Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Econometrics 140 (2007) 155–189
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

The method of elimination and substitution in the GMM estimation of mixed regressive, spatial autoregressive models

Lung-fei Lee*

Department of Economics, Ohio State University, 410 Arps Hall, 1945 N. High St., Columbus, OH 43210-1172, USA

Received 1 September 2003; received in revised form 1 June 2005; accepted 5 September 2006
Available online 11 December 2006

Abstract

This paper proposes a computationally simple GMM for the estimation of mixed regressive spatial autoregressive models. The proposed method explores the advantage of the method of elimination and substitution in linear algebra. The modified GMM approach reduces the joint (nonlinear) estimation of a complete vector of parameters into estimation of separate components. For the mixed regressive spatial autoregressive model, the nonlinear estimation is reduced to the estimation of the (single) spatial effect parameter. We identify situations under which the resulting estimator can be efficient relative to the joint GMM estimator where all the parameters are jointly estimated.

JEL classification: C13; C21; R15

Keywords: Spatial econometrics; Spatial autoregression; The method of elimination and substitution; Sequential GMM estimation; Asymptotic efficiency

1. Introduction

This paper proposes a computationally simple generalized method of moments (GMM) of Hansen (1982) for the estimation of the mixed regressive spatial autoregressive (MRSAR) models. The introduced method is designed to reduce the jointly nonlinear

*Tel.: +1 614 292 5508; fax: +1 614 292 4192.
E-mail address: lflee@econ.ohio-state.edu.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.09.006

GMM estimation of a complete vector of parameters into nonlinear estimation of a single parameter. This modified GMM procedure can substantially reduce the computational burden. We shall identify situations under which the resulting modified GMM estimator may not lose (asymptotic) efficiency relative to the joint GMM estimator.

The GMM estimation of the MRSAR model has been considered in Kelejian and Prucha (1998) and Lee (2001b). In Kelejian and Prucha (1998), the method is a two-stage least squares (2SLS). Lee (2001b) extends the moment functions of a 2SLS with additional moment functions, which capture correlation across spatial units.

A sequential (two-step) estimation procedure in the GMM framework has been mentioned in Ogaki (1993) and Newey and McFadden (1994). Ogaki (1993) and Newey and McFadden (1994) describe a system of recursive moment functions where the first set of moments contains a proper subset of parameters in the model and the remaining moments may contain all the parameters. A sequential GMM estimation is motivated by computational simplicity as a two-step estimation procedure. The sequential approach uses the first set of moment functions for the estimation of the relevant subset of parameters, and, recursively, it estimates the remaining parameters by using the second set of moment functions. Ogaki (1993) notes that the asymptotic distribution of estimated parameters in the second step GMM estimation will, in general, depend on the asymptotic distribution of the first step estimates. He derives the optimal distance matrix in the second step GMM estimation. Newey and McFadden (1994) have discussed similar issues. A sequential GMM estimator may be inefficient relative to the joint GMM estimator derived by using the complete set of moment functions with an optimal distance matrix. There are other estimation approaches which may involve two-step estimation. For example, for the estimation of a MRSAR model with spatial correlated disturbances, Kelejian and Prucha (1998) discuss a feasible generalized 2SLS approach. This approach will first estimate the MRSAR equation with a 2SLS. The spatial process of the disturbances is then estimated with the estimated residuals. The final estimate of the coefficients of the MRSAR equation comes from a feasible generalized 2SLS estimation. The system of moments in this case is not recursive in parameters. This two-step approach concerns feasible weighting issue in the generalized 2SLS approach. The two-step method in this paper is not related to weighting issues as we are considering the estimation of a MRSAR model with i.i.d. disturbances.

The systems of moment functions for the estimation of the MRSAR model in Lee (2001b), which extends the 2SLS moment functions in Kelejian and Prucha (1998) with additional moments, are not recursive in parameters. The additional moments are quadratic functions in parameters. They are designed to improve the possible efficiency of estimators. These nonlinear moments render the estimation of the complete set of coefficients of the MRSAR equation into a nonlinear estimation framework. The MRSAR equation is linear with respect to exogenous regressors but it has a nonlinear feature in its reduced form due to the spatial interactions. The coefficients of the exogenous regressors seem a nuisance in the nonlinear estimation because once the spatial interactions parameter is given, the regression coefficients can simply be estimated by the method of ordinary least squares (OLS). Therefore, one may have the desire to reduce the nonlinear estimation into the estimation of the (single) spatial effect parameter.

In this paper, we introduce the method of elimination and substitution within the GMM framework. This method eliminates the coefficients of exogenous regressors of the MRSAR from the original GMM functions. After substitution, the nonlinear estimation will focus on the spatial effect parameter in the remaining moment equations. Systematically, the modified moment equations can be cast in a sequential GMM estimation framework and the estimation becomes a two-step method. This approach is computationally simpler than the jointly nonlinear GMM estimation based on the full set of moment functions. The remaining issue concerns the possible loss of efficiency of this sequential estimation. We show that the resulting GMM estimator, or equivalently, the corresponding sequential (two-step) estimator, can be asymptotically as efficient as the joint GMM estimator under certain circumstances. They include the case where the disturbances have zero third order moment, in particular, normally distributed disturbances. They include also the case with large group interactions. The asymptotic efficiency will always be preserved under a certain class of moment functions.

This paper is organized as follows. In Section 2, we review the 2SLS, the joint GMM, and the sequential GMM approaches for estimating the MRSAR model. The method of elimination and substitution within the GMM framework is introduced. Section 3 lists some basic regularity conditions and discuss identification of the model. The consistency and asymptotic distribution of the modified GMM estimator in the presence of valid regressors are derived in Section 4. Relative efficiency of the modified GMM estimator is studied. Section 5 studies the model with large group interactions under a situation of nearly multicollinearity. Section 6 studies estimation issues when the MRSAR model possesses a feature of exact multicollinearity. Conclusions are drawn in Section 7. For easy reference, frequently used notations in the text or in the proofs are collected in Appendix A. Some useful lemmas for the proofs are in Appendix B. Proofs of the main results are collected in Appendix C. Appendix D provides a numerical identity of the modified 2SLS estimator with the conventional 2SLS estimator for the MRSAR model.

2. The MRSAR model, 2SLS, joint GMM, sequential GMM, and the method of elimination and substitution

2.1. The MRSAR model

The simplest MRSAR model is specified as

$$Y_n = \lambda_0 W_n Y_n + X_n\beta_0 + \mathscr{E}_n, \tag{2.1}$$

where $\mathscr{E}'_n = (\varepsilon_1, \ldots, \varepsilon_n)$, ε_i is i.i.d. $(0, \sigma_0^2)$, W_n is a specified $n \times n$ spatial weights matrix of constants, and X_n is a $n \times k$ matrix of exogenous variables with full column rank. The MRSAR model is an equilibrium model. The Y_n can be determined from the system as

$$Y_n = (I_n - \lambda_0 W_n)^{-1}(X_n\beta_0 + \mathscr{E}_n). \tag{2.2}$$

This model has been introduced in Cliff and Ord (1973). See, also, Anselin (1988) and Cressie (1993).

2.2. The 2SLS estimation

The MRSAR equation in (2.1) may be estimated by the 2SLS method if there are valid instrumental variables (IV) available from the model. When β_0 is zero, model (2.1) becomes a pure spatial autoregressive process, i.e., $Y_n = \lambda_0 W_n Y_n + \mathscr{E}_n$, and no valid IV constructed from X_n is available. When β_0 is not zero, functions of X_n (and W_n) can be valid IV variables. Kelejian and Prucha (1998) suggest X_n, W_nX_n and/or $W_n^2X_n$ to form the IV matrix. From (2.2), the ideal IV for W_nY_n shall be $W_n(I_n-\lambda_0W_n)^{-1}X_n\beta_0$. The W_nX_n and $W_n^2X_n$ provide approximations to this ideal IV.

Let Q_n be an $n\times r$ IV matrix with full column rank r, where $r\geqslant k+1$. The 2SLS estimator of $(\lambda_0,\beta_0')'$ is

$$\begin{pmatrix}\hat{\lambda}_n\\ \hat{\beta}_n\end{pmatrix}=\left[\begin{pmatrix}(W_nY_n)'\\ X_n'\end{pmatrix}Q_{n(\mathrm{p})}(W_nY_n,X_n)\right]^{-1}\begin{pmatrix}(W_nY_n)'\\ X_n'\end{pmatrix}Q_{n(\mathrm{p})}Y_n, \tag{2.3}$$

where, for any matrix A, $A_{(\mathrm{p})}=A(A'A)^{-1}A'$ denotes the orthogonal projector to the column space of A. The 2SLS estimator in (2.3) is a joint estimator in the sense that λ and β are jointly estimated in a single step.

2.3. A GMM approach

The 2SLS estimator (2.3) may not be efficient relative to other estimators such as the maximum likelihood (ML) estimator with some exceptions.[1] Lee (2001b) has introduced a GMM estimator derived by combining the moment functions of the 2SLS and other moment functions capturing spatial correlation. That GMM estimator can be efficient relative to the 2SLS estimator. For the MRSAR model (2.1), for any possible value $\theta=(\lambda,\beta')'$, let

$$\varepsilon_n(\theta)=(I_n-\lambda W_n)Y_n-X_n\beta.$$

The empirical moment functions for the GMM estimation in Lee (2001b) consist of $\varepsilon_n'(\theta)P_n\varepsilon_n(\theta)$ and $Q_n'\varepsilon_n(\theta)$ where P_n is a $n\times n$ constant matrix with a zero trace. As W_n is a constant matrix with zero diagonal, W_n can be used as a P_n. Other matrices generated from W_n such as $W_n^2-\frac{\mathrm{tr}(W_n^2)}{n}I_n$, etc, may also qualify.[2] With several such matrices $P_{1n},\ldots,P_{mn}$, the vector of moment functions for the GMM estimation can be

$$f_n^*(\theta)=(\varepsilon_n'(\theta)P_{1n}\varepsilon_n(\theta),\ldots,\varepsilon_n'(\theta)P_{mn}\varepsilon_n(\theta),\varepsilon_n'(\theta)Q_n)'. \tag{2.4}$$

The corresponding optimally weighted GMM estimator can be asymptotically efficient relative to the 2SLS estimator (2.3) because of the additional quadratic moment functions.[3] For cases where there are no valid IVs, $\varepsilon_n'(\theta)P_{jn}\varepsilon_n(\theta), j=1,\ldots,m$, provide the identification and estimating functions for λ (Lee, 2001a).

