin{pmatrix} tr(P_{1n}^s G_{1n}), & \cdots, & tr(P_{1n}^s G_{pn}) \\ \vdots & & \vdots \\ tr(P_{mn}^s G_{1n}), & \cdots, & tr(P_{mn}^s G_{pn}) \end{pmatrix} \tag{4.5}$$

is a $m \times p$ matrix. The following proposition summarizes the asymptotic distribution of the GMM estimator for this model, the optimal GMM estimator with a given vector of moment functions, and the best GMM estimates from $\mathcal{P}_{1n}$ or $\mathcal{P}_{2n}$.

Proposition 4.1 *Suppose P_{jn}, $j = 1, \cdots, m$, are selected from $\mathcal{P}_{1n}$ so that $a_0 \lim_{n\to\infty} \frac{h_n}{n} E(g_n(\lambda)) = 0$ has a unique root at λ_0 in Λ, where a_n converges to a_0 and $g_n(\lambda)$ is in (4.2). Then, the GMM estimator $\hat{\lambda}_n$ derived from $\min_{\lambda\in\Lambda} g_n'(\lambda)a_n' a_n g_n(\lambda)$ is a consistent estimator of λ_0, and $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \xrightarrow{D} N(0, \Sigma)$, where*

$$\Sigma = \lim_{n\to\infty} \left[(\frac{h_n}{n} D_n)' a_0' a_0 (\frac{h_n}{n} D_n) \right]^{-1} (\frac{h_n}{n} D_n)' a_0' a_0 (\frac{h_n}{n} \Omega_n) a_0' a_0 (\frac{h_n}{n} D_n) \left[(\frac{h_n}{n} D_n)' a_0' a_0 (\frac{h_n}{n} D_n) \right]^{-1}, \tag{4.6}$$

with D_n in (4.5) under the assumption that $a_0 \lim_{n\to\infty} \frac{h_n}{n} D_n$ has a full column rank p.

The optimal choice of a_n corresponds to $a_n^* = (\frac{h_n}{n}\Omega_n)^{-\frac{1}{2}}$. With the optimal a_n^*, the optimal GMM estimator $\hat{\lambda}_n^*$ with moments $g_n(\lambda)$ in (4.2) derived from $\min_{\lambda\in\Lambda} g_n'(\lambda)\Omega_n^{-1} g_n(\lambda)$ has $\sqrt{\frac{h_n}{n}}(\hat{\lambda}_n^* - \lambda_0) \xrightarrow{D} N(0, \Sigma^*)$, where $\Sigma^* = \text{plim}_{n\to\infty}(\frac{h_n}{n} D_n' \Omega_n^{-1} D_n)^{-1}$.

Furthermore, the best selections of P_ns from $\mathcal{P}_{2n}$ are $(G_{jn} - Diag(G_{jn}))$ for $j = 1, \cdots, p$. When ϵ_n is normally distributed or $\lim_{n\to\infty} h_n = \infty$, the best selection of P_n's from $\mathcal{P}_{1n}$ are $(G_{jn} - \frac{tr(G_{jn})}{n} I_n)$, $j = 1, \cdots, p$.

The following proposition demonstrates that the feasible best estimators can be constructed with a $\sqrt{\frac{n}{h_n}}$-consistent estimate $\hat{\lambda}_n$ of λ_0. It generalizes Proposition 2.10 to the high order SAR process.

Proposition 4.2 *Suppose that $\hat{\lambda}_n$ is a $\sqrt{\frac{n}{h_n}}$-consistent estimate of λ_0. Let $\hat{G}_{jn} = W_{jn} S_n^{-1}(\hat{\lambda}_n)$, $j = 1, \cdots, p$.*

Within the class of optimal GMM estimators $\mathcal{M}_{2n}$, the best estimator is the consistent root $\hat{\lambda}_{2b,n}$ derived from $\min_{\lambda\in\Lambda} g_{2n}^{*'}(\lambda) V_{2n}^{*-1} g_{2n}^*(\lambda)$ where V_{2n}^* is a $p \times p$ matrix with its (j,l)th entry being $tr[(\hat{G}_{jn} - Diag(\hat{G}_{jn}))(\hat{G}_{ln} - Diag(\hat{G}_{ln}))^s]$, and $g_{2n}^*(\lambda)$ is a p-dimensional vector with its jth entry being $Y_n' S_n'(\lambda)(\hat{G}_{jn} - Diag(\hat{G}_{jn}))S_n(\lambda)Y_n$ in the sense that $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{2b,n} - \lambda_0) \xrightarrow{D} N(0, \Sigma_{2b})$ with $\Sigma_{2b} \le \Sigma$, where Σ is the limiting variance matrix of $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0)$ in Proposition (4.1) and $\Sigma_{2b} = (\lim_{n\to\infty} \frac{h_n}{n} V_{2n}^*)^{-1}$.

In the event that $\epsilon_n \sim N(0, \sigma_0^2 I_n)$ or for the case that $\lim_{n\to\infty} h_n = \infty$, within the broader class of estimators $\mathcal{M}_{1n}$, the best estimator is the consistent root $\hat{\lambda}_{1b,n}$ derived from $\min_{\lambda\in\Lambda} g_{1n}^{*'}(\lambda) V_{1n}^{*-1} g_{1n}^*(\lambda)$ where

V^*_{1n} is a $p \times p$ matrix with its (j,l)th entry being $tr[(\hat{G}_{jn} - \frac{tr(\hat{G}_{jn})}{n} I_n)(\hat{G}_{ln} - \frac{tr(\hat{G}_{ln})}{n} I_n)^s]$, and $g^*_{1n}(\lambda)$ is a p-dimensional vector with its jth entry $Y'_n S'_n(\lambda)(\hat{G}_{jn} - \frac{tr(\hat{G}_{jn})}{n} I_n) S_n(\lambda) Y_n$, and it has the limiting distribution $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{1b,n} - \lambda_0) \xrightarrow{D} N(0, \Sigma_{1b})$ where $\Sigma_{1b} = (\lim_{n\to\infty} \frac{h_n}{n} V^*_{1n})^{-1}$.

For the high order SAR process, the moment functions for the GMM estimation are no longer quadratic functions of a single parameter. It becomes algebraically intractable to pinpoint the explicit consistent solution of a moment equation. A possible strategy to implement the feasible best GMM in Proposition 4.2 is to start the nonlinear optimization search with the initial consistent estimate $\hat{\lambda}_n$ as the starting point.[12] The initial consistent estimate can be made available as in Proposition 4.1 when enough distinct matrices P_ns are used in the GMM estimation. The constant matrices to initialize the GMM estimation can be matrices constructed from the spatial weights matrices W_{jn}s of the model, e.g., W_{jn}, $(W'_{jn} W_{ln} - \frac{tr(W'_{jn} W_{ln})}{n} I_n)$, $(W_{jn} W_{ln} - \frac{tr(W_{jn} W_{ln})}{n} I_n)$, etc., for $j, l = 1, \cdots, p$. Alternatively, the best moment functions in Proposition 4.2 can be supplemented with these inefficient moment functions to formulate an extended optimal GMM estimation. With extra moment functions, it is possible to identify uniquely the true parameters as discussed before. The additional moment functions will not increase the asymptotic efficient of the best GMM estimator but it helps to isolate the consistent root.[13]

As the moment functions are second order polynomials of several parameters, the computation of the GMM estimator can be much simpler than the corresponding QML method for the model.[14] The GMM optimization is numerically equivalent to a NLS estimation for a regression equation with nonlinear in parameters. For the p-order SAR process, the regression coefficients are linear functions of λ_j and products $\lambda_j \lambda_k$ for $j, k = 1, \cdots p$. For the feasible best GMM estimators, the additional computation is on the inverse matrix $S_n^{-1}(\hat{\lambda}_n)$ at an initiate consistent estimate $\hat{\lambda}_n$.

5. Conclusions

In this paper, we have suggested GMM for the estimation of SAR processes. The GMM can be computationally simpler than the computation of the QMLE. We consider asymptotic properties of the GMM

[12] What we have in mind are numerical algorithms such as the Newton method where the update estimate from an initial consistent estimate in each iteration is also consistent.

[13] For this extended GMM objective function, optimization search can in principle start at any arbitrary initial point.

[14] The existing literature does not have sufficient discussions on the implementation of the ML method for a high-order SAR model. The implementation of the likelihood function can be demanding as it involves the evaluation of the Jacobian of $S_n(\lambda)$ at any possible value of λ. Ord's device (Ord 1975) of evaluating the determinant of $(I_n - \rho W_{n,n})$ based on the eigenvector decomposition of $W_{n,n}$ will not be generalizable for handling the Jacobian of the likelihood function of a higher order SAR model.

estimators. We discuss the construction of optimal GMM estimators with given moment equations as well as the best selection of moment equations in some broad classes of moment equations. The best GMM estimator is shown to have the same limiting distribution of the QMLE (under normal distributional specification). The GMM can be extended to the estimation of high-order SAR processes. As contrary to the QML method, the computational complexity of the GMM estimator does not increase as more spatial lags are introduced. It can also be applied to the estimation of regression models with SAR disturbances.

In this paper, we focus solely on the estimation of SAR processes but not SAR models with mixed spatial lags and exogenous variables. For the latter models, GMM estimation methods have been proposed and discussed in various manuscripts and articles including Anselin (1988, 1990), Land and Deane (1992), Kelejian and Robinson (1993), Kelejian and Prucha (1997, 1998), Lee (1999a), among others. However, the proposed GMMs are either linear IV, 2SLS, or generalized 2SLS methods. The validity of those methods relies exclusively on the presence of exogenous variables in the model to construct their IVs. Those methods can not be applied to (pure) SAR processes as there are no relevant exogenous variables in the processes. Even though our proposed GMM framework in this paper is specifically designed for the estimation of (pure) SAR processes, it may be extended for the estimation of mixed regressive SAR models by incorporating exogenous variables in the GMM framework. Results on that direction of research shall be reported in a separate paper.

Appendix A: Some Useful Lemmas

This appendix summarizes results which are useful for the subsequent proofs of our propositions in the text. Frequent notations used in the text are assumed to be understood and will be used in the following Lemmas without interpretation. For example, W_n refers to a $n \times n$ spatial weights matrix, S_n refers to $(I_n - \lambda_0 W_n)$ or $(I_n - \sum_{j=1}^{p} \lambda_j W_{jn})$, $G_n = W_n S_n^{-1}$, etc. Elements ϵ_{ni} of the n disturbance vector ϵ_n are always assumed to be i.i.d. with zero mean, variance σ^2 and finite fourth moments μ_4 in the Lemmas.

For any $n \times n$ matrix A_n which is uniformly bounded in both row and column sums, a linear transformation of A_n which preserves the uniform boundedness property will be denoted by A_n^L. The particular transformations of A_n to $(A_n - \frac{tr(A_n)}{n} I_n)$ and $(A_n - Diag(A_n))$ are linear, and will be denoted as A_n^d to simplify presentation.

Lemma A.1 *Suppose that the elements $a_{n,ij}$ of the sequence of $n \times n$ matrices $\{A_n\}$, where $A_n = [a_{n,ij}]$, have the order $O(\frac{1}{h_n})$ uniformly in all i and j; and $\{B_n\}$ is a sequence of conformable $n \times n$ matrices.*

(1) If $\{B_n\}$ are uniformly bounded in column sums, then the elements of $A_n B_n$ have the uniform order $O(\frac{1}{h_n})$.

(2) If $\{B_n\}$ are uniformly bounded in row sums, then the elements of $B_n A_n$ have the uniform order $O(\frac{1}{h_n})$.

For both cases (1) and (2), $|tr(A_n B_n)| = |tr(B_n A_n)| = O(\frac{n}{h_n})$.