From the computational point of view, the GMM estimation is relatively more demanding than the 2SLS approach. While $Q_n'\varepsilon_n(\theta)$ is linear in parameters, $\varepsilon_n'(\theta)P_{jn}\varepsilon_n(\theta)$'s are quadratic functions of θ. With these moment functions, the GMM objective function for minimization will be a quartic function in θ, and the GMM estimator needs to be derived via a minimization routine. The minimization will involve the search for the estimator in a $(k+1)$-dimensional parameter space. For ML estimation, β can be easily

[1]For some models with large group interactions, the 2SLS estimator can be as efficient as the ML estimator, see Lee (2002).

[2]In a subsequent section, it shall be shown under some situations that the best selected P_n is $W_n(I_n-\lambda_0W_n)^{-1}-\frac{1}{n}\mathrm{tr}[W_n(I_n-\lambda_0W_n)^{-1}]I_n$. The matrices generated from W_n provide valuable approximations to the best one.

[3]For some models with large group interactions, these additional moment functions may not increase asymptotic efficiency over the 2SLS estimator. For other cases, the issue of the appropriate number of matrices P_{jn} may become less a problem because of the existence of the best one in the preceding footnote.

concentrated out and the optimization is on the single λ. On the contrary, β in the GMM estimation with the moment functions (2.4) cannot be easily concentrated out. For a given λ, the MLE of β is linear and is simply the OLS estimator of $(I_n - \lambda W_n)Y_n$ on X_n. On the other hand, the GMM objective function is a quartic function in β. Given λ, the GMM estimation of β is still nonlinear. This motivates the search for a computationally simpler GMM estimation for the model.

2.4. A method of elimination and substitution within the GMM framework

We suggest the estimation of λ and β in steps. First, β shall be evaluated given any value of λ via a subset of the linear moment equations.[4] After substitution into the remaining equations, λ will be estimated by the GMM method. Finally, β can be estimated once the estimate of λ is available.

Consider $Q_n = (Q_{n1}, X_n)$ where Q_{n1} is a matrix of possible instrumental variables excluding X_n. Given a value of λ, an estimator $\hat{\beta}_n(\lambda)$ can be derived from the following linear moment equations:

$$X_n'(Y_n - W_n Y_n \lambda - X_n \hat{\beta}_n(\lambda)) = 0, \tag{2.5}$$

which is

$$\hat{\beta}_n(\lambda) = (X_n' X_n)^{-1} X_n' (I_n - \lambda W_n) Y_n. \tag{2.6}$$

For a given λ, let the least squares residual vector be

$$\varepsilon_{x,n}(\lambda) = (I_n - \lambda W_n)Y_n - X_n \hat{\beta}_n(\lambda) = M_n (I_n - \lambda W_n) Y_n, \tag{2.7}$$

where $M_n = I_n - X_n(X_n' X_n)^{-1} X_n'$. By substituting $\hat{\beta}_n(\lambda)$ for β in (2.4), these moments become, respectively, $\varepsilon_{x,n}(\lambda)' P_{jn} \varepsilon_{x,n}(\lambda)$, $j = 1, \ldots, m$, and $Q_{n1}' \varepsilon_{x,n}(\lambda)$. For the GMM estimation of λ, the empirical moment functions are[5]

$$g_n(\lambda) = (\varepsilon_{x,n}'(\lambda) P_{1n} \varepsilon_{x,n}(\lambda), \ldots, \varepsilon_{x,n}'(\lambda) P_{mn} \varepsilon_{x,n}(\lambda), \varepsilon_{x,n}'(\lambda) Q_{n1})'. \tag{2.8}$$

At λ_0, denote $\varepsilon_{x,n} = \varepsilon_{x,n}(\lambda_0)$. Because $\varepsilon_{x,n} = M_n \mathscr{E}_n$, $E(\varepsilon_{x,n}) = 0$. These moment functions resemble those in (2.4) with $\varepsilon_{x,n}(\lambda)$ and Q_{n1} replacing, respectively, $\varepsilon_n(\theta)$ and Q_n.[6] The $g_n(\lambda)$ is still a quartic function of λ. But a GMM minimization will involve only a one-dimensional search over the parameter space of λ.

The modified moment functions after elimination and substitution can be cast into the sequential GMM framework of Ogaki (1993). In the general sequential GMM framework, $f(z_i, \theta) = (f_1'(z_i, \theta_1), f_2'(z_i, \theta_1, \theta_2))'$ consists of recursive sets of moments. The first set of moment functions $f_1(z_i, \theta_1)$ depends on the parameter component θ_1, where z_i is the ith vector of sample observations. The θ_1 is a subvector of θ, where $\theta = (\theta_1', \theta_2')'$. The second set of moment functions $f_2(z_i, \theta_1, \theta_2)$ depends on both θ_1 and θ_2. The sequential GMM estimation approach separates the estimation of θ_1 and θ_2 in two steps. It may be

[4] If one intends to eliminate β from the whole set of linear and nonlinear moment equations for a given λ, this would concentrate out the β. However, the latter would involve nonlinear estimation of β under the assumption that each of P_{jn} with $j = 1, \ldots, m$ has a zero trace.

[5] The linear moments involve only Q_{n1} instead of Q_n because $X_n' \varepsilon_{x,n}(\lambda)$ is identically zero for all λ by construction.

[6] We note that there is a minor difference of these recursive moment functions from those in (2.4) in that $E(\varepsilon_{x,n}' P_{jn} \varepsilon_{x,n}) = \sigma_0^2 \operatorname{tr}(P_{jn} M_n)$, $j = 1, \ldots, m$, are not necessarily zero. However, one may show that they can be close to zero when divided by the large sample size n under the assumption that P_{jn}'s have a zero trace.

computationally convenient to estimate θ_1 from $f_1(z_i, \theta_1)$ first, and then estimate θ_2 from $f_2(z_i, \hat{\theta}_1, \theta_2)$, where $\hat{\theta}_1$ is the first step estimate of θ_1, in a secondstep.[7] For the modified GMM estimation based on (2.6) and (2.8), the complete set of recursive moments is

$$f_n(\theta) = (\varepsilon'_{x,n}(\lambda)P_{1n}\varepsilon_{x,n}(\lambda), \ldots, \varepsilon'_{x,n}(\lambda)P_{mn}\varepsilon_{x,n}(\lambda), \varepsilon'_{x,n}(\lambda)Q_{n1}, \varepsilon'_n(\theta)X_n). \tag{2.9}$$

The estimation with the moment functions in (2.8) for λ corresponds to the first step estimation with θ_1 being λ. The least squares estimation in (2.6) corresponds to the second step GMM estimation with the moment functions $X'_n\varepsilon_n(\theta)$ and θ_2 being β.

Because a sequential GMM estimation may lose efficiency relative to the joint GMM estimation, in the subsequent sections, we shall investigate asymptotic properties of the modified GMM estimator and compare its efficiency with that of the joint GMM estimator based on the original moments (2.4) as well as that based on the complete recursive moments (2.9). We shall identify circumstances under which the sequential estimation may be efficient as the joint GMM estimation of the MRSAR model.

3. Some basic regularity conditions and model identification

3.1. Some basic regularity conditions

In order to justify asymptotic properties of the modified GMM estimator, it is essential to have restrictions on the sequence of spatial weights matrices $\{W_n\}$. In the empirical literature, there are generally two different types of spatial weights matrices. For geographical problems, W_n may be a sparse matrix as neighboring units for each spatial unit are defined by only a few adjacent ones. For social interactions problems in group setting, each unit may be influenced by all members in the group but each member of a group may have only small influence on other members. In many cases, the group size can be large. An example of the latter is the model in Case (1991). To allow both scenarios, the following assumptions are the basic regularity conditions for the model (2.1).

Assumption 1. The disturbances ε_i's of (2.1) are i.i.d. with zero mean, variance σ_0^2 and its moments of higher than the fourth order exists.

Assumption 2. The elements of exogenous variables in the $n \times k$ matrix X_n are uniformly bounded constants, X_n has the full column rank k, and $\lim_{n\to\infty} \frac{1}{n} X'_n X_n$ exists and is nonsingular.

Assumption 3. The spatial weights matrices $\{W_n\}$ and $\{(I_n - \lambda W_n)^{-1}\}$ at $\lambda = \lambda_0$ are uniformly bounded in absolute value in both row and column sums.

Assumption 4. The elements of $W_n = (w_{n,ij})$ are of order $O(\frac{1}{h_n})$ uniformly in i and j, where $\{h_n\}$ can be a bounded or a divergent nonnegative constant sequence.

Assumption 5. If $\{h_n\}$ is a divergent sequence, $\lim_{n\to\infty} \frac{h_n}{n} = 0$.

[7]The main issues discussed in Ogaki (1993) and Newey and McFadden (1994) concern the asymptotic distribution of the second step estimator of θ_2. Newey and McFadden (1994) have focused their attention on an exactly identified moment system. Ogaki (1993) describes the optimum selection of the distance matrix in the second step GMM estimation. The second step estimator of θ_2 will, in general, be affected by the asymptotic distribution of the first step estimator of θ_1 unless $E(\frac{\partial f_2(z,\theta_0)}{\partial \theta_1}) = 0$.

Assumption 6. The parameter space Λ of λ_0 is a compact interval of the real line with λ_0 in its interior.

These six assumptions are the basic structures of the model. The variances of quadratic forms $\varepsilon_n' P_{jn} \varepsilon_n$ exist as the fourth moment of ε_n exists. The existence of a moment higher than the fourth order is needed for the application of central limit theorems for quadratic forms (Kelejian and Prucha, 2001). Assumption 2 is a convenient assumption, which can be replaced by proper finite moment conditions if x is stochastic without a bounded range.