Proof: Consider (1). Let $a_{n,ij} = \frac{c_{n,ij}}{h_n}$. Because $a_{n,ij} = O(\frac{1}{h_n})$ uniformly in i and j, there exists a constant c_1 so that $|c_{n,ij}| \leq c_1$ for all i, j and n. Because $\{B_n\}$ is uniformly bounded in column sums, there exists a constant c_2 so that $\sum_{k=1}^{n} |b_{n,kj}| \leq c_2$ for all n and j. Let $a_{i,n}$ be the ith row of A_n and $b_{n,l}$ be the lth column of B_n. It follows that

$$|a_{i,n} b_{n,l}| \leq \frac{1}{h_n} \sum_{j=1}^{n} |c_{n,ij} b_{n,jl}| \leq \frac{c_1}{h_n} \sum_{j=1}^{n} |b_{n,jl}| \leq \frac{c_1 c_2}{h_n},$$

for all i and l. Furthermore, $|tr(A_n B_n)| = |\sum_{i=1}^{n} a_{i,n} b_{n,i}| \leq \sum_{i=1}^{n} |a_{i,n} b_{n,i}| \leq c_1 c_2 \frac{n}{h_n}$. These prove the results in (1). The results in (2) follow from (1) because $(B_n A_n)' = A_n' B_n'$ and the uniform boundedness in row sums of $\{B_n\}$ is equivalent to the uniform boundedness in column sums of $\{B_n'\}$. Q.E.D.

Lemma A.2 *Let $A_n = [a_{ij}]$ be an n-dimensional square matrix. Then*

1) $E(\epsilon_n' A_n \epsilon_n) = \sigma^2 tr(A_n)$,

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

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

Proof: See Lee (1999b). Q.E.D.

Lemma A.3 *Suppose that* $\{A_n\}$ *are uniformly bounded in both row and column sums, and the elements of* $A_n = [a_{n,ij}]$ *have the order* $a_{n,ij} = O(\frac{1}{h_n})$ *uniformly in all* i *and* j*. Then,* $E(\epsilon_n' A_n \epsilon_n) = O(\frac{n}{h_n})$*,* $\text{var}(\epsilon_n' A_n \epsilon_n) = O(\frac{n}{h_n})$*, and* $\epsilon_n' A_n \epsilon_n = O_P(\frac{n}{h_n})$*. Furthermore, if* $\lim_{n\to\infty} \frac{h_n}{n} = 0$*, then* $\frac{h_n}{n}\epsilon_n' A_n \epsilon_n - \frac{h_n}{n}E(\epsilon_n' A_n \epsilon_n) = o_P(1)$*.*

Proof: $E(\epsilon_n' A_n \epsilon_n) = \sigma^2 tr(A_n) = O(\frac{n}{h_n})$ by Lemma A.1. From Lemma A.2, the variance of $\epsilon_n' A_n \epsilon_n$ is $\text{var}(\epsilon_n' A_n \epsilon_n) = (\mu_4 - 3\sigma^4)\sum_{i=1}^n a_{n,ii}^2 + \sigma^4[tr(A_n A_n') + tr(A_n^2)]$. Lemma A.1 implies that $tr(A_n^2)$ and $tr(A_n A_n')$ are of order $O(\frac{n}{h_n})$. As $\sum_{i=1}^n a_{n,ii}^2 \leq tr(A_n A_n')$, it follows that $\sum_{i=1}^n a_{n,ii}^2 = O(\frac{n}{h_n})$. Hence, $\text{var}(\epsilon_n' A_n \epsilon_n) = O(\frac{n}{h_n})$.

When $\frac{h_n}{n} = o(1)$, $E((\epsilon_n' A_n \epsilon_n)^2) = \text{var}(\epsilon_n' A_n \epsilon_n) + E^2(\epsilon_n' A_n \epsilon_n) = O(\max[\frac{n}{h_n}, (\frac{n}{h_n})^2]) = O((\frac{n}{h_n})^2)$. The generalized Chebyshev inequality implies that $P(\frac{h_n}{n}|\epsilon_n' A_n \epsilon_n| \geq M) \leq \frac{1}{M^2}(\frac{h_n}{n})^2 E(|\epsilon_n' A_n \epsilon_n|^2) = \frac{1}{M^2}O(1)$ and, hence, $\frac{h_n}{n}\epsilon_n' A_n \epsilon_n = O_P(1)$. Finally, because $\text{var}(\frac{h_n}{n}\epsilon_n' A_n \epsilon_n) = O(\frac{h_n}{n}) = o(1)$, the Chebyshev inequality implies that $\frac{h_n}{n}\epsilon_n' A_n \epsilon_n - \frac{h_n}{n}E(\epsilon_n' A_n \epsilon_n) = o_P(1)$. Q.E.D.

Lemma A.4 *Suppose that* $\{A_n\}$ *is a sequence of symmetric matrices with row and column sums uniformly bounded in absolute value and the entries* $a_{n,ij}$ *of* A_n *are of order* $O(\frac{1}{h_n})$*. The* $\epsilon_{n1}, \cdots, \epsilon_{nn}$ *are i.i.d. random variables with zero mean and finite variance* σ^2*, and its moment* $E(|\epsilon|^{4+2\delta})$ *for some* $\delta > 0$ *exists. Let* $\sigma_{Q_n}^2$ *be the variance of* Q_n *where* $Q_n = V_n' A_n V_n - \sigma^2 tr(A_n)$*. Assume that the variance* $\sigma_{Q_n}^2$ *is bounded away from zero at the rate* $\frac{n}{h_n}$*. If* $\lim_{n\to\infty} \frac{h_n^{1+\frac{2}{\delta}}}{n} = 0$*, then* $\frac{Q_n}{\sigma_{Q_n}} \xrightarrow{D} N(0,1)$*.*

Proof: See Kelejian and Prucha (1999b) and Lee (1999b).

Lemma A.5 *Suppose that the elements of the sequences of* n*-dimensional column vectors* Z_{1n} *and* Z_{2n} *are uniformly bounded. If* $\{A_n\}$ *are uniformly bounded in either row or column sums, then* $|Z_{1n}' A_n Z_{2n}| = O(n)$*.*

Proof: Trivial.

Lemma A.6 *Suppose that* A_n *is a* $n \times n$ *matrix with its column sums being uniformly bounded and elements of the* $n \times k$ *matrix* C_n *are uniformly bounded. Then,* $\frac{1}{\sqrt{n}}C_n' A_n \epsilon_n = O_P(1)$*, Furthermore, if the limit of* $\frac{1}{n}C_n' A_n A_n' C_n$ *exists and is positive definite, then* $\frac{1}{\sqrt{n}}C_n' A_n \epsilon_n \xrightarrow{D} N(0, \sigma_0^2 \lim_{n\to\infty} \frac{1}{n}C_n' A_n A_n' C_n)$*.*

Proof: See Lee (1999b). Q.E.D.

Lemma A.7 *Consider* $S_n(\lambda) = I_n - \sum_{j=1}^p \lambda_j W_{jn}$*. Suppose that* $\{\| S_n^{-1} \|\}$ *and* $\{\| W_{jn} \|\}$ *for* $j = 1, \cdots, p$*, where* $\| \cdot \|$ *is a matrix norm, are bounded. Then,* $\{\| S_n(\lambda)^{-1} \|\}$ *are uniformly bounded in a* *neighborhood of* λ_0*.*

Proof:[15] Let c be a constant so that $\| S_n^{-1} \| \leq c$ and $\| W_{jn} \| \leq c$ for $j = 1, \cdots, p$, and all n. We note that $S_n^{-1}(\lambda) = (S_n - \sum_{j=1}^p (\lambda_j - \lambda_{j0}) W_{jn})^{-1} = S_n^{-1}(I_n - R_n(\lambda))^{-1}$ where $R_n(\lambda) = \sum_{j=1}^p (\lambda_j - \lambda_{j0}) G_{jn}$. By the submultiplicative property of a matrix norm, $\| G_{jn} \| \leq \| W_{jn} \| \cdot \| S_n^{-1} \| \leq c^2$ for all j and n.

Let $B_1(\lambda_0) = \{\lambda : \sum_{j=1}^p |\lambda_j - \lambda_0| < \frac{1}{c^2}\}$. It follows that, for any $\lambda \in B_1(\lambda_0)$, $\| R_n(\lambda) \| \leq \sum_{j=1}^p |\lambda_j - \lambda_{j0}| \cdot \| G_{jn} \| < 1$. Hence, $(I_n - R_n(\lambda))$ for $\lambda \in B_1(\lambda_0)$ is invertible and it has the expansion that $(I_n - R_n(\lambda))^{-1} = \sum_{k=0}^\infty R_n^k(\lambda)$ (Horn and Johnson 1985). We note that

$$\| R_n(\lambda) \|^k \leq (\sum_{j=1}^p |\lambda_j - \lambda_{j0}| \cdot \| G_{jn} \|)^k \leq (\max_{j=1,\dots,p} \| G_{jn} \| \cdot \sum_{j=1}^p |\lambda_j - \lambda_{j0}|)^k$$

$$\leq \max_{j=1,\dots p} \| G_{jn} \|^k \cdot (\sum_{j=1}^p |\lambda_j - \lambda_{j0}|)^k \leq c^{2k} (\sum_{j=1}^p |\lambda_j - \lambda_0|)^k.$$

Hence, $\| I_n - R_n(\lambda) \| \leq \sum_{k=0}^\infty \| R_n(\lambda) \|^k \leq \sum_{k=0}^\infty (c^2 \sum_{j=1}^p |\lambda_j - \lambda_0|)^k = \frac{1}{1 - c^2 \sum_{j=1}^p |\lambda_j - \lambda_0|} < \infty$ for $\lambda \in B_1(\lambda_0)$. The final result follows by taking a close neighborhood $B(\lambda_0)$ of λ_0 contained within $B_1(\lambda_0)$. In $B(\lambda_0)$, $\sup_{\lambda \in B(\lambda_0)} c^2 \sum_{j=1}^p |\lambda_j - \lambda_{j0}| < 1$. Therefore,

$$\sup_{\lambda \in B(\lambda_0)} \| S_n^{-1}(\lambda) \| \leq \| S_n^{-1} \| \cdot \sup_{\lambda \in B(\lambda_0)} \| I_n - R_n(\lambda) \| \leq \sup_{\lambda \in B(\lambda_0)} \frac{c}{1 - c^2 \sum_{j=1}^p |\lambda_j - \lambda_0|} < \infty.$$

Q.E.D.

Lemma A.8 *Suppose that* $\frac{h_n}{n}(g_n(\lambda) - E(g_n(\lambda)) = o_P(1)$ *uniformly in* $\lambda \in \Lambda$*, and* $\frac{h_n}{n} E(g_n(\lambda)) = 0$ *has a unique root at* λ_0 *in* Λ *as* n *goes to infinity. The* $\hat{\lambda}_n$ *and* $\hat{\lambda}_n^*$ *are, respectively, the roots of the moment equations* $g_n(\hat{\lambda}_n) = 0$ *and* $g_n^*(\hat{\lambda}_n^*) = 0$*. If* $\frac{h_n}{n}(g_n^*(\lambda) - g_n(\lambda)) = o_P(1)$ *uniformly in* $\lambda \in \Lambda$*, then both* $\hat{\lambda}_n$ *and* $\hat{\lambda}_n^*$ *converge in probability to* λ_0*.*

In addition, suppose that $\frac{h_n}{n} \frac{\partial g_n(\lambda)}{\partial \lambda}$ *converges in probability to a well defined limit function* $Q(\lambda)$ *uniformly in* $\lambda \in \Lambda$ *with* $Q(\lambda_0) \neq 0$*, and* $\sqrt{\frac{h_n}{n}} g_n(\lambda_0) = O_P(1)$*. If* $\frac{h_n}{n}(\frac{\partial g_n^*(\lambda)}{\partial \lambda} - \frac{\partial g_n(\lambda)}{\partial \lambda}) = o_P(1)$ *uniformly in* $\lambda \in \Lambda$*, and* $\sqrt{\frac{h_n}{n}}(g_n^*(\lambda_0) - g_n(\lambda_0)) = o_P(1)$*, then both* $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0)$ *and* $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n^* - \lambda_0)$ *have the same limiting distribution.*

Proof: The convergence of $\hat{\lambda}_n$ to λ_0 follows from the uniform convergence of $\frac{h_n}{n}(g_n(\lambda) - E(g_n(\lambda)))$ to zero in probability and the identification uniqueness condition at λ_0. As $\frac{h_n}{n}[g_n^*(\lambda) - E(g_n(\lambda))] = \frac{h_n}{n}(g_n^*(\lambda) - g_n(\lambda)) + \frac{h_n}{n}[g_n(\lambda) - E(g_n(\lambda))] = o_P(1)$ uniformly in Λ, the consistency of $\hat{\lambda}_n^*$ follows.