The notion of uniform boundedness in absolute value in row and column sums of a sequence of matrices in Assumption 3 is important. For any sequence of $n \times n$ matrices $\{A_n\}$, where $A_n = (a_{n,ij})$, A_n is uniformly bounded in absolute value in row sums (respectively, column sums) if there exists a finite constant c such that $\max_{1 \leqslant i \leqslant n} \sum_{j=1}^n |a_{n,ij}| \leqslant c$ (respectively, $\max_{1 \leqslant j \leqslant n} \sum_{i=1}^n |a_{n,ij}| \leqslant c$) for all n. These conditions can be expressed as boundedness in matrix norms for $\{A_n\}$. The $\|A_n\|_1 = \max_{1 \leqslant j \leqslant n} \sum_{i=1}^n |a_{n,ij}|$ is known as the maximum column sum matrix norm of A_n, and $\|A_n\|_\infty = \max_{1 \leqslant i \leqslant n} \sum_{j=1}^n |a_{n,ij}|$ is its maximum row sum matrix norm (Horn and Johnson, 1985).[8] As a matrix norm satisfies the submultiplicative property, when two sequences $\{A_n\}$ and $\{B_n\}$ are uniformly bounded in absolute value in row (column) sums, its product sequence $\{A_n B_n\}$ will also be uniformly bounded in absolute value in row (column) sums. The variance of Y_n from (2.2) is $\sigma_0^2 (I_n - \lambda_0 W_n)^{-1} (I_n - \lambda_0 W_n')^{-1}$. Along with Assumptions 1 and 2, Assumption 3 guarantees, for example, that the elements of Y_n have finite variances and the sequence of variances is bounded as n goes to infinity. These uniform boundedness conditions originated in Kelejian and Prucha (1998, 1999, 2001). We note that the uniform boundedness condition for $\{(I_n - \lambda W_n)^{-1}\}$ is imposed at $\lambda = \lambda_0$. It can be shown, in general, under this condition the uniform boundedness property will also hold uniformly w.r.t. λ at a small neighborhood of λ_0 (Lee, 2004).[9]

Assumption 4 includes both the conventional spatial scenario of few neighbors and the social scenario with a large number of small interactions in Case (1991).[10] In Case's model, 'neighbors' refer to farmers who live in the same district. Suppose that there are R districts and there are m farmers in each district (for simplicity). The sample size is $n = mR$. Case assumed that in a district, each neighbor of a farmer is given equal weight. In that case, $W_n = I_R \otimes B_m$, where $B_m = \frac{1}{(m-1)}(l_m l_m' - I_m)$, $\otimes$ is the Kronecker product, and l_m is a m-dimensional column vector of ones. In this example, $h_n = (m-1)$. If sample size n increases because m is increasing, then $\{h_n\}$ will be a divergent sequence. When $\lambda < 1$, it is easy to see that the uniform boundedness conditions in Assumption 3 are satisfied for this large interactions scenario. As $\frac{h_n}{n} = (\frac{m-1}{m}) \cdot \frac{1}{R} = O(\frac{1}{R})$. If sample size n increases by increasing both R and m, then h_n goes to infinity and $\frac{h_n}{n}$ goes to zero as n tends to infinity. Assumption 5 rules out extreme cases that h_n diverges to infinity at the rate n. If h_n is divergent to infinity at the rate n, one can give an example that the GMM estimator is inconsistent. The same phenomenon is observed for the ML estimation of the model (Lee, 2004).

[8]The norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are related to L_1 and L_∞ for integrable functions. This is so because they are the matrix norms induced by the corresponding vector norms for finite dimensional vectors.

[9]Uniform boundedness of $\{(I_n - \lambda W_n)^{-1}\}$ uniform w.r.t. λ in its whole parameter space Λ would be a stronger requirement. For the GMM approach, the stronger requirement is not needed.

[10]We note that the uniform boundedness condition for W_n in Assumption 3 rules out the possibility that $\{h_n\}$ goes to zero. This is so, because the uniform boundedness condition implies that all elements of $W_n = (w_{n,ij})$ are uniformly bounded for all i, j.

For general nonlinear extremum estimation, a parameter space is generally assumed to be a compact set (Amemiya, 1985). This is so for our model by Assumption 6. For our case, the nonlinearity of λ appears in a quartic form of the objective minimization function.

As the IV matrices Q_{n1} and P_{jn}'s are usually generated from X_n and W_n, they shall have similar properties as those of X_n and W_n in Assumptions 2–5. In particular, it is reasonable to assume that they shall possess the properties in the following assumption.

Assumption 7. The elements of Q_n are uniformly bounded in absolute value. The constant matrices P_{jn}'s with $\operatorname{tr}(P_{jn}) = 0$ are uniformly bounded in absolute value in both row and column sums. The elements of P_{jn}'s, where $P_{jn} = (p_{n,ij})$, are of order $O(\frac{1}{h_n})$ uniformly in i and j.

Additional regularity conditions shall be subsequently specified.

3.2. Identification

Consider the identification of λ via $g_n(\lambda)$ in (2.8). From (2.7), in terms of its relation to $\mathscr{E}_n$,

$$\begin{aligned}\varepsilon_{x,n}(\lambda) &= M_n(I_n - \lambda W_n)(I_n - \lambda_0 W_n)^{-1}(X_n\beta_0 + \mathscr{E}_n)\\ &= (\lambda_0 - \lambda)M_n W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0 + M_n(I_n - \lambda W_n)(I_n - \lambda_0 W_n)^{-1}\mathscr{E}_n, \end{aligned} \quad (3.1)$$

where the first term on the right-hand side of the last equality follows because $(I_n - \lambda W_n)(I_n - \lambda_0 W_n)^{-1} = I_n + (\lambda_0 - \lambda)W_n(I_n - \lambda_0 W_n)^{-1}$ and $M_n X_n = 0$. Therefore,

$$E(Q'_{n1}\varepsilon_{x,n}(\lambda)) = (\lambda_0 - \lambda)Q'_{n1}M_n W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0 \quad (3.2)$$

and

$$\begin{aligned}E(\varepsilon'_{x,n}(\lambda)P_{jn}\varepsilon_{x,n}(\lambda)) = (\lambda_0 - \lambda)^2(W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0)'M_n P_{jn} M_n W_n(I_n - \lambda_0 W_n)^{-1}\\ \times X_n\beta_0 + \sigma_0^2 \operatorname{tr}[(I_n - \lambda_0 W_n)'^{-1}(I_n - \lambda W_n)'M_n P_{jn}\\ \times M_n(I_n - \lambda W_n)(I_n - \lambda_0 W_n)^{-1}]. \end{aligned} \quad (3.3)$$

From (3.2), $E(Q'_{n1}\varepsilon_{x,n}(\lambda)) = 0$ has a unique solution at $\lambda = \lambda_0$ if $Q'_{n1}M_n W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0 \neq 0$.

If $W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0$ depends linearly on X_n for large n, $M_n W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0 = 0$ and the identification of λ_0 will rely on the moment equations in (3.3). An obvious example of multicollinearity is that all the regressors in X_n may be irrelevant, i.e., $\beta_0 = 0$. Another relevant example is that W_n is row-normalized and $X_n = (l_n, X_{2n})$, where l_n is a vector of ones, but X_{2n} is a subvector of spatially varying regressors with zero coefficients. In this situation, $X_n\beta_0 = l_n\beta_{01}$, where β_{01} is the intercept and $W_n(I_n - \lambda_0 W_n)^{-1}\ X_n\beta_0 = \frac{\beta_{01}}{1-\lambda_0}l_n$. For any square matrix A, denote $A^{s} = A + A'$, i.e., the sum of A and its transpose. A^{s} is symmetric. When linear dependence occurs, (3.3) becomes

$$\begin{aligned}&E(\varepsilon'_{x,n}(\lambda)P_{jn}\varepsilon_{x,n}(\lambda))\\ &\quad = \sigma_0^2 \operatorname{tr}[(I_n - \lambda_0 W_n)'^{-1}(I_n - \lambda W_n)'M_n P_{jn} M_n(I_n - \lambda W_n)(I_n - \lambda_0 W_n)^{-1}]\\ &\quad = \sigma_0^2\{(\lambda_0 - \lambda)\operatorname{tr}(P^{s}_{jn}W_n(I_n - \lambda_0 W_n)^{-1})\\ &\qquad + (\lambda_0 - \lambda)^2 \operatorname{tr}((I_n - \lambda_0 W'_n)^{-1} \times W'_n P_{jn} W_n(I_n - \lambda_0 W_n)^{-1})\} + O(1),\end{aligned}$$

where the simplification comes from Lemmas B.4 and B.5 in Appendix B, and $\mathrm{tr}(P_{jn}) = 0$. The identification of λ_0 shall be based on $E(\varepsilon_n'(\lambda)P_{jn}\varepsilon_n(\lambda)) = 0$, $j = 1, \ldots, m$, with $m \geqslant 2$, similar to those for the identification of the pure SAR process in Lee (2001a).

The identification of λ_0 (and hence β_0) can be summarized:

(i) either $\lim_{n\to\infty} \frac{1}{n} Q_n'[W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0, X_n]$ has the full rank $(k+1)$, or,
(ii) $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ has the full rank k, $\lim_{n\to\infty} \frac{h_n}{n} \mathrm{tr}(P_{jn} W_n(I_n - \lambda_0 W_n)^{-1}) \neq 0$ for some j, and

$$\lim_{n\to\infty} \frac{h_n}{n}[\mathrm{tr}(P_{1n}^{\mathrm{s}} W_n(I_n - \lambda_0 W_n)^{-1}), \ldots, \mathrm{tr}(P_{mn}^{\mathrm{s}} W_n(I_n - \lambda_0 W_n)^{-1})]'$$

is linearly independent of

$$\lim_{n\to\infty} \frac{h_n}{n}[\mathrm{tr}((I_n - \lambda_0 W_n')^{-1} W_n' P_{1n} W_n(I_n - \lambda_0 W_n)^{-1}), \ldots, \mathrm{tr}((I_n - \lambda_0 W_n')^{-1} \times W_n' P_{mn} W_n(I_n - \lambda_0 W_n)^{-1})]'.$$

The identification condition (i) explores the existence of relevant regressors and valid instruments. By substituting the reduced form of W_nY_n into (2.1), the reduced form equation (2.2) can be expressed as

$$Y_n = \lambda_0[W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0] + X_n\beta_0 + (I_n - \lambda_0 W_n)^{-1}\mathscr{E}_n. \tag{3.4}$$

This condition (i) will fail if $W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0$ and X_n are multicollinear (in the limit). For example, without the presence of valid regressors X_n, $\beta_0 = 0$ will imply the failure of this condition. The condition (ii) can remedy this situation as it explores the correlation of Y_n via the correlation of the reduced form disturbances $(I_n - \lambda_0 W_n)^{-1}\mathscr{E}_n$. While condition (i) is invariant with the size of group interactions, namely h_n, condition (ii) involves explicitly h_n. This is so because $\mathrm{tr}(P_{jn}^{\mathrm{s}} W_n(I_n - \lambda_0 W_n)^{-1}) = \mathrm{O}(\frac{n}{h_n})$ and $\mathrm{tr}((I_n - \lambda_0 W_n')^{-1} W_n' P_{jn} W_n(I_n - \lambda_0 W_n)^{-1}) = \mathrm{O}(\frac{n}{h_n})$ as shown in Appendix B. For subsequent analysis of asymptotic properties of the modified GMM estimator, it is desirable to separate the situations of condition (i) and condition (ii).