For the limiting distribution, the Taylor expansion of $g_n(\hat{\lambda}_n) = 0$ at λ_0 implies that

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = (-\frac{h_n}{n} \frac{\partial g_n(\bar{\lambda}_n)}{\partial \lambda})^{-1} \sqrt{\frac{h_n}{n}} g_n(\lambda_0) = -Q(\lambda_0)^{-1} \sqrt{\frac{h_n}{n}} g_n(\lambda_0) + o_P(1),$$

[15] This Lemma and its proof generalize those for the first-order SAR model in Lee (1999b).

because $\bar{\lambda}_n$ lying between $\hat{\lambda}_n$ and λ_0 converges in probability to λ_0. For $\hat{\lambda}_n^*$, the Taylor expansion of $g_n^*(\hat{\lambda}_n^*) = 0$ implies that

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n^* - \lambda_0) = -(\frac{h_n}{n}\frac{\partial g_n^*(\bar{\lambda}_n^*)}{\partial \lambda})^{-1}\sqrt{\frac{h_n}{n}}g_n^*(\lambda_0) = -(\frac{h_n}{n}\frac{\partial g_n(\bar{\lambda}_n^*)}{\partial \lambda} + o_P(1))^{-1}(\sqrt{\frac{h_n}{n}}g_n(\lambda_0) + o_P(1))$$

$$= -Q(\lambda_0)^{-1}\sqrt{\frac{h_n}{n}}g_n(\lambda_0) + o_P(1).$$

These show that $\hat{\lambda}_n^*$ has the same limiting distribution as $\hat{\lambda}_n$. Q.E.D.

Lemma A.9 *Let A_n and B_n be $n \times n$ matrices, uniformly bounded in both row and column sums. Let $C_n(\lambda) = W_{ln}S_n^{-1}(\lambda)$ for some l, where $S_n(\lambda) = I_n - \sum_{j=1}^m \lambda_j W_{jn}$. Suppose that $\hat{\lambda}_n$ is a $\sqrt{\frac{n}{h_n}}$-consistent estimator of λ_0 and $\frac{h_n^{1+\delta}}{n} = o(1)$ for some $\delta > 0$. Then,*

(i) $\frac{h_n}{n}\epsilon_n' A_n'(C_n^d(\hat{\lambda}_n) - C_n^d(\lambda_0))B_n\epsilon_n = o_P(1)$, *and*

(ii) $\sqrt{\frac{h_n}{n}}\epsilon_n'(C_n^d(\hat{\lambda}_n) - C_n^d(\lambda_0))\epsilon_n = o_P(1)$.

Proof: For any $n \times n$ matrix M, $M^d = M - Diag(M)$ or $= M - \frac{tr(M)}{n}I_n$ is a transformed $n \times n$ matrix. This transformation is linear because $Diag$ is a linear transformation and tr is a linear function.

As $S_n - S_n(\hat{\lambda}_n) = \sum_{j=1}^p (\hat{\lambda}_{nj} - \lambda_{j0})W_{jn}$, it follows that

$$S_n^{-1}(\hat{\lambda}_n) - S_n^{-1} = S_n^{-1}(\hat{\lambda}_n)[S_n - S_n(\hat{\lambda}_n)]S_n^{-1} = S_n^{-1}(\hat{\lambda}_n)[\sum_{j=1}^p (\hat{\lambda}_{nj} - \lambda_{j0})G_{jn}].$$

By induction,

$$S_n^{-1}(\hat{\lambda}_n) - S_n^{-1} = S_n^{-1}\sum_{k=1}^m[\sum_{j=1}^p (\hat{\lambda}_{nj} - \lambda_{j0})G_{jn}]^k + S_n^{-1}(\hat{\lambda}_n)[\sum_{j=1}^p (\hat{\lambda}_{nj} - \lambda_{j0})G_{jn}]^{m+1}, \qquad (A.1)$$

for any positive integer m.

Let $T_n = \frac{h_n}{n}\epsilon_n' A_n'(C_n^d(\hat{\lambda}_n) - C_n^d(\lambda_0))B_n\epsilon_n$. With the above expansion, $T_n = T_{n1} + T_{n2}$ where

$$T_{n1} = \frac{h_n}{n}\epsilon_n' A_n'(G_{ln}\sum_{k=1}^m[\sum_{j=1}^p (\hat{\lambda}_{nj} - \lambda_{j0})G_{jn}]^k)^d B_n\epsilon_n$$

and $T_{n2} = \frac{h_n}{n}\epsilon_n' A_n'(W_{ln}S_n^{-1}(\hat{\lambda}_n)[\sum_{j=1}^p (\hat{\lambda}_{nj} - \lambda_{j0})G_{jn}]^{m+1})^d B_n\epsilon_n$. The term T_{n1} can be rewritten as

$$T_{n1} = \sum_{k=1}^m \sum_{j_1=1}^p \cdots \sum_{j_k=1}^p (\hat{\lambda}_{j_1} - \lambda_{j_1 0})\cdots(\hat{\lambda}_{j_k 0} - \lambda_{j_k 0})\frac{h_n}{n}\epsilon_n' A_n'(G_{ln}G_{j_1 n}\cdots G_{j_k n})^d B_n\epsilon_n = o_P(1),$$

because $\frac{h_n}{n}\epsilon_n' A_n'(G_{ln}G_{j_1 n}\cdots G_{j_k n})^d B_n\epsilon_n = O_p(1)$ by Lemma A.3, and $\hat{\lambda}_{jn} - \lambda_{j0} = o_P(1)$. For T_{n2}, let $\| \cdot \|$ be either the maximum row sum norm or the maximum column sum norm. One has

$$|T_{n2}| \leq \sum_{j_1=1}^p \cdots \sum_{j_{m+1}=1}^p |\hat{\lambda}_{nj_1} - \lambda_{j_1 0}|\cdots|\hat{\lambda}_{nj_{m+1}} - \lambda_{j_{m+1}0}|\frac{h_n}{n} \| \epsilon_n' \| \cdot \| \epsilon_n \|$$

$$\cdot \| A_n'(W_{ln}S_n^{-1}(\hat{\lambda}_n)G_{j_1}\cdots G_{j_{m+1}})^d B_n \|$$

$$\leq c\sum_{j_1=1}^p \cdots \sum_{j_{m+1}=1}^p |\hat{\lambda}_{nj_1} - \lambda_{j_1 0}|\cdots|\hat{\lambda}_{nj_{m+1}} - \lambda_{j_{m+1}0}|\frac{h_n}{n} \| \epsilon_n' \| \cdot \| \epsilon_n \|$$

for some constant c, where the last inequality holds because the uniform boundedness of S_n^{-1} in row (resp. column) sums implies that $S_n^{-1}(\lambda)$ is uniformly bounded in row (resp. column) sums, uniformly in a small neighborhood of λ_0 by Lemma A.7; and the product of relevant matrices is uniformly bounded in either row or column sums. We note that, for any finite positive k_1 and k_2, $n^{k_1} h_n^{k_2} (\frac{h_n}{n})^m = \frac{h_n^{m+k_2}}{n^{m-k_1}} \leq (\frac{h_n^{1+\delta}}{n})^{m-k_1} = o(1)$ for large enough m. Hence,

$$\sum_{j_1=1}^{p} \cdots \sum_{j_{m+1}=1}^{p} |\hat{\lambda}_{nj_1} - \lambda_{j_1 0}| \cdots |\hat{\lambda}_{nj_{m+1}} - \lambda_{j_{m+1}0}| \frac{h_n}{n} \parallel \epsilon_n' \parallel \cdot \parallel \epsilon_n \parallel$$

$$\leq n h_n (\frac{h_n}{n})^{\frac{m+1}{2}} \sum_{j_1=1}^{p} \cdots \sum_{j_{m+1}=1}^{p} |\sqrt{\frac{n}{h_n}} (\hat{\lambda}_{nj_1} - \lambda_{j_1 0})| \cdots |\sqrt{\frac{n}{h_n}} (\hat{\lambda}_{nj_{m+1}} - \lambda_{j_{m+1}0})| (\frac{1}{n} \sum_{i=1}^{n} |\epsilon_{ni}|)^2 = o_P(1),$$

because $\frac{1}{n}\sum_{i=1}^n |\epsilon_{ni}|$ converges in probability to the absolute first moment of ϵ_{ni} and $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = O_P(1)$. These show that $T_{n2} = o_P(1)$.

Similarly, let $U_n = \sqrt{\frac{h_n}{n}} \epsilon_n' (C_n^d(\hat{\lambda}_n) - C_n^d(\lambda_0))\epsilon_n$. Then, $U_n = U_{n1} + U_{n2}$ where

$$\begin{aligned} U_{n1} &= \sqrt{\frac{h_n}{n}} \epsilon_n' (G_{ln} \sum_{k=1}^{m} [\sum_{j=1}^{p} (\hat{\lambda}_{nj} - \lambda_{j0}) G_{jn}]^k)^d \epsilon_n \\ &= \sum_{k=1}^{m} \sum_{j_1=1}^{p} \cdots \sum_{j_k=1}^{p} (\hat{\lambda}_{nj_1} - \lambda_{j_1 0}) \cdots (\hat{\lambda}_{nj_k} - \lambda_{j_k 0}) \cdot \sqrt{\frac{h_n}{n}} \epsilon_n' (G_{ln} G_{j_1 n} \cdots G_{j_k n})^d \epsilon_n = o_P(1) \end{aligned}$$

because $\sqrt{\frac{h_n}{n}} \epsilon_n' (G_{ln} G_{j_1 n} \cdots G_{j_k n})^d \epsilon_n = O_P(1)$ by Lemma A.4; and

$$\begin{aligned} U_{n2} &= \sqrt{\frac{h_n}{n}} \epsilon_n' (W_{ln} S_n^{-1}(\hat{\lambda}_n) [\sum_{j=1}^{p} (\hat{\lambda}_{nj} - \lambda_{j0}) G_{jn}]^{m+1})^d \epsilon_n \\ &= \sqrt{\frac{h_n}{n}} \sum_{j_1=1}^{p} \cdots \sum_{j_{m+1}=1}^{p} (\hat{\lambda}_{nj_1} - \lambda_{j_1 0}) \cdots (\hat{\lambda}_{nj_{m+1}} - \lambda_{j_{m+1}0}) \epsilon_n' (W_{ln} S_n^{-1}(\hat{\lambda}_n) G_{j_1 n} \cdots G_{j_{m+1} n})^d \epsilon_n. \end{aligned}$$

The term $U_{n2} = o_P(1)$ because

$$\begin{aligned} \parallel U_{n2} \parallel &\leq c \sum_{j_1=1}^{p} \cdots \sum_{j_{m+1}=1}^{p} |\sqrt{\frac{n}{h_n}} (\hat{\lambda}_{nj_1} - \lambda_{j_1 0})| \cdots |\sqrt{\frac{n}{h_n}} (\hat{\lambda}_{nj_{m+1}} - \lambda_{j_{m+1}0})| h_n^{\frac{1}{2}} n^{\frac{3}{2}} (\frac{h_n}{n})^{\frac{m+1}{2}} (\frac{1}{n} \sum_{i=1}^{n} |\epsilon_{ni}|)^2 \\ &= o_P(1). \end{aligned}$$

Q.E.D.

Lemma A.10 *Suppose that the elements of the $n \times k$ matrix X_n are uniformly bounded, and the limit $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular, then*

(i) the projectors M_n and $(I_n - M_n)$, where $M_n = X_n (X_n' X_n)^{-1} X_n'$, are uniformly bounded in both row and column sums; and

(ii) $\epsilon_n' A_n' M_n B_n \epsilon_n = O_P(1)$ for any $n \times n$ matrices A_n and B_n uniformly bounded in column sums.

Proof: Part (i) is a result in Lee (1999b). For (ii), because $\frac{1}{\sqrt{n}}X_n' A_n \epsilon_n$ and $\frac{1}{\sqrt{n}}X_n' B_n \epsilon_n$ are of order $O_P(1)$ by Lemma A.6, $\epsilon_n' A_n' M_n B_n \epsilon_n = \frac{1}{\sqrt{n}}\epsilon_n' A_n' X_n(\frac{1}{n}X_n' X_n)^{-1}\frac{1}{\sqrt{n}}X_n' B_n \epsilon_n = O_P(1)$. Q.E.D.

Lemma A.11 *Suppose that the elements of the $n \times k$ matrix C_n are uniformly bounded, the $n \times n$ matrix A_n is uniformly bounded in column sums, $\hat{\lambda}_n$ is a $\sqrt{\frac{n}{h_n}}$-consistent estimator, and $\frac{h_n^{1+\delta}}{n} = o(1)$ for some $\delta > 0$. Then, $\frac{1}{\sqrt{n}}C_n'(G_{ln}(\hat{\lambda}_n))^L A_n \epsilon_n = O_P(1)$, where $G_{ln}(\lambda) = W_{ln}S_n^{-1}(\lambda)$ with $S_n(\lambda) = I_n - \sum_{j=1}^p \lambda_j W_{jn}$.*

Proof: With (A.1), $G_{ln}(\hat{\lambda}_n)$ can be expanded as

$$G_{ln}(\hat{\lambda}_n) = G_{ln} + G_{ln}\sum_{k=1}^m [\sum_{j=1}^p (\hat{\lambda}_{nj} - \lambda_{j0})G_{jn}]^k + G_{ln}(\hat{\lambda}_n)[\sum_{j=1}^p (\hat{\lambda}_{nj} - \lambda_{j0})G_{jn}]^{m+1}.$$

It follows that $\frac{1}{\sqrt{n}}C_n'(G_{ln}(\hat{\lambda}_n))^L A_n \epsilon_n = \frac{1}{\sqrt{n}}C_n' G_{ln}^L A_n \epsilon_n + R_{n1} + R_{n2}$ where

$$R_{n1} = \sum_{k=1}^m \sum_{j_1=1}^p \cdots \sum_{j_k=1}^p (\hat{\lambda}_{nj_1} - \lambda_{j_1 0})\cdots(\hat{\lambda}_{nj_k} - \lambda_{j_k 0}) \cdot \frac{1}{\sqrt{n}}C_n'(G_{ln}G_{j_1 n}\cdots G_{j_k n})^L A_n \epsilon_n,$$

and $R_{n2} = \frac{1}{\sqrt{n}}\sum_{j_1=1}^p \cdots \sum_{j_{m+1}=1}^p (\hat{\lambda}_{nj_1} - \lambda_{j_1 0})\cdots(\hat{\lambda}_{nj_{m+1}} - \lambda_{j_{m+1}0})C_n'(G_{ln}(\hat{\lambda}_n)G_{j_1 n}\cdots G_{j_{m+1}n})^L A_n \epsilon_n$. The term R_{n1} is of order $o_P(1)$ because $\frac{1}{\sqrt{n}}C_n'(G_{ln}G_{j_1 n}\cdots G_{j_k n})^L A_n \epsilon_n = O_P(1)$ by Lemma A.6 and $\hat{\lambda}_n - \lambda_0 = o_P(1)$. For R_{n2}, with either maximum row or column sum norm $\| \cdot \|$,

$$\begin{aligned}
\| R_{n2} \| &\leq n^{-1/2} \| C_n' \| \cdot \| \epsilon_n \| \\
&\quad \cdot \sum_{j_1=1}^p \cdots \sum_{j_{m+1}=1}^p |\hat{\lambda}_{nj_1} - \lambda_{j_1 0}| \cdots |\hat{\lambda}_{nj_{m+1}} - \lambda_{j_{m+1}0}| \cdot \| (G_{ln}(\hat{\lambda}_n)G_{j_1 n}\cdots G_{j_{m+1}n})^L A_n \| \\
&\leq cn^{2-1/2}(\frac{h_n}{n})^{m+1}\sum_{i=1}^k (\frac{1}{n}\sum_{j=1}^n |c_{n,ij}|) \cdot \frac{1}{n}\sum_{l=1}^n |\epsilon_{nl}| \\
&\quad \cdot \sum_{j_1=1}^p \cdots \sum_{j_{m+1}=1}^p |\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{nj_1} - \lambda_{j_1 0})| \cdots |\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{nj_{m+1}} - \lambda_{j_{m+1}0})| = o_P(1),
\end{aligned}$$

by using a large enough m in expansion. Hence $\frac{1}{\sqrt{n}}C_n'(G_{ln}(\hat{\lambda}_n))^L A_n \epsilon_n = \frac{1}{\sqrt{n}}C_n' G_{ln}^L A_n \epsilon_n + o_P(1)$. The final result follows from Lemma A.6. Q.E.D.

Lemma A.12 *Suppose that A_n, B_n and C_n are matrices uniformly bounded in column sums, X_n satisfies the assumptions in Lemma A.10, $\hat{\lambda}_n$ is $\sqrt{\frac{n}{h_n}}$-consistent, and $\frac{h_n^{1+\delta}}{n} = o_P(1)$ for some $\delta > 0$. Then, $\epsilon_n' A_n'(G_{ln}(\hat{\lambda}_n))^L B_n' M_n C_n \epsilon_n = O_P(1)$, where $M_n = X_n(X_n' X_n)^{-1}X_n'$ and $G_{ln}(\lambda) = W_{ln}S_n^{-1}(\lambda)$ where $S_n(\lambda) = I_n - \sum_{j=1}^p \lambda_j W_{jn}$.*

Proof: As B_n' is uniformly bounded in row sums and elements of X_n are uniformly bounded, elements of $B_n' X_n$ are uniformly bounded. Hence, by Lemmas A.6 and A.11,

$$\epsilon_n' A_n'(G_{ln}(\hat{\lambda}_n))^L B_n' M_n C_n \epsilon_n = (\frac{1}{\sqrt{n}}\epsilon_n' A_n'(G_{ln}(\hat{\lambda}_n))^L B_n' X_n)\left(\frac{1}{n}X_n' X_n\right)^{-1}(\frac{1}{\sqrt{n}}X_n' C_n \epsilon_n) = O_P(1).$$

Q.E.D.

Appendix B: Proofs

Proof of Proposition 2.1: $E[(P_nS_nY_n)'\epsilon_n] = E[(P_n\epsilon_n)'\epsilon_n] = E(\epsilon_n'P_n\epsilon_n) = \sigma_0^2 tr(P_n) = 0.$ Q.E.D.

Proof of Proposition 2.2: Consider (2.7). Denote $a_n = Y_n'W_n'P_nW_nY_n$, $b_n = Y_n'P_n^sW_nY_n$, and $c_n = Y_n'P_nY_n$. As $b_n = \epsilon_n'S_n^{'-1}P_n^sG_n\epsilon_n$ and elements of G_n have the uniform order $O(\frac{1}{h_n})$, Lemma A.3 together with Lemma A.1 imply that $\frac{h_n}{n}(b_n - E(b_n)) = o_P(1)$, where $E(b_n) = \sigma_0^2 tr(S_n^{'-1}P_n^sG_n)$. Similarly, as $a_n = \epsilon_n'G_n'P_nG_n\epsilon_n$, $\frac{h_n}{n}(a_n - E(a_n)) = o_P(1)$, where $E(a_n) = \sigma_0^2 tr(G_n'P_nG_n)$. For c_n, because $S_n^{-1} = I_n + \lambda_0 G_n$, one has the expansion that $c_n = \epsilon_n'S_n^{'-1}P_nS_n^{-1}\epsilon_n = \epsilon_n'(P_n + \lambda_0 P_n^sG_n + \lambda_0^2 G_n'P_nG_n)\epsilon_n$. Because elements of P_n have the uniform order $O(\frac{1}{h_n})$ and P_n is uniformly bounded in both row and column sums by Assumption 4, $\frac{h_n}{n}[\epsilon_n'P_n\epsilon_n - \sigma_0^2 tr(P_n)] = o_p(1)$ by Lemma A.3. Similarly, the probability convergence holds for the other two terms in the expansion of c_n. Hence, $\frac{h_n}{n}(c_n - E(c_n)) = o_P(1)$, where $E(c_n) = \sigma_0^2 tr(S_n^{'-1}P_nS_n^{-1}) = \sigma_0^2\lambda_0[tr(P_n^sG_n) + \lambda_0 tr(G_n'P_nG_n)]$, by using $tr(P_n) = 0$. These implies that

$$\hat{\lambda}_n - \left\{\frac{h_n}{n}E(b_n) - \left[(\frac{h_n}{n}E(b_n))^2 - 4\frac{h_n}{n}E(a_n)\frac{h_n}{n}E(c_n)\right]^{1/2}\right\} / \left(2\frac{h_n}{n}E(a_n)\right) = o_p(1) \qquad (B.1)$$

(White (1984), Prop. 2.30). Because $S_n^{-1} = I_n + \lambda_0 G_n$, it follows that

$$\begin{aligned}&\left(\frac{h_n}{n}E(b_n)\right)^2 - 4\frac{h_n}{n}E(a_n)\frac{h_n}{n}E(c_n)\\&= \sigma_0^4(\frac{h_n}{n})^2[tr^2(S_n^{'-1}P_n^sG_n) - 4tr(G_n'P_nG_n)tr(S_n^{'-1}P_nS_n^{-1})] = \sigma_0^4(\frac{h_n}{n})^2 tr^2(P_n^sG_n)\end{aligned}$$

and, hence, if $tr(P_n^sG_n)$ were positive,

$$\begin{aligned}&\frac{h_n}{n}E(b_n) - \left[\left(\frac{h_n}{n}E(b_n)\right)^2 - 4\frac{h_n}{n}E(a_n)\frac{h_n}{n}E(c_n)\right]^{1/2}\\&= \sigma_0^2\frac{h_n}{n}[tr(S_n^{'-1}P_n^sG_n) - tr(P_n^sG_n)] = \sigma_0^2\frac{h_n}{n}\lambda_0 tr(G_n'P_n^sG_n) = 2\frac{h_n}{n}E(a_n)\lambda_0.\end{aligned}$$

Therefore, $\hat{\lambda}_n - \lambda_0 = o_P(1)$ from (B.1) and $\hat{\lambda}_n$ converges in probability to λ_0. Otherwise, $\hat{\lambda}_n$ in (2.8) will be the consistent one.

With $\hat{\lambda}_n = Y_n'P_nY_n/Y_n'P_n^sW_nY_n$ in the remaining part of the proposition, Lemma A.3 implies that $\hat{\lambda}_n - E(c_n)/E(b_n) = o_P(1)$. Suppose that $\lim_{n\to\infty}\frac{h_n}{n}tr(G_n'P_nG_n) = 0$. It follows that

$$\frac{E(c_n)}{E(b_n)} = \lambda_0\frac{tr(P_n^sG_n)}{tr(S_n^{'-1}P_n^sG_n)} + \lambda_0^2\frac{tr(G_n'P_nG_n)}{tr(S_n^{'-1}P_n^sG_n)} = \lambda_0\frac{tr(P_n^sG_n)}{tr(S_n^{'-1}P_n^sG_n)} + o(1) = \lambda_0 + o(1),$$

because $\frac{h_n}{n}tr(S_n^{'-1}P_n^sG_n) = \frac{h_n}{n}tr(P_n^sG_n) + \lambda_0\frac{h_n}{n}tr(G_n'P_n^sG_n) = \frac{h_n}{n}tr(P_n^sG_n) + o(1)$ by using $S_n^{-1} = I_n + \lambda_0 G_n$. Therefore, $\hat{\lambda}_n - \lambda_0 = o_P(1)$. Q.E.D.