4. Consistency and asymptotic distribution of the modified GMM estimator in the presence of valid regressors

In this section, we consider the situation (i). When $W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0$ and X_n are linearly independent, (i) can be valid as long as the IV matrix Q_n is properly chosen. Because $Q_n = (Q_{n1}, X_n)$, condition (i) is equivalent to that $\lim_{n\to\infty} \frac{1}{n} Q_{n1}' M_n W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0 \neq 0$ and $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ has the full rank k.[11]

Assumption 8. $\lim_{n\to\infty} \frac{1}{n} Q_{n1}' M_n W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0 \neq 0$.

Because $E(W_nY_n) = W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0$, Assumption 8 means that the IV's in Q_{n1} shall be chosen such that they are correlated with $E(W_nY_n)$ after the influence of X_n has

[11] This follows because $\begin{pmatrix} A & B \\ D' & C \end{pmatrix} = \begin{pmatrix} A - BC^{-1}D' & B \\ 0 & C \end{pmatrix}\begin{pmatrix} I_1 & 0 \\ C^{-1}D' & I_2 \end{pmatrix}$ where C is nonsingular.

been taken out. If $M_nE(W_nY_n)$ were zero, this assumption would fail. Cases where Assumption 8 is violated will be considered in a subsequent section.

Suppose that Q_n is a $n \times r$ matrix. The matrix Q_{n1} will then be a $n \times (r-k)$ matrix. Let F_n be a constant (stochastic) $s \times (m+r-k)$ matrix of full row rank s, where $s \geqslant 1$, which converges (in probability) to a constant matrix F_0 with full row rank s. The modified GMM estimator $\hat{\lambda}_n$ is derived from $\min_\lambda g_n'(\lambda)F_n'F_ng_n(\lambda)$ with $g_n(\lambda)$ in (2.8).

4.1. Asymptotic distribution of the modified GMM estimator

Consistency of the modified GMM estimator $\hat{\lambda}_n$ can be established in Theorem 1 below. For the asymptotic distribution, the first order condition of $\hat{\lambda}_n$ is $\frac{\partial g_n'(\hat{\lambda}_n)}{\partial \lambda}F_n'F_ng_n(\hat{\lambda}_n) = 0$. By the mean value theorem of $g_n(\hat{\lambda}_n)$ at λ_0,

$$\hat{\lambda}_n - \lambda_0 = -\left(\frac{\partial g_n'(\hat{\lambda}_n)}{\partial \lambda}F_n'F_n\frac{\partial g_n(\bar{\lambda}_n)}{\partial \lambda'}\right)^{-1}\frac{\partial g_n'(\hat{\lambda}_n)}{\partial \lambda}F_n'F_ng_n(\lambda_0), \tag{4.1}$$

where $\bar{\lambda}_n$ lies between λ_0 and $\hat{\lambda}_n$ and

$$\frac{\partial g_n(\lambda)}{\partial \lambda} = -(P_{1n}^s\varepsilon_{x,n}(\lambda), \ldots, P_{mn}^s\varepsilon_{x,n}(\lambda), Q_{n1})'M_nW_n(I_n - \lambda_0W_n)^{-1}(X_n\beta_0 + \mathscr{E}_n). \tag{4.2}$$

For any sequence of constant vectors $\{b_n\}$ with all its elements being bounded, $\frac{1}{n}b_n'\mathscr{E}_n = o_P(1)$ and, for any sequence of constant matrices $\{A_n\}$ uniformly bounded in absolute value in either row or column sums, $\frac{1}{n}\mathscr{E}_n'A_n\mathscr{E}_n = \frac{\sigma_0^2}{n}\operatorname{tr}(A_n) + o_P(1)$ by Lemma B.8. It follows that

$$\frac{1}{n}\frac{\partial g_n(\lambda_0)}{\partial \lambda} = -\frac{1}{n}D_n + o_P(1), \tag{4.3}$$

where

$$D_n = [\sigma_0^2C_{mn}, (W_n(I_n - \lambda_0W_n)^{-1}X_n\beta_0)'M_nQ_{n1}]'$$

and $C_{mn} = [\operatorname{tr}(P_{1n}^sW_n(I_n - \lambda_0W_n)^{-1}), \ldots, \operatorname{tr}(P_{mn}^sW_n(I_n - \lambda_0W_n)^{-1})]$.

The asymptotic distribution of $\hat{\lambda}_n$ can be derived from (4.1) and (4.3). The proof is in Appendix C.

Theorem 1. *Under the regularity Assumptions* (1–8) *and that, for any* $\lambda \neq \lambda_0$, $\lim_{n\to\infty}\frac{1}{n}E(g_n(\lambda))$, *where* $g_n(\lambda)$ *is based on* (2.8), *does not lie in the orthogonal space of the rows of* F_0, *the modified GMM estimator* $\hat{\lambda}_n$ *is consistent and*

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) = \left(\frac{1}{n}D_n'F_n'F_nD_n\right)^{-1}D_n'F_n'F_n\frac{1}{\sqrt{n}}g_n(\lambda_0) + o_P(1) \xrightarrow{d} N(0, \Sigma_\lambda), \tag{4.4}$$

where

$$\Sigma_\lambda = \lim_{n\to\infty}\left(\frac{1}{n}D_n'F_n'F_nD_n\right)^{-1}D_n'F_n'F_n\frac{1}{n}\,Var(g_n(\lambda_0))F_n'F_nD_n\left(\frac{1}{n}D_n'F_n'F_nD_n\right)^{-1}, \tag{4.5}$$

which is assumed to exit.

The condition in Theorem 1 that, for any $\lambda \neq \lambda_0$, $\lim_{n\to\infty}\frac{1}{n}E(g_n(\lambda))$ does not lie in the orthogonal space of the rows of F_0, shall hold under the identification condition in

Assumption 8 as long as $\{F_n\}$ is chosen not to wipe out all the information in $g_n(\lambda)$ (Ruud, 2000). The asymptotic distribution of $\hat{\lambda}_n$ has the familiar expression in the typical GMM framework. The variances and covariances of $g_n(\lambda_0)$ can be derived from the formulas for a quadratic function (see, e.g., Lee, 2001a). For any $n \times n$ square matrix $A = [a_{ij}]$, let $\text{Diag}(A) = diag(a_{11}, \ldots, a_{nn})$ be the diagonal matrix and $\text{vec}_D(A) = (a_{11}, \ldots, a_{nn})'$ be the vector formed by the diagonal elements $a_{11}, \ldots, a_{nn}$ of A. Appendix A provides a summary of these notations and others used in this paper for easy reference. The components of $var(g_n(\lambda_0))$ are

$$\text{var}(\mathscr{E}'_n M_n P_{jn} M_n \mathscr{E}_n) = (\mu_4 - 3\sigma_0^4)\,\text{tr}(\text{Diag}^2(M_n P_{jn} M_n)) + \sigma_0^4\,\text{tr}(M_n P_{jn} M_n P^{\text{s}}_{jn}),$$

$$\begin{aligned} E(\mathscr{E}'_n M_n P_{jn} M_n \mathscr{E}_n \cdot \mathscr{E}'_n M_n P_{kn} M_n \mathscr{E}_n) &= (\mu_4 - 3\sigma_0^4)\, vec'_D(M_n P_{jn} M_n) vec_D(M_n P_{kn} M_n) \\ &\quad + \sigma_0^4\,\text{tr}(M_n P_{jn} M_n P^{\text{s}}_{kn} M_n), \end{aligned}$$

$\text{var}(Q'_{n1} M_n \mathscr{E}_n) = \sigma_0^2 Q'_{n1} M_n Q_{n1}$, and $E(Q'_{n1} M_n \mathscr{E}_n \cdot \mathscr{E}'_n M_n P_{jn} M_n \mathscr{E}_n) = \mu_3 Q'_{n1} M_n \text{vec}_D(M_n P_{jn} M_n)$, where $\mu_3 = E(\varepsilon_i^3)$ and $\mu_4 = E(\varepsilon_i^4)$. When ε_i has a symmetric distribution (or more general, $\mu_3 = 0$), quadratic moments $\varepsilon'_{x,n} P_{jn} \varepsilon_{x,n}$ will be uncorrelated with linear moments $Q_{n1}\varepsilon_{x,n}$ and the variance matrix of $g_n(\lambda_0)$ will be block-diagonal. The following lemma can simplify the expression of $\frac{1}{n} Var(g_n(\lambda_0))$ in the limit. It says that in many relevant quantities, the presence of M_n can be ignored in large samples. The proof of this lemma is given in Appendix C.

Lemma 1. *Suppose that the sequence of $n \times n$ constant matrices $\{P_n\}$ is uniformly bounded in absolute value in both row and column sums. The elements of $\{P_n\}$ and the sequence of vectors $\{q_n\}$ are uniformly bounded. Then, under Assumption 2 for X_n,*

(1) $\text{tr}(P_n M_n) = \text{tr}(M_n P_n) = \text{tr}(P_n) + \text{O}(1)$,
(2) $\text{tr}(M_n P_n M_n P^{\text{s}}_n) = \text{tr}(P_n P^{\text{s}}_n) + \text{O}(1)$,
(3) $vec'_D(M_n P_n) q_n = vec'_D(P_n) q_n + \text{O}(1)$ *and* $vec'_D(M_n P_n M_n) q_n = vec'_D(P_n) q_n + \text{O}(1)$, *and*
(4) $vec'_D(M_n P_{1n} M_n)\, vec_D(M_n P_{2n} M_n) = vec'_D(P_{1n})\, vec_D(P_{2n}) + \text{O}(1)$.
Furthermore, when $P_{n,ij} = \text{O}(\frac{1}{h_n})$ uniformly for all i, j, where h_n is a rate not larger than the rate n,
(5) $\text{tr}(Diag^2(M_n P_n M_n)) = \sum_{i=1}^n P^2_{n,ii} + \text{O}(\frac{1}{h_n})$.