Proof of Proposition 2.3: Lemmas A.1 and A.3 imply that $\frac{h_n}{n}[\epsilon_n' H_n \epsilon_n - \sigma_0^2 tr(H_n)] = o_p(1)$ where $H_n = S_n'^{-1} P_n S_n^{-1}$, $G_n' P_n^s S_n^{-1}$, or $G_n' P_n G_n$ in this proposition. As $g_n(\lambda) = \epsilon_n' S_n'^{-1} P_n S_n^{-1}\epsilon_n - \lambda \epsilon_n' G_n' P_n^s S_n^{-1}\epsilon_n + \lambda^2 \epsilon_n' G_n' P_n G_n \epsilon_n$, it follows that $\frac{h_n}{n} g_n(\lambda) - \frac{h_n}{n} E(g_n(\lambda)) = o_P(1)$ and, hence, $(\frac{h_n}{n} g_n(\lambda))^2 - Q_n(\lambda) = o_P(1)$, where $Q_n(\lambda) = (\frac{h_n}{n} E(g_n(\lambda))^2$, uniformly in λ in any bounded subset of λ.

As $E(g_n(\lambda_0)) = 0$ because $tr(P_n) = 0$, it remains to show that λ_0 is a strict local minimizer of $Q_n(\lambda)$ for large n. The first and second order derivatives of $Q_n(\lambda)$ are $\frac{dQ_n(\lambda)}{d\lambda} = 2(\frac{h_n}{n})^2 E(g_n(\lambda)) \frac{dE(g_n(\lambda))}{d\lambda}$ and $\frac{d^2 Q_n(\lambda)}{d\lambda^2} = 2(\frac{h_n}{n})^2 \{(\frac{dE(g_n(\lambda)}{d\lambda})^2 + E(g_n(\lambda)) \frac{d^2 E(g_n(\lambda))}{d\lambda^2}\}$. At λ_0,

$$\frac{d^2 Q_n(\lambda_0)}{d\lambda^2} = 2(\frac{h_n}{n})^2 \left(\frac{dE(g_n(\lambda_0))}{d\lambda}\right)^2 = 2(\frac{h_n}{n})^2 \sigma_0^4 [2\lambda_0 tr(G_n' P_n G_n) - tr(G_n' P_n^s S_n^{-1})]^2 = 2\sigma_0^4 \left[\frac{h_n}{n} tr(P_n^s G_n)\right]^2,$$

where the last expression follows because $2\lambda_0 tr(G_n' P_n G_n) - tr(G_n' P_n^s S_n^{-1}) = -tr(P_n^s G_n)$ by using $S_n^{-1} = I_n + \lambda_0 G_n$. Under the assumed regularity condition, $(\frac{h_n}{n} tr(P_n^s G_n))^2 > 0$ for large n and, hence, λ_0 is a strict local minimizer.

The consistency of $\hat{\lambda}_n$ follows from the uniform convergence in probability of $(\frac{h_n}{n} g_n(\lambda) - Q_n(\lambda))$ to zero and the local identification of λ_0 in Λ (White 1994, Theorem 3.4). Q.E.D.

Proof of Proposition 2.4: From the Taylor expansion $0 = g_n(\hat{\lambda}_n) = g_n(\lambda_0) + \frac{\partial g_n(\bar{\lambda}_n)}{\partial \lambda}(\hat{\lambda}_n - \lambda_0)$ where $\bar{\lambda}_n$ lies between $\hat{\lambda}_n$ and λ_0, and (2.9),

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = \left[\frac{h_n}{n} Y_n' S_n'(\bar{\lambda}_n) P_n^s W_n Y_n\right]^{-1} \sqrt{\frac{h_n}{n}} Y_n' S_n' P_n S_n Y_n = \left[\frac{h_n}{n} Y_n' S_n'(\bar{\lambda}_n) P_n^s W_n Y_n\right]^{-1} \sqrt{\frac{h_n}{n}} \epsilon_n' P_n \epsilon_n,$$

because $Y_n' S_n' P_n S_n Y_n = \epsilon_n' P_n \epsilon_n$. Explicitly,

$$\frac{h_n}{n} Y_n' S_n'(\bar{\lambda}_n) P_n^s W_n Y_n = \frac{h_n}{n} Y_n' W_n' P_n^s S_n Y_n - (\bar{\lambda}_n - \lambda_0) \frac{h_n}{n} Y_n' W_n' P_n^s W_n Y_n.$$

Lemma A.3 implies that $\frac{h_n}{n} Y_n' W_n' P_n^s W_n Y_n = \frac{h_n}{n} \epsilon_n' G_n' P_n^s G_n \epsilon_n = O_p(1)$ and

$$\frac{h_n}{n}[Y_n' W_n' P_n^s S_n Y_n - E(Y_n' W_n' P_n^s S_n Y_n)] = \frac{h_n}{n}[\epsilon_n' G_n' P_n^s \epsilon_n - \sigma_0^2 tr(P_n^s G_n)] = o_p(1).$$

As $\bar{\lambda}_n - \lambda_0 = o_p(1)$, it follows that

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = \left[\sigma_0^2 \frac{h_n}{n} tr(P_n^s G_n) + o_p(1)\right]^{-1} \sqrt{\frac{h_n}{n}} \epsilon_n' P_n \epsilon_n. \qquad (B.2)$$

As $E(\epsilon_n' P_n \epsilon_n) = 0$, and $\frac{h_n}{n} \text{var}(\epsilon_n' P_n \epsilon_n) = (\mu_4 - 3\sigma_0^4) \frac{h_n}{n} \sum_{i=1}^n p_{n,ii}^2 + \sigma_0^4 \frac{h_n}{n} tr(P_n P_n^s) = O(1)$ from Lemmas A.2 and A.1, Lemma A.4 implies that $\epsilon_n' P_n \epsilon_n / \text{var}^{\frac{1}{2}}(\epsilon_n' P_n \epsilon_n) \xrightarrow{D} N(0,1)$. The asymptotic distribution for $\hat{\lambda}_n$ follows from (B.2).

For the special case that $\epsilon_n \sim N(0, \sigma_0^2 I_n)$, $\kappa_4 - 3 = 0$. When $P_n \in \mathcal{P}_{2n}$, $p_{n,ii} = 0$ for all $i = 1, \cdots, n$. Finally, as $\sum_{i=1}^n p_{n,ii}^2 = O(\frac{n}{h_n^2})$, $\sum_{i=1}^n p_{n,ii}^2/(\frac{h_n}{n} tr^2(P_n^s G_n)) = O(\frac{1}{h_n}) = o(1)$ if $\lim_{n\to\infty} h_n = \infty$. Q.E.D.

Proof of Lemma 2.1: As $\epsilon_n' A\epsilon_n \epsilon_n' B\epsilon_n = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij} b_{kl} \epsilon_{ni}\epsilon_{nj}\epsilon_{nk}\epsilon_{nl}$, the mutual independence of ϵ_{ni}s implies that $E(\epsilon_{ni}\epsilon_{nj}\epsilon_{nk}\epsilon_{nl}) \neq 0$ only if $(i = j = k = l)$, $(i = j, k = l)$, $(i = k, j = l)$, or $(i = l, j = k)$. It follows that

$$\begin{aligned} E(\epsilon_n' A\epsilon_n \cdot \epsilon_n' B\epsilon_n) &= \sum_{i=1}^n a_{ii} b_{ii} E(\epsilon_{ni}^4) + \sum_{i=1}^n \sum_{j\neq i}^n (a_{ii} b_{jj} + a_{ij} b_{ij} + a_{ij} b_{ji}) E(\epsilon_{ni}^2 \epsilon_{nj}^2) \\ &= (\mu_4 - 3\sigma_0^4) \sum_{i=1}^n a_{ii} b_{ii} + \sigma_0^4 \sum_{i=1}^n \sum_{j=1}^n (a_{ii} b_{jj} + a_{ij} b_{ij} + a_{ij} b_{ji}) \\ &= (\mu_4 - 3\sigma_0^4) vec_D'(A) vec_D(B) + \sigma_0^4 [tr(A) tr(B) + tr(AB') + tr(AB)]. \end{aligned}$$

Q.E.D.

Proof of Proposition 2.5: By a similar argument in the proof of Proposition (2.3), $\frac{h_n}{n} a_n [g_n(\lambda) - E(g_n(\lambda))] = o_p(1)$ uniformly in λ in any bounded set. It follows that

$$(\frac{h_n}{n})^2 g_n'(\lambda) a_n' a_n g_n(\lambda) - (\frac{h_n}{n})^2 E(g_n'(\lambda)) a_n' a_n E(g_n(\lambda)) = o_p(1)$$

uniformly in $\lambda \in \Lambda$. As λ_0 is the unique minimizer of $\lim_{n\to\infty} (\frac{h_n}{n})^2 E(g_n(\lambda)) a_n' a_n E(g_n(\lambda))$, the consistency of $\hat{\lambda}_n$ follows from the uniform convergence of λ in Λ and the identification uniqueness of λ_0 in Λ.

As $\frac{\partial g_n'(\lambda)}{\partial \lambda} = -D_n'(\lambda)$ where $D_n(\lambda) = (Y_n' S_n'(\lambda) P_{1n}^s W_n Y_n, \cdots, Y_n' S_n'(\lambda) P_{mn}^s W_n Y_n)'$, the Taylor expansion of $\frac{\partial g_n'(\hat{\lambda}_n)}{\partial \lambda} a_n' a_n g_n(\hat{\lambda}_n) = 0$ at λ_0 implies that

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = \left[\frac{h_n}{n} D_n'(\bar{\lambda}_n) a_n' a_n \frac{h_n}{n} D_n(\bar{\lambda}_n)\right]^{-1} \frac{h_n}{n} D_n'(\bar{\lambda}_n) a_n' a_n \sqrt{\frac{h_n}{n}} (\epsilon_n' P_{1n}\epsilon_n, \cdots, \epsilon_n' P_{mn}\epsilon_n)'.$$

Because $\frac{h_n}{n} Y_n' S_n'(\bar{\lambda}_n) P_{ln}^s W_n Y_n = \sigma_0^2 \frac{h_n}{n} tr(P_{ln}^s G_n) + o_P(1)$,

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = [\frac{h_n}{n} \sigma_0^4 d_n' a_n' a_n d_n + o_p(1)]^{-1} (\sigma_0^2 \frac{h_n}{n} d_n' + o_p(1)) a_n' a_n \sqrt{\frac{h_n}{n}} (\epsilon_n' P_{1n}\epsilon_n, \cdots, \epsilon_n' P_{mn}\epsilon_n)'. \quad (B.3)$$

As $a_n \sqrt{\frac{n}{h_n}} (\epsilon_n' P_{1n}\epsilon_n, \cdots, \epsilon_n' P_{mn}\epsilon_n)' = \sqrt{\frac{h_n}{n}} \epsilon_n' (\sum_{j=1}^m a_{nj} P_{jn})\epsilon_n$, where $a_n = (a_{n1}, \cdots, a_{nm})$, the central limit theorem in Lemma A.4 is applicable to this quadratic form. Lemma A.2 implies that

$$\begin{aligned} &\text{var}(\epsilon_n' (\sum_{j=1}^m a_{nj} P_{jn})\epsilon_n) \\ &= (\mu_4 - 3\sigma_0^4) \sum_{i=1}^n ((\sum_{j=1}^m a_{nj} P_{jn})_{ii})^2 + \sigma_0^4 tr[(\sum_{j=1}^m a_{nj} P_{jn})(\sum_{l=1}^m a_{nl} P_{ln}^s)] \\ &= \sigma_0^4 \{(\kappa_4 - 3) a_n (vec_D(P_{1n}), \cdots, vec_D(P_{mn}))' (vec_D(P_{1n}), \cdots, vec_D(P_{mn})) a_n' + a_n V_n a_n'\} \\ &= \sigma_0^4 a_n \Omega_n a_n'. \end{aligned}$$

Hence, $\sqrt{\frac{h_n}{n}} a_n(\epsilon_n' P_{1n}\epsilon_n, \cdots, \epsilon_n' P_{mn}\epsilon_n)' \xrightarrow{D} N(0, \sigma_0^4 a_0(\lim_{n\to\infty} \frac{h_n}{n}\Omega_n)a_0')$. The asymptotic distribution of $\hat{\lambda}_n$ follows from the expansion $(B.3)$. Q.E.D.

Proof of Proposition 2.6: From (2.16), the generalized Schwartz inequality shows that the optimal weighting matrix of $a_0'a_0$ is the limit of $\frac{h_n}{n}\Omega_n$. When ϵ_n is normally distributed, $\kappa_4 = 3$ and $\Omega_n = V_n$. If P_{jn}s are from $\mathcal{P}_{2n}$, $vec_D(P_{jn}) = 0$ and $\Omega_n = V_n$. For the case that $\lim_{n\to\infty} h_n = \infty$, $\frac{h_n}{n} vec_D'(P_{jn})vec_D(P_{ln}) = \frac{h_n}{n} O(\frac{n}{h_n^2}) = O(\frac{1}{h_n}) = o(1)$ because the elements of P_ns are of order $O(\frac{1}{h_n})$, and, hence, $\lim_{n\to\infty} \frac{h_n}{n}(\Omega_n - V_n) = 0$.

The consistency and asymptotic distribution shall follow by showing that the stochastic $\hat{\Omega}_n$ can be replaced by the nonstochastic Ω_n. For consistency, because

$$\frac{h_n}{n} g_n'(\lambda)\hat{\Omega}_n^{-1} g_n(\lambda) = \frac{h_n}{n} g_n'(\lambda)\Omega_n^{-1} g_n(\lambda) + \frac{h_n}{n} g_n'(\lambda)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\lambda),$$

it is sufficient to show that $\frac{h_n}{n} g_n'(\lambda)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\lambda) = o_p(1)$ uniformly in $\lambda \in \Lambda$. Let $\| \ \|$ be the Euclidean or maximum row sum norm for vectors and matrices. Then

$$\| \frac{h_n}{n} g_n'(\lambda)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\lambda) \| \leq (\frac{h_n}{n} \| g_n(\lambda) \|)^2 \| (\frac{h_n}{n}\hat{\Omega}_n)^{-1} - (\frac{h_n}{n}\Omega_n)^{-1} \| .$$

From the proofs of preceding propositions, $\frac{h_n}{n}[g_n(\lambda) - E(g_n(\lambda))]$ has the order $o_P(1)$ uniformly in $\lambda \in \Lambda$, i.e., $\sup_{\lambda\in\Lambda} \| \frac{h_n}{n}(g_n(\lambda) - E(g_n(\lambda)) \| = o_P(1)$. On the other hand,

$$\begin{aligned} \| \frac{h_n}{n} E(g_n(\lambda)) \| &\leq \sigma_0^2 \max_{j=1,\cdots,m} \left\{ \left| \frac{h_n}{n} tr(S_n'^{-1} P_{jn} S_n^{-1}) \right| + \lambda \left| \frac{h_n}{n} tr(G_n' P_{jn}^s S_n^{-1}) \right| + \lambda^2 \left| \frac{h_n}{n} tr(G_n' P_{jn} G_n) \right| \right\} \\ &= O(1) \end{aligned}$$

uniformly in $\lambda \in \Lambda$. Hence, $\| \frac{h_n}{n} g_n(\lambda) \| \leq \| \frac{h_n}{n} E(g_n(\lambda)) \| + \| \frac{h_n}{n}[g_n(\lambda) - E(g_n(\lambda))] \| = O_P(1)$. Therefore, $\frac{h_n}{n} g_n'(\lambda)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\lambda) = o_p(1)$ uniformly in $\lambda \in \Lambda$.

For its limiting distribution, it is sufficient to show that $\hat{\Omega}_n^{-1}$ can be replaced by Ω_n^{-1} in the Taylor expansion that $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{v,n} - \lambda_0) = \left[\frac{h_n}{n} D_n'(\bar{\lambda}_n)\hat{\Omega}_n^{-1} D_n(\bar{\lambda}_n)\right]^{-1} D_n'(\bar{\lambda}_n)\hat{\Omega}_n^{-1} \cdot \sqrt{\frac{h_n}{n}}(\epsilon_n' P_{1n}\epsilon_n, \cdots, \epsilon_n' P_{mn}\epsilon_n)'$. As in the proof of the preceding proposition, one has $\frac{h_n}{n} D_n(\bar{\lambda}_n) = O_P(1)$. Thus,

$$D_n'(\bar{\lambda}_n)\hat{\Omega}_n^{-1} = D_n'(\bar{\lambda}_n)\Omega_n^{-1} + \frac{h_n}{n} D_n'(\bar{\lambda}_n)((\frac{h_n}{n}\hat{\Omega}_n)^{-1} - (\frac{h_n}{n}\Omega_n)^{-1}) = D_n'(\bar{\lambda}_n)\Omega_n^{-1} + o_P(1)$$

and $\frac{h_n}{n} D_n'(\bar{\lambda}_n)\hat{\Omega}_n^{-1} D_n(\bar{\lambda}_n) = \frac{h_n}{n} D_n'(\bar{\lambda}_n)\Omega_n^{-1} D_n(\bar{\lambda}_n) + o_P(1)$. The asymptotic distribution follows.

For the overidentification test, by the Taylor expansion, $g_n(\hat{\lambda}_{v,n}) = g_n(\lambda_0) - D_n(\bar{\lambda}_n)(\hat{\lambda}_{v,n} - \lambda_0)$. It follows that

$$\begin{aligned} \sqrt{\frac{h_n}{n}} g_n(\hat{\lambda}_{v,n}) &= \sqrt{\frac{h_n}{n}} g_n(\lambda_0) - \frac{h_n}{n} D_n(\bar{\lambda}_n) \cdot \sqrt{\frac{n}{h_n}}(\hat{\lambda}_{v,n} - \lambda_0) \\ &= \{I_n - D_n(\bar{\lambda}_n)[D_n'(\bar{\lambda}_n)\hat{\Omega}_n^{-1} D_n(\bar{\lambda}_n)]^{-1} D_n'(\bar{\lambda}_n)\hat{\Omega}_n^{-1}\}\sqrt{\frac{h_n}{n}} g_n(\lambda_0), \end{aligned}$$

and

$$g_n'(\hat{\lambda}_{v,n})(\hat{\Omega}_n)^{-1}g_n(\hat{\lambda}_{v,n}) = \sqrt{\frac{h_n}{n}}g_n'(\lambda_0)(\frac{h_n}{n}\hat{\Omega}_n)^{-1/2}\{I_n - \hat{\Omega}_n^{-1/2}D_n(\bar{\lambda}_n)[D_n'(\bar{\lambda}_n)\hat{\Omega}_n^{-1}D_n(\bar{\lambda}_n)]^{-1} \cdot D_n'(\bar{\lambda}_n)\hat{\Omega}_n^{-1/2}\}(\frac{h_n}{n}\hat{\Omega}_n)^{-1/2}\sqrt{\frac{h_n}{n}}g_n(\lambda_0).$$

From the proof of Proposition 2.5, $\sqrt{\frac{h_n}{n}}g_n(\lambda_0) \xrightarrow{D} N(0, \sigma_0^4 \lim_{n\to\infty} \frac{h_n}{n}\Omega_n)$. Hence, $g_n'(\hat{\lambda}_{v,n})(\hat{\Omega}_n)^{-1}g_n(\hat{\lambda}_{v,n}) \xrightarrow{D} \sigma_0^4\chi^2(m-1)$. Q.E.D.

Proof of Lemma 2.2: (i) follows because $tr(A \cdot \frac{tr(B)}{n}I_n) = \frac{tr(B)}{n}tr(A) = 0$. (ii) follows because $tr(A \cdot Diag(B)) = tr(Diag(A) \cdot Diag(B)) = 0$. Q.E.D.

Proof or Proposition 2.7: For any two squares matrices A and B, the Cauchy-Schwartz inequality implies that $|tr(AB)|^2 \leq tr(A^2)tr(B^2)$ (see, e.g., Zhang 1999, p.25). As P_n^s is symmetric, $tr(P_n^sG_n) = tr(G_n'P_n^s) = tr(P_n^sG_n')$ and, hence, $tr(P_n^sG_n) = \frac{1}{2}tr(P_n^sG_n^s)$.

Suppose $P_n \in \mathcal{P}_{1n}$, Lemma 2.2 implies that $tr(P_n^sG_n^s) = tr[P_n^s(G_n^s - 2\frac{tr(G_n)}{n}I_n)] = tr[P_n^s(G_n - \frac{tr(G_n)}{n}I_n)^s]$. The Cauchy-Schwartz inequality implies that $tr^2(P_n^s(G_n - \frac{tr(G_n)}{n}I_n)^s) \leq tr(P_n^{s2})tr[(G_n - \frac{tr(G_n)}{n}I_n)^{s2}]$, and, hence, $tr^2(P_n^s(G_n - \frac{tr(G_n)}{n}I_n)^s)/tr(P_n^{s2}) \leq tr[(G_n - \frac{tr(G_n)}{n}I_n)^{s2}]$. The last equality holds because $tr[(G_n - \frac{tr(G_n)}{n}I_n)^sG_n^s] = tr[(G_n - \frac{tr(G_n)}{n}I_n)^{s2}]$.

Similar arguments are applicable to $P_n \in \mathcal{P}_{2n}$ using the property $tr(P_n^sG_n^s) = tr(P_n^s(G_n - Diag(G_n))^s)$ from Lemma 2.2. Q.E.D.

Proof of Proposition 2.8: For any matrix $P_n \in \mathcal{M}_{2n}$, $tr(P_n^sG_n) = \frac{1}{2}tr[P_n^s(G_n^s - 2Diag(G_n))]$. The matrix Σ_{vn} in Proposition 2.6 can be rewritten as

$$\Sigma_{vn} = \frac{1}{4}(tr[P_{1n}^s(G_n^s - 2Diag(G_n))], \cdots, tr[P_{mn}^s(G_n^s - 2Diag(G_n))]) \cdot V_n^{-1}(tr[P_{1n}^s(G_n^s - 2Diag(G_n))], \cdots, tr[P_{mn}^s(G_n^s - 2Diag(G_n))])'. \quad (B.4)$$

Note that $\Sigma_b = (\lim_{n\to\infty} \frac{h_n}{n}\Sigma_{bn})^{-1}$ where

$$\Sigma_{bn} = tr^2[(G_n^s - 2Diag(G_n))G_n]/tr[(G_n - Diag(G_n))(G_n^s - 2Diag(G_n))] = \frac{1}{2}tr[(G_n^s - 2Diag(G_n))(G_n^s - 2Diag(G_n)]$$

from Proposition (2.6). This is so, because, as $G_n - Diag(G_n) \in \mathcal{P}_{2n}$, its corresponding $\Omega_n = V_n = tr[(G_n - Diag(G_n))(G_n - Diag(G_n))^s]$ in (2.14). So it is sufficient to compare Σ_{bn} with Σ_{vn}. We note that for any conformable matrices A and B, $tr(AB) = vec'(A')vec(B)$. Let $C = vec(G_n^s - 2Diag(G_n))$ and $D = (vec(P_{1n}^s), \cdots, vec(P_{mn}^s))$. Then, from (B.4) and the second expression of V_n in (2.14), $\Sigma_{vn} = \frac{1}{2}C'D(D'D)^{-1}D'C$ and $\Sigma_{bn} = \frac{1}{2}C'C$. By the generalized Schwartz inequality, $\Sigma_{vn} \leq \Sigma_{bn}$.

Similar argument is applicable to $\mathcal{M}_{1n}$ by using $tr(P_n^s G_n) = \frac{1}{2} tr(P_n^s(G_n^s - 2\frac{tr(G_n)}{n} I_n))$ for $P_n \in \mathcal{P}_{1n}$ when $\epsilon_n \sim N(0, \sigma_0^2 I_n)$ or $\lim_{n\to\infty} h_n = \infty$. This is so, because when $\epsilon_n \sim N(0, \sigma_0^2 I_n)$, $\Omega_n = V_n$; and for the case that $\lim_{n\to\infty} h_n = \infty$, $\lim_{n\to\infty} \frac{h_n}{n}(\Omega_n - V_n) = 0$. Q.E.D.