Denote $\omega_n = (vec_D(P_{1n}), \ldots, vec_D(P_{mn}))$, which is a $n \times m$ matrix. Let

$$\Omega_n = \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega'_n \omega_n + \sigma_0^4 \Delta_n & \mu_3 \omega'_n M_n Q_{n1} \\ \mu_3 Q'_{n1} M_n \omega_n & \sigma_0^2 Q'_{n1} M_n Q_{n1} \end{pmatrix}, \tag{4.6}$$

where

$$\Delta_n = \begin{pmatrix} \text{tr}(P_{1n} P^{\text{s}}_{1n}) & \ldots & \text{tr}(P_{1n} P^{\text{s}}_{mn}) \\ \vdots & & \vdots \\ \text{tr}(P_{1n} P^{\text{s}}_{mn}) & \ldots & \text{tr}(P_{mn} P^{\text{s}}_{mn}) \end{pmatrix}. \tag{4.7}$$

Lemma 1shows that $\frac{1}{n} Var(g_n(\lambda_0)) - \frac{1}{n}\Omega_n = \text{O}(\frac{1}{n}) = \text{o}(1)$.[12]

[12]The term $M_n Q_{n1}$ cannot be simplified because it is the projection of Q_{n1} into the orthogonal space of the column space of X_n.

For large group interaction cases, i.e., $\lim_{n\to\infty} h_n = \infty$, $\lim_{n\to\infty}\frac{1}{n}\Omega_n$ can be further simplified. As elements of P_n's are $O(\frac{1}{h_n})$, $\frac{1}{n}\omega_n'\omega_n = O(\frac{1}{h_n^2})$, $\frac{1}{n}\Delta_n = O(\frac{1}{h_n})$ and $\frac{1}{n}\omega_n' M_n Q_{n1} = O(\frac{1}{h_n})$ by Lemma B.2. Therefore, as $h_n \to \infty$, $\frac{1}{n}\Omega_n - \frac{1}{n}\Omega_n^* = O(\frac{1}{h_n}) = o(1)$, where

$$\Omega_n^* = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_0^2 Q_{n1}' M_n Q_{n1} \end{pmatrix}. \tag{4.8}$$

Let $F_n = (F_{n1}, \ldots, F_{nm}, F_{nx})$, where F_{nj}, $j = 1, \ldots, m$, are s-dimensional column vectors and F_{nx} is a $s \times (r-k)$ matrix. The above simplification implies that $\sqrt{n}(\hat{\lambda}_n - \lambda_0) \xrightarrow{d} N(0, \Sigma_\lambda)$, where

$$\begin{aligned}
\Sigma_\lambda = \lim_{n\to\infty} & \left[\frac{1}{n^2}(W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0)' M_n Q_{n1} F_{nx}' F_{nx} Q_{n1}' M_n W_n (I_n - \lambda_0 W_n)^{-1} X_n \beta_0\right]^{-1} \\
& \times \frac{\sigma_0^2}{n^3}(W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0)' M_n Q_{n1} F_{nx}' F_{nx} Q_{n1}' M_n Q_{n1} F_{nx}' F_{nx} Q_{n1}' \\
& \times M_n W_n (I_n - \lambda_0 W_n)^{-1} X_n \beta_0 \left[\frac{1}{n^2}(W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0)' M_n Q_{n1} F_{nx}' F_{nx} Q_{n1}' \right. \\
& \left. \times M_n W_n (I_n - \lambda_0 W_n)^{-1} X_n \beta_0 \right]^{-1}
\end{aligned} \tag{4.9}$$

in Theorem 1. An implication of (4.9) is that, for large group interactions, $\varepsilon_{x,n}'(\lambda) P_{jn} \varepsilon_{x,n}(\lambda)$ does not provide an asymptotic effect in addition to that of $Q_{n1}\varepsilon_{x,n}$ for the GMM estimation of λ under Assumption 8.

It remains to consider the asymptotic distribution of the modified estimator $\hat{\beta}_n$ of β_0. Because

$$\hat{\beta}_n = (X_n' X_n)^{-1} X_n' S_n(\hat{\lambda}_n) Y_n = (X_n' X_n)^{-1} X_n' [I_n - (\hat{\lambda}_n - \lambda_0) W_n (I_n - \lambda_0 W_n)^{-1}](X_n \beta_0 + \mathscr{E}_n), \tag{4.10}$$

it follows that

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \left(\frac{1}{n} X_n' X_n\right)^{-1} \frac{1}{\sqrt{n}} X_n' \mathscr{E}_n - (X_n' X_n)^{-1} X_n' W_n (I_n - \lambda_0 W_n)^{-1} X_n \beta_0 \sqrt{n}(\hat{\lambda}_n - \lambda_0) + o_P(1). \tag{4.11}$$

From Theorem 1 and (4.11),

$$\begin{aligned}
& \begin{pmatrix} \frac{1}{n^2} D_n' F_n' F_n D_n & 0 \\ \frac{1}{n} X_n' W_n (I_n - \lambda_0 W_n)^{-1} X_n \beta_0 & \frac{1}{n} X_n' X_n \end{pmatrix} \sqrt{n}(\hat{\theta}_n - \theta_0) \\
& = \begin{pmatrix} \frac{1}{n} D_n' F_n' F_n \frac{1}{\sqrt{n}} g_n(\lambda_0) \\ \frac{1}{\sqrt{n}} \end{pmatrix} X_n' \mathscr{E}_n + o_P(1).
\end{aligned}$$

As $E(X_n'\mathscr{E}_n \cdot g_n'(\lambda_0)) = \eta_n' + O(1)$ where $\eta_n = (\mu_3 X_n'\omega_n, 0)'$, it follows that $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Sigma_\theta)$ where

$$\Sigma_\theta = \lim_{n\to\infty} \begin{pmatrix} \frac{1}{n^2} D_n'F_n'F_nD_n & 0 \\ \frac{1}{n}X_n'W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0 & \frac{1}{n}X_n'X_n \end{pmatrix}^{-1}$$
$$\times \begin{pmatrix} \frac{1}{n^2}D_n'F_n'F_n\frac{1}{n}Var(g_n(\lambda_0))F_n'F_nD_n & \frac{1}{n^2}D_n'F_n'F_n\eta_n \\ \frac{1}{n^2}\eta_n'F_n'F_nD_n & \sigma_0^2\frac{1}{n}X_n'X_n \end{pmatrix}$$
$$\times \begin{pmatrix} \frac{1}{n^2}D_n'F_n'F_nD_n & \frac{1}{n}(W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0)'X_n \\ 0 & \frac{1}{n}X_n'X_n \end{pmatrix}^{-1}. \tag{4.12}$$

In summary, the estimates $\hat{\lambda}_n$ and $\hat{\beta}_n$ are $\sqrt{n}$-consistent, and asymptotically normal under the crucial condition in Assumption 8 in addition to the basic regularity conditions in Assumptions 1–7.

The GMM estimation $\min_\lambda g_n'(\lambda)F_n'F_ng_n(\lambda)$ can be implemented by a nonlinear least squares programming. The $F_ng_n(\lambda)$ can be expanded in λ:

$$F_ng_n(\lambda) = \left(\sum_{j=1}^m F_{nj}Y_n'M_nP_{jn}M_nY_n + F_{nx}Q_{n1}'M_nY_n\right) - \left(\sum_{j=1}^m F_{nj}Y_n'W_n'M_nP_{jn}^{s}M_nY_n + F_{nx}Q_{n1}'M_nW_nY_n\right)\lambda + \left(\sum_{j=1}^m F_{nj}Y_n'W_n'M_nP_{jn}M_nW_nY_n\right)\lambda^2$$

and can be regarded as a 'residual' vector of a nonlinear least squares equation with s number of observations and two regressors. The vector of the dependent variable is $(\sum_{j=1}^m F_{nj}Y_n'M_nP_{jn}M_nY_n + F_{nx}Q_{n1}'M_nY_n)$ and $(\sum_{j=1}^m F_{nj}Y_n'W_n'M_nP_{jn}^{s}M_nY_n + F_{nx}Q_{n1}'M_n W_nY_n)$ and $(\sum_{j=1}^m F_{nj}Y_n'W_n'M_nP_{jn}M_nW_nY_n)$ are the two vectors of regressors with nonlinear coefficients λ and $-\lambda^2$. The GMM minimization corresponds to the minimization of the sum of squared residuals.

4.2. The modified GMM estimator with optimum weighting

With $g_n(\lambda)$ in (2.8) and the variance matrix Ω_n of $g_n(\lambda_0)$ in (4.6), the optimum GMM estimation for λ_0 in the modified approach is

$$\min_\lambda g_n'(\lambda)\Omega_n^{-1}g_n(\lambda). \tag{4.13}$$

Theorem 2. *Under Assumptions* (1–8) *where* $\{h_n\}$ *is a bounded sequence, and the conditions that* $\lim_{n\to\infty}\frac{1}{n}\Omega_n$ *and* $\lim_{n\to\infty}\frac{1}{n}D_n'\Omega_n^{-1}D_n$ *exist and are nonsingular, the modified optimum*

GMM estimator $\tilde{\lambda}_n$ *from* $\min_{\lambda\in\Lambda} g_n'(\lambda)\hat{\Omega}_n^{-1} g_n(\lambda)$, *where* $g_n(\lambda)$ *is based on* (2.8) *and* $\frac{1}{n}\hat{\Omega}_n - \frac{1}{n}\Omega_n = o_P(1)$, *is consistent and*

$$\sqrt{n}(\tilde{\lambda}_n - \lambda_0) \xrightarrow{d} N\left(0, \lim_{n\to\infty}\left(\frac{1}{n}D_n'\Omega_n^{-1}D_n\right)^{-1}\right). \tag{4.14}$$

The corresponding modified optimum GMM estimator $\hat{\theta}_n$ *of* θ_0 *is asymptotically normal with the asymptotic variance*

$$Asy.Var(\hat{\theta}_n) = \begin{pmatrix} D_n'\Omega_n^{-1}D_n & 0 \\ X_n'W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0 & X_n'X_n \end{pmatrix}^{-1} \begin{pmatrix} D_n'\Omega_n^{-1}D_n & D_n'\Omega_n^{-1}\eta_n \\ \eta_n'\Omega_n^{-1}D_n & \sigma_0^2 X_n'X_n \end{pmatrix} \times \begin{pmatrix} D_n'\Omega_n^{-1}D_n & (W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0)'X_n \\ 0 & X_n'X_n \end{pmatrix}^{-1}. \tag{4.15}$$

The results of the above theorem can be valid only for the case with small group interactions, i.e., $\{h_n\}$ is a bounded sequence, because the limit of $\frac{1}{n}\Omega_n$ needs to be nonsingular. For the large group interaction case with h_n being a divergent sequence, the limiting matrix of $\frac{1}{n}\Omega_n$ will be singular as seen from (4.8). In this case, the quadratic moments $\varepsilon_{x,n}'(\lambda)P_{jn}\varepsilon_{x,n}(\lambda)$ are dominated by the linear moment $Q_{n1}\varepsilon_{x,n}(\lambda)$ (see the discussion below (4.9)). The corresponding modified optimum GMM estimator turns out to be the familiar 2SLS estimator as shown in Appendix C.