Proof of Proposition 2.9: The asymptotic distribution of $\hat{\lambda}_{1b,n}$ follows from (2.10) with $P_n = G_n - \frac{tr(G_n)}{n} I_n$. It is shown in Lee (1999b) that the QMLE estimator $\hat{\lambda}_{QM,n}$ has the asymptotic distribution that $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_{QM,n} - \lambda_0) \xrightarrow{D} N(0, \Sigma_{\lambda\lambda} + \Sigma_{\lambda\lambda}\Omega\Sigma_{\lambda\lambda})$ where $\Sigma_{\lambda\lambda} = (\lim_{n\to\infty} \frac{h_n}{n}[tr(C_n C_n') + tr(C_n^2)])^{-1}$ and $\Omega = (\kappa_4 - 3)\lim_{n\to\infty} \frac{h_n}{n} \sum_{i=1}^n C_{n,ii}^2$ with $C_n = G_n - \frac{tr(G_n)}{n} I_n$. The limiting distributions of $\hat{\lambda}_{1b,n}$ and $\hat{\lambda}_{QM,n}$ are exactly the same. Q.E.D.

Proof of Proposition 2.10: For consistency, it is sufficient to show that $\frac{h_n}{n}\hat{g}_n(\lambda) - \frac{h_n}{n} g_n(\lambda) = o_P(1)$ uniformly in $\lambda \in \Lambda$. Explicitly, $\frac{h_n}{n}(\hat{g}_n(\lambda) - g_n(\lambda)) = T_{n1} - \lambda T_{n2} + \lambda^2 T_{n3}$ where $T_{n1} = \frac{h_n}{n}\epsilon_n' S_n'^{-1}(\hat{G}_n - G_n)^d S_n^{-1}\epsilon_n$, $T_{n2} = \frac{h_n}{n}\epsilon_n' G_n'(\hat{G}_n^s - G_n^s)^d S_n^{-1}\epsilon_n$, and $T_{n3} = \frac{h_n}{n}\epsilon_n' G_n'(\hat{G}_n - G_n)^d G_n \epsilon_n$. The terms T_{nj}, $j = 1, 2, 3$, are all of order $o_P(1)$ by Lemma A.9. Hence $\frac{h_n}{n}\hat{g}_n(\lambda) - \frac{h_n}{n} g_n(\lambda) = o_P(1)$ uniformly in $\lambda \in \Lambda$. The consistency of $\hat{\lambda}_n$ follows from the first part of Lemma A.8.

For the asymptotic distribution, consider $\frac{h_n}{n}\frac{\partial g_n(\lambda)}{\partial \lambda}$ and $\sqrt{\frac{n}{h_n}} g_n(\lambda_0)$. As $S_n(\lambda) = S_n - (\lambda - \lambda_0) W_n$,

$$\begin{aligned}\frac{h_n}{n} Y_n' S_n'(\lambda)(\hat{G}_n^s)^d W_n Y_n &= \frac{h_n}{n}\epsilon_n' G_n'(\hat{G}_n^s)^d \epsilon_n - (\lambda - \lambda_0)\frac{h_n}{n}\epsilon_n' G_n'(\hat{G}_n^s)^d G_n \epsilon_n \\ &= \frac{h_n}{n}\epsilon_n' G_n'(G_n^s)^d \epsilon_n - (\lambda - \lambda_0)\frac{h_n}{n}\epsilon_n' G_n'(G_n^s)^d G_n \epsilon_n + R_{n1} + R_{n2},\end{aligned}$$

where $R_{n1} = \frac{h_n}{n}\epsilon_n' G_n'(\hat{G}_n^s)^d \epsilon_n - \frac{h_n}{n}\epsilon_n' G_n'(G_n^s)^d \epsilon_n$ and $R_{n2} = \frac{h_n}{n}\epsilon_n' G_n'(\hat{G}_n^s)^d G_n \epsilon_n - \frac{h_n}{n}\epsilon_n' G_n'(G_n^s)^d G_n \epsilon_n$. Lemma A.9 implies that both $R_{n1} = o_P(1)$ and $R_{n2} = o_P(1)$. Hence,

$$\frac{h_n}{n} Y_n' S_n'(\lambda)(\hat{G}_n^s)^d W_n Y_n = \frac{h_n}{n} Y_n' S_n'(\lambda)(G_n^s)^d W_n Y_n + o_P(1),$$

uniformly in $\lambda \in \Lambda$, i.e., $\frac{h_n}{n}(\frac{\partial \hat{g}_n(\lambda)}{\partial \lambda} - \frac{\partial g_n(\lambda)}{\partial \lambda}) = o_P(1)$ uniformly in $\lambda \in \Lambda$. For the other term,

$$\sqrt{\frac{h_n}{n}} Y_n' S_n'(\hat{G}_n)^d S_n Y_n = \sqrt{\frac{h_n}{n}}\epsilon_n' G_n^d \epsilon_n + \sqrt{\frac{h_n}{n}}\epsilon_n'[(\hat{G}_n)^d - G_n^d]\epsilon_n = \sqrt{\frac{h_n}{n}}\epsilon_n' G_n^d \epsilon_n + o_P(1),$$

by Lemma A.9 (ii), i.e., $\sqrt{\frac{n}{h_n}}(\hat{g}_n(\lambda_0) - g_n(\lambda_0)) = o_P(1)$. Hence, by Lemma A.8, the feasible GMM estimator derived from $\min_{\lambda\in\Lambda}[Y_n' S_n'(\lambda)(\hat{G}_n)^d S_n(\lambda) Y_n]^2$ has the same limiting distribution as that derived from $\min_{\lambda\in\Lambda}[Y_n' S_n'(\lambda) G_n^d S_n(\lambda) Y_n]^2$. Q.E.D.

Proof of Proposition 2.11: This is a special case of Proposition 2.2 with the qualification that G_n can be replaced by $\hat{G}_n$ as in Proposition 2.10. Q.E.D.

Proof of Proposition 3.1: To prove this proposition, we shall show that GMM moment equations using u_n^* and those using u_n (as if it is observed) satisfy the conditions in Lemma A.8.

As $u_n^* = (I_n - M_n)u_n$,

$$\begin{aligned} u_n^{*'} S_n'(\lambda) P_n S_n(\lambda) u_n^* &= u_n'(I_n - M_n) S_n'(\lambda) P_n S_n(\lambda)(I_n - M_n) u_n \\ &= u_n' S_n'(\lambda) P_n S_n(\lambda) u_n - u_n' M_n S_n'(\lambda) P_n^s S_n(\lambda) u_n + u_n' M_n S_n'(\lambda) P_n S_n(\lambda) M_n u_n \\ &= u_n' S_n'(\lambda) P_n S_n(\lambda) u_n + O_p(1) \end{aligned}$$

uniformly in $\lambda \in \Lambda$ (as λ is linear in $S_n(\lambda)$) by Lemma A.10 after substituting u_n by $S_n^{-1}\epsilon_n$. Hence, in particular, $\frac{h_n}{n}(u_n^{*'} S_n'(\lambda) P_n S_n(\lambda) u_n^* - u_n' S_n'(\lambda) P_n S_n(\lambda) u_n) = o_P(1)$, and $\sqrt{\frac{h_n}{n}}(u_n^{*'} S_n' P_n S_n u_n^* - u_n' S_n' P_n S_n u_n) = o_P(1)$, as $\frac{h_n}{n} = o(1)$.

For the derivative of the moment function, as $\frac{\partial}{\partial \lambda}(u_n^{*'} S_n'(\lambda) P_n S_n(\lambda) u_n^*) = -u_n^{*'} S_n'(\lambda) P_n^s W_n u_n^*$, $\frac{\partial g_n'(\lambda)}{\partial \lambda} = -D_n'(\lambda)$ where $D_n(\lambda) = (u_n^{*'} S_n'(\lambda) P_{1n}^s W_n u_n^*, \cdots, u_n^{*'} S_n'(\lambda) P_{mn}^s W_n u_n^*)'$. By expansion,

$$\begin{aligned} & u_n^{*'} S_n'(\lambda) P_n^s W_n u_n^* \\ &= u_n'(I_n - M_n) S_n'(\lambda) P_n^s W_n (I_n - M_n) u_n \\ &= u_n' S_n'(\lambda) P_n^s W_n u_n - u_n' M_n S_n'(\lambda) P_n^s W_n u_n - u_n' S_n'(\lambda) P_n^s W_n M_n u_n + u_n' M_n S_n'(\lambda) P_n^s W_n M_n u_n \\ &= u_n' S_n'(\lambda) P_n^s W_n u_n + o_P(1) \end{aligned}$$

uniformly in $\lambda \in \Lambda$ by Lemma A.10. In particular, $u_n^{*'} S_n'(\lambda) P_n^s W_n u_n^* = u_n' S_n'(\lambda) P_n^s W_n u_n + o_P(1)$ uniformly in $\lambda \in \Lambda$. The consistency and the asymptotic distribution of $\hat{\lambda}_n$ follow from Lemma A.8 and Proposition 2.5. Q.E.D.

Proof of Proposition 3.2: Because $\hat{\lambda}_n - \lambda_0 = o_P(1)$,

$$\begin{aligned} \frac{1}{n} X_n' \hat{S}_n' \hat{S}_n X_n &= \frac{1}{n} X_n' X_n - \hat{\lambda}_n \frac{1}{n} X_n' W_n^s X_n + \hat{\lambda}_n^2 \frac{1}{n} X_n' W_n' W_n X_n \\ &= \frac{1}{n} X_n' X_n - \lambda_0 \frac{1}{n} X_n' W_n^s X_n + \lambda_0^2 \frac{1}{n} X_n' W_n' W_n X_n + o_P(1) \\ &= \frac{1}{n} X_n' S_n' S_n X_n + o_P(1), \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\sqrt{n}} X_n' \hat{S}_n' \hat{S}_n u_n &= \frac{1}{\sqrt{n}} X_n' S_n' S_n u_n - (\hat{\lambda}_n - \lambda_0) \frac{1}{\sqrt{n}} X_n' W_n^s u_n + (\hat{\lambda}_n^2 - \lambda_0^2) \frac{1}{\sqrt{n}} X_n' W_n' W_n u_n \\ &= \frac{1}{\sqrt{n}} X_n' S_n' \epsilon_n + o_P(1) \xrightarrow{D} N(0, \sigma_0^2 (\lim_{n\to\infty} X_n' S_n' S_n X_n)^{-1}), \end{aligned}$$

by Lemma A.6. Hence,

$$\sqrt{n}(\hat{\beta}_{G,n} - \beta_0) = \left(\frac{1}{n} X_n' \hat{S}_n' \hat{S}_n X_n \right)^{-1} \frac{1}{\sqrt{n}} X_n' \hat{S}_n' \hat{S}_n u_n \xrightarrow{D} N\left(0, \sigma_0^2 \left(\lim_{n\to\infty} \frac{1}{n} X_n' S_n' S_n X_n \right)^{-1} \right).$$

Q.E.D.

Proof of Proposition 3.3: Denote $g_n^*(\lambda) = u_n^{*'} S_n'(\lambda)(\hat{G}_n)^d S_n(\lambda) u_n^*$. It is sufficient to show that this moment function and its derivative are close enough to those of $g_n(\lambda)$ where $g_n(\lambda) = u_n' S_n'(\lambda)(\hat{G}_n)^d S_n(\lambda) u_n$

so that Lemma A.8 is applicable. Specifically, it shall be shown that $g_n^*(\lambda) - g_n(\lambda) = O_p(1)$ and $\frac{\partial g_n^*(\lambda)}{\partial \lambda} - \frac{\partial g_n(\lambda)}{\partial \lambda} = O_P(1)$ uniformly in $\lambda \in \Lambda$. These properties are stronger than those sufficient conditions in Lemma A.8.