The modified GMM estimation is a sequential procedure. Instead of this sequential estimation, the relevant moment functions can be stacked together and λ and β can be jointly estimated. The following subsections compare the efficiency of the various GMM approaches. For the divergent h_n case, the modified optimum GMM estimator is the familiar 2SLS estimator. So it remains to consider the bounded h_n case.

4.3. The modified GMM estimator vs joint optimum GMM estimators

4.3.1. The recursive moment functions (2.9)

The variance matrix $Var(f_n(\theta_0))$ of (2.9) is asymptotically equal to

$$\Omega_{c,n} = \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega_n'\omega_n + \sigma_0^4\Delta_n & \mu_3\omega_n'M_nQ_{n1} & \mu_3\omega_n'X_n \\ \mu_3 Q_{n1}'M_n\omega_n & \sigma_0^2 Q_{n1}'M_nQ_{n1} & 0 \\ \mu_3 X_n'\omega_n & 0 & \sigma_0^2 X_n'X_n \end{pmatrix}, \tag{4.16}$$

in that $\frac{1}{n}Var(f_n(\theta_0)) - \frac{1}{n}\Omega_{c,n} = o(1)$. Let $\hat{\theta}_{c,n}$ be the optimum GMM estimator from $\min_\theta f_n'(\theta)\Omega_{c,n}^{-1}f_n(\theta)$. With similar arguments as for Theorem 2, it can be shown that

$$\sqrt{n}(\hat{\theta}_{c,n} - \theta_0) \xrightarrow{D} N\left(0, \left(\lim_{n\to\infty}\frac{1}{n}D_{c,n}'\Omega_{c,n}^{-1}D_{c,n}\right)^{-1}\right), \tag{4.17}$$

where

$$D_{c,n} = \begin{pmatrix} \sigma_0^2 C_{mn} & (W_n(I_n - \lambda_0 W_n)^{-1} X_n \beta_0)' M_n Q_{n1} & (W_n(I_n - \lambda_0 W_n)^{-1} X_n \beta_0)' X_n \\ 0 & 0 & X_n' X_n \end{pmatrix}'. \tag{4.18}$$

Consider the distance matrix $\Omega_{s,n}^{-1}$, where

$$\Omega_{s,n} = \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega_n'\omega_n + \sigma_0^4 \Delta_n & \mu_3 \omega_n' M_n Q_{n1} & 0 \\ \mu_3 Q_{n1}' M_n \omega_n & \sigma_0^2 Q_{n1}' M_n Q_{n1} & 0 \\ 0 & 0 & \sigma_0^2 X_n' X_n \end{pmatrix}. \tag{4.19}$$

The matrix $\Omega_{s,n}$ differs from $\Omega_{c,n}$ in (4.16) by replacing the component $\mu_3 \omega_n' X_n$ in $\Omega_{c,n}$ with a zero submatrix. The matrix $\Omega_{s,n}$ is a block diagonal matrix.

Theorem 3. *The modified GMM estimation of λ and β is equivalent to the joint GMM estimation* $\min_\theta f_n'(\theta)\Omega_{s,n}^{-1} f_n(\theta)$, *where $f_n(\theta)$ is based on* (2.9).

As any possible nonzero correlation between recursive moment functions is ignored in the weighting matrix $\Omega_{s,n}$, the modified GMM estimator $\hat{\theta}_n$ can be inefficient relative to the joint optimum GMM estimator $\hat{\theta}_{c,n}$. However, when $\mu_3 = 0$ or that P_n's have zero diagonals, there is a zero correlation and $\Omega_{s,n} = \Omega_{c,n}$. Under such circumstances, the modified estimator is as efficient as the joint optimum GMM estimator.

Corollary 1. *Under the conditions in Theorem* 2, *when $\mu_3 = 0$ or for the case that P_n's have zero diagonals, the modified estimator $\hat{\theta}_n$ is as efficient as the joint optimum GMM estimator $\hat{\theta}_{c,n}$.*

The case $\mu_3 = 0$ occurs, for example, when the density of ε is symmetric. The other case holds when P_{ln}'s are designed to have zero diagonals.

4.4. The moment functions (2.4) versus the Recursive Moment Functions (2.9)

The $f_n^*(\theta)$ in (2.4) extends the moments $Q_n'\varepsilon_n(\theta)$, where $Q_n = (Q_{n1}, X_n)$, of a *2SLS* approach by incorporating $\varepsilon_n'(\theta) P_{jn} \varepsilon_n(\theta)$, $j = 1, \ldots, m$, in the estimation. It is of interest to compare the optimum GMM based on $f_n^*(\theta)$ with that based on $f_n(\theta)$ in (2.9). The optimum GMM estimator $\hat{\theta}_{c,n}^*$ from

$$\min_\theta f_n^{*'}(\theta)\Omega_{c,n}^{*-1} f_n^*(\theta) \tag{4.20}$$

is, from Lee (2001b),

$$\sqrt{n}(\hat{\theta}_{c,n}^* - \theta_0) \xrightarrow{D} \mathrm{N}\left(0, \left(\lim_{n\to\infty} \frac{1}{n} D_{c,n}^{*'} \Omega_{c,n}^{*-1} D_{c,n}^*\right)^{-1}\right), \tag{4.21}$$

where the variance matrix of $f_n^*(\theta_0)$ is

$$\Omega_{c,n}^* = \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega_n'\omega_n + \sigma_0^4 \Delta_n & \mu_3 \omega_n' Q_n \\ \mu_3 Q_n' \omega_n & \sigma_0^2 Q_n' Q_n \end{pmatrix} \tag{4.22}$$

and

$$D^*_{c,n} = \begin{pmatrix} \sigma_0^2 C_{mn} & (W_n(I_n - \lambda_0 W_n)^{-1} X_n \beta_0)' Q_n \\ 0 & X_n' Q_n \end{pmatrix}'. \tag{4.23}$$

The following theorem shows that asymptotic variances of the two optimum GMM estimators $\hat{\theta}_{c,n}$ in (4.17) and $\hat{\theta}^*_{c,n}$ in (4.21) are algebraically identical.

Theorem 4. *When $Q_n = (Q_{n1}, X_n)$, the identity $D^{*'}_{c,n} \Omega^{*-1}_{c,n} D^*_{c,n} = D'_{c,n} \Omega^{-1}_{c,n} D_{c,n}$ holds, under the conditions that $\Omega_{c,n}$ and $\Omega^*_{c,n}$ are nonsingular.*

Consequently, $\hat{\theta}_{c,n}$ and $\hat{\theta}^*_{c,n}$ from (4.17) and (4.21) have the same limiting distribution. From this result and Theorem 3, we conclude that, in general, $\hat{\theta}_n$ may be inefficient relative to $\hat{\theta}^*_{c,n}$ when $\mu_3 \omega_n \neq 0$. However, for the cases that $\mu_3 = 0$ or $diag(P_{ln}) = 0$, $l = 1, \ldots, m$, $\hat{\theta}_n$ is asymptotically efficient as $\hat{\theta}^*_{c,n}$.

4.5. The best modified GMM estimator

When $\mu_3 = 0$, the asymptotic variance of $\hat{\lambda}_n$ is

$$Var(\hat{\lambda}_n) = (D_n' \Omega_n^{-1} D_n)^{-1} = \left\{ C_{mn} \left[\left(\frac{\mu_4}{\sigma_0^4} - 3 \right) \omega_n' \omega_n + \Delta_n \right]^{-1} C'_{mn} \right.$$
$$\left. + \frac{1}{\sigma_0^2} (W_n(I_n - \lambda_0 W_n)^{-1} X_n \beta_0)' (M_n Q_{n1})_{(\mathrm{p})} (W_n(I_n - \lambda_0 W_n)^{-1} X_n \beta_0) \right\}^{-1}, \tag{4.24}$$

from (4.14) and (4.6), and the asymptotic variance of $\hat{\beta}_n$ from (4.15) becomes

$$Var(\hat{\beta}_n) = \sigma_0^2 (X_n' X_n)^{-1} + (X_n' X_n)^{-1} X_n' W_n (I_n - \lambda_0 W_n)^{-1} X_n \beta_0$$
$$\times (D_n' \Omega_n^{-1} D_n)^{-1} (W_n(I_n - \lambda_0 W_n)^{-1} X_n \beta_0)' X_n (X_n' X_n)^{-1}. \tag{4.25}$$

Because M_n is idempotent, by the generalized Schwartz inequality, $(M_n Q_{n1})_{(\mathrm{p})} \leqslant M_n$. Hence, the best Q_{n1} to minimize $Var(\hat{\lambda}_n)$ in (4.24) corresponds to $Q^*_{n1} = W_n(I_n - \lambda_0 W_n)^{-1} X_n \beta_0$. This is intuitively appealing because it is $E(W_n Y_n)$, which is also the best IV for $W_n Y_n$ in the 2SLS estimation (Lee, 2003). In consequence, Q^*_{n1} minimizes also $Var(\hat{\beta}_n)$ in (4.25).[13]

The best selection of P_n is available when $\mu_3 = 0$ and $\mu_4 = 3\sigma_0^4$, e.g., ε is normally distributed. In this situation, (4.24) becomes

$$D_n' \Omega_n^{-1} D_n = C_{mn} \Delta_n^{-1} C'_{mn} + \frac{1}{\sigma_0^2} (W_n(I_n - \lambda_0 W_n)^{-1} X_n \beta_0)' (M_n Q_{n1})_{(\mathrm{p})} W_n (I_n - \lambda_0 W_n)^{-1} X_n \beta_0.$$

By the representation $\mathrm{tr}(AB) = vec'(A') vec(B)$ for conformable matrices,

$$C_{mn} = \frac{1}{2} vec' \left(\left[W_n(I_n - \lambda_0 W_n)^{-1} - \frac{\mathrm{tr}(W_n(I_n - \lambda_0 W_n)^{-1})}{n} I_n \right]^{\mathrm{s}} \right) (vec(P^{\mathrm{s}}_{1n}) \ldots vec(P^{\mathrm{s}}_{mn}))$$

[13] In practice, one may use $W_n(I_n - \hat{\lambda}_n W_n)^{-1} X_n \hat{\beta}_n$, based on some initial consistent estimates $\hat{\lambda}_n$ and $\hat{\beta}_n$, as a feasible Q^*_{n1}. This is so also for the best P^*_n below.

and

$$\Delta_n = \tfrac{1}{2}(vec(P^s_{1n})\ldots vec(P^s_{mn}))'(vec(P^s_{1n})\ldots vec(P^s_{mn})).$$

The generalized Schwartz inequality implies

$$C_{mn}\Delta_n^{-1}C'_{mn} \leqslant \mathrm{tr}\left(\left[W_n(I_n-\lambda_0 W_n)^{-1} - \frac{\mathrm{tr}(W_n(I_n-\lambda_0 W_n)^{-1})}{n} I_n\right]^s W_n(I_n-\lambda_0 W_n)^{-1}\right).$$

Hence, the best P_n matrix is $P_n^* = W_n(I_n-\lambda_0 W_n)^{-1} - \frac{\mathrm{tr}(W_n(I_n-\lambda_0 W_n)^{-1})}{n} I_n$.[14] With the best P_n^* and Q_n^*, the asymptotic variance of the best modified GMM estimator $\hat{\lambda}_n$ is

$$(D'_n\Omega_n^{-1}D_n)^{-1} = \left\{\mathrm{tr}\left(\left[W_n(I_n-\lambda_0 W_n)^{-1} - \frac{\mathrm{tr}(W_n(I_n-\lambda_0 W_n)^{-1})}{n} I_n\right]^s W_n(I_n-\lambda_0 W_n)^{-1}\right)\right.$$
$$\left. + \frac{1}{\sigma_0^2}(W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0)'M_n W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0\right\}^{-1}. \quad (4.26)$$

When P_n is restricted to the smaller class of matrices with $diag(P_n)=0$, then $\omega_n = 0$ and $\Omega^*_{c,n}$ is a diagonal matrix similar to that of the case with $\mu_3 = 0$ and $\mu_4 = 3\sigma_0^4$. The generalized Schwartz inequality shows that the best IV matrix is $Q_n^* = (W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0, X_n)$, and the best P_n^* with a zero diagonal is $P_n^* = W_n(I_n-\lambda_0 W_n)^{-1} - Diag(W_n(I_n-\lambda_0 W_n)^{-1})$. The asymptotic variance of the best GMM estimator $\hat{\lambda}_n$ is

$$(D'_n\Omega_n^{-1}D_n)^{-1} = \left\{\mathrm{tr}\left(\left[W_n(I_n-\lambda_0 W_n)^{-1} - diag(W_n(I_n-\lambda_0 W_n)^{-1})\right]^s W_n(I_n-\lambda_0 W_n)^{-1}\right)\right.$$
$$\left. + \frac{1}{\sigma_0^2}(W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0)'M_n W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0\right\}^{-1}. \quad (4.27)$$

5. Estimation when $\lim_{n\to\infty}\frac{h_n}{n}(W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0)'M_n W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0$ is a finite positive constant in the case $\lim_{n\to\infty}h_n = \infty$

Assumption 8 assumes that $W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0$ and X_n are linearly independent in the limit. In a certain case, they may be linearly independent for all finite n but are nearly multicollinear in the limit in the sense that $\lim_{n\to\infty}\frac{1}{n}(W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0)' M_n W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0 = 0$. In this section, we consider the following situation:

Assumption 8′. $\lim_{n\to\infty}\frac{h_n}{n}(W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0)'M_n W_n(I_n-\lambda_0 W_n)^{-1}X_n\beta_0 = c$, where $0<c<\infty$, as $h_n\to\infty$.

An example satisfies Assumption 8′ is in the scenario of large group interactions in Case (1991) mentioned before. Suppose there are R districts and, for simplicity, there are m spatial units in each district. In this case, $h_n = m$ and $\frac{h_n}{n} = R$. If x contains an intercept term and x's are i.i.d. across spatial units and districts, it can be shown (see in Lee, 2004,

[14] This best P_n^* is derived over sets of $P_{jn}, j=1,\ldots,m$ for any finite m. It is not just the best one over the set of a single P_n. Therefore, any P_n in addition to P_n^* for GMM estimation will not improve the asymptotic efficiency of the best GMM estimator of θ_0.

footnote 15) that, as $m \to \infty$,

$$\operatorname*{plim}_{n\to\infty} \frac{h_n}{n}(W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0)' M_n W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0 = \frac{1}{(1-\lambda_0)^2}\beta_0' \Sigma_x \beta_0,$$

where $\Sigma_x = E[(x-\mu)'(x-\mu)]$ is the variance matrix of x, is finite and nonzero when $\beta_0 \neq 0$.

Under the situation in Assumption 8′, elements of $M_n W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0$ would likely have the order $O(\frac{1}{\sqrt{h_n}})$. It is natural to rescale IV matrix Q_{n1} so that its elements have also order $O(\frac{1}{\sqrt{h_n}})$. Any rescale will, in principle, not change the moment equation $E(Q_{n1}'\varepsilon_n) = 0$. But it does make quadratic and linear moments in a similar scale. Therefore, Q_{n1} shall be rescaled by dividing $\sqrt{h_n}$ if necessary, so that the following setting holds:

Assumption 9. The elements of Q_{n1} have order $O(\frac{1}{\sqrt{h_n}})$.

This implies, in particular, elements of $\frac{h_n}{n} Q_{n1}' M_n Q_{n1}$ have order O(1). Under the circumstance in this section, as $h_n \to \infty$, in order for the relevant central limit theorem in Lemma B.9 to be applicable, Assumption 5 needs to be slightly strengthened.

Assumption 5′. If $\{h_n\}$ is a divergent sequence, $\lim_{n\to\infty} \frac{h_n^{1+\frac{2}{\delta}}}{n} = 0$ where $\delta > 0$ such that $E(|\varepsilon|^{4+2\delta})$ exists.

Theorem 5. *Under Assumptions* 1–4, 5′, 6, 7, 8′, 9, $\lim_{n\to\infty} h_n = \infty$, *and that* $\lim_{n\to\infty} \frac{h_n}{n} E(g_n(\lambda))$ *does not lie in the orthogonal space of the column space of* F_0 *for any* $\lambda \neq \lambda_0$, *then the modified GMM estimator* $\hat{\lambda}_n$ *from* $\min_{\lambda\in\Lambda} g_n'(\lambda) F_n' F_n g_n(\lambda)$, *where* $g_n(\lambda)$ *is based on* (2.8), *is consistent and*

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = \left(\frac{h_n}{n} D_n' F_n' F_n D_n\right)^{-1} D_n' F_n' \sqrt{\frac{h_n}{n}} F_n g_n(\lambda_0) + o_P(1) \xrightarrow{d} N(0, \Sigma_\lambda), \tag{5.1}$$

where

$$\Sigma_\lambda = \lim_{n\to\infty} \left(\frac{h_n}{n} D_n' F_n' F_n D_n\right)^{-1} D_n' F_n' \left(\frac{h_n}{n} F_n \Omega_n F_n'\right) F_n D_n \left(\frac{h_n}{n} D_n' F_n' F_n D_n\right)^{-1}. \tag{5.2}$$

An implication of Theorem 5 is that the modified GMM estimator of $\hat{\lambda}_n$ under Assumption 8′ has the slower $\sqrt{\frac{n}{h_n}}$-rate of convergence. This is so also for the corresponding 2SLS estimator of λ_0 because of the near multicollinearity of $W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0$ and X_n. In the situation of Assumption 8′, linear moments do not dominate the quadratic moments and the 2SLS estimator of λ_0 can be improved upon by the additional quadratic moments.

From the asymptotic distribution of Theorem 5, the optimum distance matrix is apparently Ω_n^{-1}. Under Assumptions 3, 4 and 9, $|\frac{h_n}{n} vec_D'(P_n) M_n Q_{n1}| \leqslant (\frac{h_n}{n} vec_D'(P_n) vec_D(P_n))^{1/2} (\frac{h_n}{n} Q_{n1}' M_n Q_{n1})^{1/2} = O(\frac{1}{\sqrt{h_n}}) = o(1)$, which implies that $\sqrt{\frac{h_n}{n}} Q_{n1}' \mathscr{E}_n$ and $\sqrt{\frac{h_n}{n}} \mathscr{E}_n' P_n \mathscr{E}_n$ are asymptotically uncorrelated, regardless whether μ_3 is zero or not. Thus, the best IV Q_{n1} for the linear moment is $W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0$. Also, because

$\frac{1}{n}\omega_n'\omega_n = O(\frac{1}{h_n^2})$, $\frac{h_n}{n}\Omega_n = \frac{h_n}{n}\begin{pmatrix}\sigma_0^4\Delta_n & 0 \\ 0 & \sigma_0^2 Q_{n1}' M_n Q_{n1}\end{pmatrix} + o(1)$, the best P_n is $(W_n(I_n - \lambda_0 W_n)^{-1} - \frac{\text{tr}(W_n(I_n-\lambda_0 W_n)^{-1})}{n} I_n)$ even if ε's are not normally distributed.

As for the modified GMM estimator $\hat{\beta}_n$ in (4.10), its asymptotic distribution can be derived from (4.11):

$$\sqrt{\frac{n}{h_n}}(\hat{\beta}_n - \beta_0) = -\left(\frac{X_n'X_n}{n}\right)^{-1} \frac{X_n' W_n(I_n - \lambda_0 W_n)^{-1} X_n\beta_0}{n} \sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) + o_P(1). \tag{5.3}$$

The asymptotic distribution of $\hat{\beta}_n$ may have the same $\sqrt{\frac{n}{h_n}}$-rate of convergence as that of $\hat{\lambda}_n$.

In summary, when the circumstance in Assumption 8 does not hold but has been changed to that in Assumption 8′, the estimates $\hat{\lambda}_n$ and $\hat{\beta}_n$ are asymptotically normal but their rate of convergence is the slower $\sqrt{\frac{n}{h_n}}$-rate instead of the usual $\sqrt{n}$-rate for the case $\lim_{n\to\infty} h_n = \infty$.