Because $u_n^* = (I_n - M_n)u_n$, $g_n^*(\lambda) = g_n(\lambda) + E_n(\lambda)$ where

$$E_n(\lambda) = -u_n' S_n'(\lambda)(\hat{G}_n^s)^d S_n(\lambda) M_n u_n + u_n' M_n S_n'(\lambda)(\hat{G}_n)^d S_n(\lambda) M_n u_n.$$

Substituting $u_n = S_n^{-1}\epsilon_n$ in the terms of $E_n(\lambda)$, Lemma A.12 is applicable and all the three terms of $E_n(\lambda)$ are of order $O_P(1)$ uniformly in $\lambda \in \Lambda$. The uniform order holds because λ is linear in $S_n(\lambda)$. Hence, $g_n^*(\lambda) = g_n(\lambda) + O_P(1)$ uniformly in $\lambda \in \Lambda$. Consequently, one has, in particular, that $\frac{h_n}{n} g_n^*(\lambda) = \frac{h_n}{n} g_n(\lambda) + o_P(1)$ and $\sqrt{\frac{h_n}{n}} g_n^*(\lambda_0) = \sqrt{\frac{h_n}{n}} g_n^*(\lambda_0) + o_P(1)$ because $\frac{h_n}{n} = o(1)$.

The first order derivative of $g_n^*(\lambda)$ is

$$\begin{aligned} \frac{\partial g_n^*(\lambda)}{\partial \lambda} &= -u_n^{*'} S_n'(\lambda)(\hat{G}_n^s)^d S_n(\lambda) u_n^* \\ &= -u_n'(I_n - M_n) W_n'(\hat{G}_n^s)^d S_n(\lambda)(I_n - M_n) u_n \\ &= \frac{\partial g_n(\lambda)}{\partial \lambda} + R_n(\lambda) \end{aligned}$$

where $R_n(\lambda) = u_n' M_n W_n'(\hat{G}_n^s)^d S_n(\lambda) u_n + u_n' W_n'(\hat{G}_n^s)^d S_n(\lambda) M_n u_n - u_n' M_n W_n'(\hat{G}_n^s)^d S_n(\lambda) M_n u_n$. By a similar argument, $R_n(\lambda) = O_P(1)$ uniformly in $\lambda \in \Lambda$ by Lemma A.12. This implies, in turn, that $\frac{h_n}{n}\frac{\partial g_n^*(\lambda)}{\partial \lambda} = \frac{h_n}{n}\frac{\partial g_n(\lambda)}{\partial \lambda} + o_P(1)$ uniformly in $\lambda \in \Lambda$.

The consistency of the estimator $\hat{\lambda}_n$ and its asymptotic distribution follow from Lemma A.8. Q.E.D.

Proof of Proposition 4.1: Consider each component of $g_n(\lambda)$. The lth component of $g_n(\lambda)$ is $Y_n' S_n'(\lambda) P_{ln} S_n(\lambda) Y_n = \epsilon_n' S_n^{'-1} S_n'(\lambda) P_{ln} S_n(\lambda) S_n^{-1} \epsilon_n$. By expansion,

$$S_n'(\lambda) P_{ln} S_n(\lambda) = P_{ln} - \sum_{j=1}^p \lambda_j W_{jn}' P_{ln} - P_{ln} \sum_{j=1}^p \lambda_j W_{jn} + \sum_{k=1}^p \sum_{j=1}^p \lambda_j \lambda_k W_{jn}' P_{ln} W_{kn}.$$

Lemmas A.1 and A.3 imply $\frac{h_n}{n}(\epsilon_n' H_n \epsilon_n - \sigma_0^2 tr(H_n)) = o_P(1)$ where $H_n = S_n^{'-1} P_{ln} S_n^{-1}$, $S_n^{'-1} P_{ln}^s W_{jn} S_n^{-1}$, and $S_n^{'-1} W_{jn}' P_{ln} W_{kn} S_n^{-1}$. It follows that $\frac{h_n}{n}(Y_n' S_n'(\lambda) P_{ln} S_n(\lambda) Y_n - E(Y_n' S_n'(\lambda) P_{ln} S_n(\lambda) Y_n)) = o_P(1)$ and $\frac{h_n}{n} a_n (g_n(\lambda) - E(g_n(\lambda))) = o_P(1)$ uniformly in $\lambda \in \Lambda$. With remaining arguments similar to those of the proof of Proposition 2.5, the consistency of $\hat{\lambda}_n$ follows from uniform convergence and the identification uniqueness condition.

The Taylor expansion of $\frac{\partial g_n'(\hat{\lambda})}{\partial \lambda} a_n' a_n g_n(\hat{\lambda}_n) = 0$ at λ_0 implies that

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = -\left[\frac{h_n}{n}\frac{\partial g_n'(\bar{\lambda}_n)}{\partial \lambda} a_n' a_n \frac{h_n}{n}\frac{\partial g_n(\bar{\lambda}_n)}{\partial \lambda'}\right]^{-1} \frac{h_n}{n}\frac{\partial g_n'(\bar{\lambda}_n)}{\partial \lambda'} a_n' a_n \sqrt{\frac{h_n}{n}}(\epsilon_n' P_{1n}\epsilon_n, \cdots, \epsilon_n' P_{mn}\epsilon_n).$$

As $E(\epsilon_n' P_{ln}^s W_{jn} S_n^{-1}\epsilon_n) = \sigma_0^2 tr(P_{ln}^s G_{jn})$, $\frac{h_n}{n} Y_n' S_n' P_{ln}^s W_{jn} Y_n = \frac{h_n}{n}\epsilon_n' P_{ln}^s W_{jn} S_n^{-1}\epsilon_n = \sigma_0^2 \frac{h_n}{n} tr(P_{ln}^s G_{jn}) + o_P(1)$. Hence, with uniform convergence of $\frac{h_n}{n}(\frac{\partial g_n(\lambda)}{\partial \lambda} - E(\frac{\partial g_n(\lambda)}{\partial \lambda}))$ to zero in probability uniformly in $\lambda \in \Lambda$, $\frac{\partial g_n(\bar{\lambda}_n)}{\partial \lambda} = -D_n + o_P(1)$, and

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = [\sigma_0^4 \frac{h_n}{n} D_n' a_n' a_n \frac{h_n}{n} D_n + o_P(1)]^{-1}(\sigma_0^2 \frac{h_n}{n} D_n' a_n' a_n + o_P(1))\sqrt{\frac{h_n}{n}}(\epsilon_n' P_{1n}\epsilon_n, \cdots, \epsilon_n' P_{mn}\epsilon_n).$$

The distribution of $\hat{\lambda}_n$ follows from Lemma A.4.

The optimal GMM weight for $a_n' a_n$ is $(\frac{h_n}{n}\Omega_n)^{-1}$ by the generalized Schwartz inequality. From the asymptotic variance matrix of the optimal GMM with P_ns, $(G_{jn} - Diag(G_{jn}))$ with $j = 1, \cdots, p$ are the best P_ns from $\mathcal{P}_{2n}$ by the generalized Schwartz inequality. When ϵ is normally distributed, $\Omega_n = V_n$. For the case that $\lim_{n\to\infty} h_n = \infty$, $\Omega_n = V_n + o_P(1)$. For both cases, the best selections from $\mathcal{P}_{1n}$ are $(G_{jn} - \frac{tr(G_{jn})}{n} I_n)$, $j = 1, \cdots, p$. Q.E.D.

Proof of Proposition 4.2 This proposition states that the feasible best estimators with $\hat{G}_{jn}$s will have the same asymptotic distributions as those best estimators with G_{jn}s.

Because the number of the best moment functions p is equal to the number of the unknown parameters λs, asymptotically, the minimization of $g_{jn}^{*'}(\lambda) V_{jn}^{*-1} g_{jn}^*(\lambda)$ is equivalent to solve the corresponding p moment equations $g_{jn}^*(\lambda) = 0$ for each $j = 1, 2$. The difference $\frac{h_n}{n}(g_{jn}^*(\lambda) - g_{jn}(\lambda))$ is a vector of dimension p. Its lth component is

$$\begin{aligned} \frac{h_n}{n} Y_n' S_n'(\lambda)[G_{ln}(\hat{\lambda}_n) - G_{ln}]^d S_n(\lambda) Y_n &= \frac{h_n}{n}\epsilon_n' S_n^{'-1}(I_n - \sum_{i=1}^p \lambda_i W_{in}')[G_{ln}(\hat{\lambda}_n) - G_{ln}]^d (I_n - \sum_{j=1}^p \lambda_j W_{jn}) S_n^{-1}\epsilon_n \\ &= T_{n1,l} - \sum_{i=1}^p \lambda_i T_{n2,li} + \sum_{i=1}^p \sum_{j=1}^p \lambda_i \lambda_j T_{n3,lij}, \end{aligned}$$

where $T_{n1,l} = \frac{h_n}{n}\epsilon_n' S_n^{'-1}[G_{ln}(\hat{\lambda}_n) - G_{ln}]^d S_n^{-1}\epsilon_n$, $T_{n2,li} = \frac{h_n}{n}\epsilon_n' G_{in}'[G_{ln}^s(\hat{\lambda}_n) - G_{ln}^s]^d S_n^{-1}\epsilon_n$, and $T_{n3,lij} = \frac{h_n}{n}\epsilon_n' G_{in}'[G_{ln}(\hat{\lambda}_n) - G_{ln}]^d G_{jn}\epsilon_n$. This decomposition slightly generalizes that in the proof of Proposition 2.10 for the first SAR process. The proof of this proposition can parallel to that of Proposition 2.10 for each component of the p moment equations and their derivatives. Q.E.D.

References

Amemiya, T. (1974), "The Nonlinear Two-Stage Least-Squares Estimator". *Journal of Econometrics* 2, 105-110.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht.

Anselin, L. (1990), "Some Robust Approaches to Testing and Estimation in Spatial Econometrics". *Regional Science and Urban Economics* 20, 141-163.

Anselin, L. and A.K. Bera (1998), "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics". In: *Handbook of Applied Economic Statistics*, A. Ullah and D.E.A. Giles, Eds., pp. 237-289, Marcel Dekker, NY.

Case, A.C. (1991), "Spatial Patterns in Household Demand". *Econometrica* 59, 953-965.

Case, A.C. (1992), "Neighborhood Influence and Technological Change". *Regional Science and Urban Economics* 22, 491-508.

Cliff, A. and J.K. Ord (1973), *Spatial Autocorrelation*, Pion, London.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators". *Econometrica* 50, 1029-1054.

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

Kelejian, H.H. and D.P. Robinson (1993), "A Suggested Method of Estimation for Spatial Interdependence Models with Autocorrelated Errors, and an Application to a County Expenditure Model". *Papers in Regional Science* 72, 297-312.

Kelejian, H.H. and I.R. Prucha (1997), "Estimation of Spatial Regression Models with Autoregressive Errors by Two-Stage Least Squares Procedures: A Serious Problem". *International Regional Science Review* 20, 1 and 2, pp. 103-111.

Kelejian, H.H. and I.R. Prucha (1998), "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbance". *Journal of Real Estate Finance and Economics* 17, 99-121.

Kelejian, H.H. and I.R. Prucha (1999a), "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model". *International Economic Review* 40, pp. 509-533.

Kelejian, H.H. and I.R. Prucha (1999b), "On the Asymptotic Distribution of the Moran I Test Statistic with Applications". Manuscript, Department of Economics, University of Maryland, forthcoming in *Journal* *of Econometrics*.

Land, K. and G. Deane (1992), "On the Large-Sample Estimation of Regression Models with Spatial-or Network-Effects Terms: A Two Stage Least Squares Approach". In P. Marsden (ed.), *Sociological Methodology*, pp. 221-248, Jossey Bass, San Francisco.

Lee, L.F. (1999a), "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances". Department of Economics, HKUST, Hong Kong.

Lee, L.F. (1999b), "Asymptotic Distributions of Maximum Likelihood Estimators for Spatial Autoregressive Models". Manuscript, revised 2001, Department of Economics, OSU, Columbus, OH.

Matyas, L. (1999), *Generalized Method of Moments Estimation*. Cambridge University Press, Cambridge, U.K.

Ord, J. (1975), "Estimation Methods for Models of Spatial Interaction", *Journal of the American Statistical Association* 70, 120-126.

Pace, R. and R. Barry (1997), "Sparse Spatial Autoregressions", *Statistics and Probability Letters* 33, 291-297.

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, London.

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, New York.

Whittle, P. (1954), "On Stationary Processes in the Plane". *Biometrika* 41, 434-449.

Zhang, F. (1999), *Matrix Theory*. Springer-Verlag, New York.