6. Estimation of MRSAR models under multicollinearity of $W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0$ and X_n

Both Assumptions 8 and 8′ rule out the cases that $W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0$ and X_n can be perfectly multicollinear. In this section, we shall consider the more general situation, which includes the multicollinearity case. It considers the remaining situations not covered under Assumptions 8 and 8′ for both bounded or divergent $\{h_n\}$.

Assumption 8″. $\lim_{n\to\infty} \frac{h_n}{n}(W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0)' M_n (W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0) = 0.$

When $W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0$ and X_n are multicollinear for large n, there will be no valid IV for the estimation of λ because the reduced form (3.4) is simply a regression equation in X_n. Therefore, any linear moment $Q_{n1}\varepsilon_{x,n}(\lambda)$ would not be useful and the 2SLS method is not applicable. When $W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0$ and X_n are not perfectly multicollinear but have the feature in Assumption 8″, the quadratic moments shall dominate any linear moments, that is an implication from the preceding section. The appropriate approach shall include only the quadratic moments. Thus the empirical joint moments shall have $Q_n = X_n$ and

$$f_n^*(\theta) = (\varepsilon_n'(\theta)P_{1n}\varepsilon_n(\theta), \ldots, \varepsilon_n'(\theta)P_{mn}\varepsilon_n(\theta), \varepsilon_n'(\theta)X_n)' \tag{6.1}$$

in place of (2.4). The method of elimination and substitution shall have $\hat{\beta}_n(\lambda) = (X_n'X_n)^{-1} X_n'(I_n - \lambda W_n)Y_n$ and

$$g_n(\lambda) = (\varepsilon_{x,n}'(\lambda)P_{1n}\varepsilon_{x,n}(\lambda), \ldots, \varepsilon_{x,n}'(\lambda)P_{mn}\varepsilon_{x,n}(\lambda)) \tag{6.2}$$

after substitution. The corresponding (simplified) variance matrix of the moment function $g_n(\lambda_0)$ is

$$\Omega_n = (\mu_4 - 3\sigma_0^4)\omega_n'\omega_n + \sigma_0^4\Delta_n, \tag{6.3}$$

where Δ_n is in (4.7). Lemma B.2 implies that elements of Ω_n are of order $O(\frac{n}{h_n})$. The identification of λ_0 will depend on correlation across spatial units in $(I_n - \lambda_0 W_n)^{-1}\mathscr{E}_n$:

Assumption 10. The $\lim_{n\to\infty}\frac{h_n}{n}\operatorname{tr}(P_{jn}W_n(I_n-\lambda_0 W_n)^{-1})\neq 0$ for some j and

$$\lim_{n\to\infty}\frac{h_n}{n}[\operatorname{tr}(P^{s}_{1n}W_n(I_n-\lambda_0 W_n)^{-1}),\ldots,\operatorname{tr}(P^{s}_{mn}W_n(I_n-\lambda_0 W_n)^{-1})]'$$

is linearly independent of

$$\lim_{n\to\infty}\frac{h_n}{n}[\operatorname{tr}((I_n-\lambda_0 W'_n)^{-1}W'_nP_{1n}W_n(I_n-\lambda_0 W_n)^{-1}),\ldots,\operatorname{tr}((I_n-\lambda_0 W'_n)^{-1}\times W'_nP_{mn}W_n(I_n-\lambda_0 W_n)^{-1})]'.$$

The consistency and asymptotic distribution of this modified GMM estimator $\hat{\lambda}_n$ are in the following theorem.

Theorem 6. *Under Assumptions* 1–5, 5′, 6, 7, 8″, 10 *and that* $\lim_{n\to\infty}\frac{h_n}{n}E(g_n(\lambda))$ *does not lie in the orthogonal space of the columns of* F_0 *for any* $\lambda\neq\lambda_0$*, the modified GMM estimator* $\hat{\lambda}_n$ *from* $\min_{\lambda\in\Lambda} g'_n(\lambda)F'_nF_ng_n(\lambda)$*, where* $g_n(\lambda)$ *is in* (6.2)*, is consistent and*

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n-\lambda_0)=\left(\frac{h_n}{n}\sigma_0^2C'_{mn}F'_nF_nC_{mn}\right)^{-1}C'_{mn}F'_nF_n\sqrt{\frac{h_n}{n}}g_n(\lambda_0)+o_P(1)\xrightarrow{d}N(0,\Sigma_\lambda),$$

where

$$\Sigma_\lambda=\lim_{n\to\infty}\left(\frac{h_n}{n}\sigma_0^2C'_{mn}F'_nF_nC_{mn}\right)^{-1}C'_{mn}F'_n\left(\frac{h_n}{n}F_n\Omega_nF'_n\right)F_nC_{mn}\left(\frac{h_n}{n}\sigma_0^2C'_{mn}F'_nF_nC_{mn}\right)^{-1}. \tag{6.4}$$

Theorem 6 is applicable for both bounded or divergent $\{h_n\}$. When $\{h_n\}$ is a divergent sequence, the modified GMM estimator $\hat{\lambda}_n$ has a slower than $\sqrt{n}$-rate of convergence. In any case, the linear moments do not have effects on the asymptotic distribution of $\hat{\lambda}_n$ in (6.4) because C_{mn} and Ω_n in (6.3) do not depend on X_n.

From (6.4), the generalized Schwartz inequality implies that the optimum weighting matrix is $(\frac{h_n}{n}\Omega_n)^{-1}$. The modified optimum GMM estimator $\tilde{\lambda}_n$ from

$$\min_{\lambda\in\Lambda} g'_n(\lambda)\hat{\Omega}_n^{-1}g_n(\lambda), \tag{6.5}$$

where $\frac{h_n}{n}\hat{\Omega}_n-\frac{h_n}{n}\Omega_n=o_P(1)$, will have the asymptotic distribution:

$$\sqrt{\frac{n}{h_n}}(\tilde{\lambda}_n-\lambda_0)\xrightarrow{d}N\left(0,\frac{1}{\sigma_0^4}\lim_{n\to\infty}\left(\frac{h_n}{n}C'_{mn}\Omega_n^{-1}C_{mn}\right)^{-1}\right). \tag{6.6}$$

When $\mu_4=3\sigma_0^4$ or P_n's have zero diagonal, $\Omega_n=\sigma_0^4\Delta_n$ in (6.3). For the case that $h_n\to\infty$, $\frac{h_n}{n}(\Omega_n-\sigma_0^4\Delta_n)=o(1)$ because $\omega'_n\omega_n=O(\frac{n}{h_n^2})$. Thus, when ε's are normally distributed or h_n goes to infinity, the best P_n shall be $(W_n(I_n-\lambda_0W_n)^{-1}-\frac{\operatorname{tr}(W_n(I_n-\lambda_0W_n)^{-1})}{n}I_n)$. For the class of moment functions with $Diag(P_n)=0$, the best P_n is $(W_n(I_n-\lambda_0W_n)^{-1}-Diag(W_n(I_n-\lambda_0W_n)^{-1}))$.

The asymptotic distribution of $\hat{\beta}_n$ follows from (4.11):

$$\sqrt{\frac{n}{h_n}}(\hat{\beta}_n - \beta_0) = \frac{1}{\sqrt{h_n}}\left(\frac{1}{n}X_n'X_n\right)^{-1}\frac{1}{\sqrt{n}}X_n'\mathscr{E}_n - (X_n'X_n)^{-1}X_n'W_n(I_n - \lambda_0 W_n)^{-1} \times X_n\beta_0 \cdot \sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) + o_P\left(\frac{1}{\sqrt{h_n}}\right). \tag{6.7}$$

The first and second terms on the right-hand side of (6.7) would be uncorrelated if $\mu_3 = 0$. For the case that $\lim_{n\to\infty} h_n = \infty$, the first term vanishes and the limiting distribution of $\hat{\beta}_n$ will be determined by $\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0)$ if all the components of $\lim_{n\to\infty}(X_n'X_n)^{-1}X_n'W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0$ are nonzero. In the event that $W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0$ and X_n are multicollinear, $W_n(I_n - \lambda_0 W_n)^{-1}X_n\beta_0 = X_n c_n$ for some column vector $c_n \neq 0$. Let $c_n = (c_{1n}', c_{2n}')'$ where all the components of c_{1n} are nonzero in the limit and $c_{2n} = 0$. Let J_n be the selection matrix such that $J_n c_n = c_{2n}$. Then, for the case $\lim_{n\to\infty} h_n = \infty$,

$$\sqrt{\frac{n}{h_n}}(\hat{\beta}_{1n} - \beta_{10}) = -c_{1n}\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) + o_P(1) \tag{6.8}$$

and

$$\sqrt{n}(\hat{\beta}_{2n} - \beta_{20}) = J_n\left(\frac{1}{n}X_n'X_n\right)^{-1}\frac{1}{\sqrt{n}}X_n'\mathscr{E}_n + o_P(1), \tag{6.9}$$

where $\beta = (\beta_1', \beta_2')$. The component $\hat{\beta}_{1n}$ has the $\sqrt{\frac{n}{h_n}}$-rate of convergence but the component $\hat{\beta}_{2n}$ has the usual $\sqrt{n}$-rate.

In summary, under the circumstance in Assumption 8$''$, for the case with h_n being a bounded sequence, the estimates $\hat{\lambda}_n$ and $\hat{\beta}_n$ can still be asymptotically normal with the $\sqrt{n}$-rate of convergence. However, for the case with $h_n \to \infty$, $\hat{\lambda}_n$ and $\hat{\beta}_n$ are asymptotically normal but, in general, $\hat{\lambda}_n$ and certain components of $\hat{\beta}_n$ may have the slower $\sqrt{\frac{n}{h_n}}$-rate of convergence.

7. Conclusion

This paper introduces a modified GMM method based on the method of elimination and substitution for estimating the MRSAR model. This GMM approach isolates the nonlinear estimation of the MRSAR model on the spatial effect parameter. The parameters of exogenous regressors can be estimated by the least squares method once the estimate of the spatial effect parameter is available. This approach is computationally simpler than other GMM approaches which extend the 2SLS estimation in Kelejian and Prucha (1998) and estimate jointly the spatial effect parameter and the regression coefficients of the model.

For the ML method, the likelihood function (under normality assumption) involves a Jacobian term. The computation of the Jacobian, i.e., the determinant of $(I_n - \lambda W_n)$, has received much attention in the literature (e.g. Ord, 1975; Pace and Barry, 1997; Smirnov and Anselin, 2001). The GMM approach has the feature that neither the determinant nor the inverse of $(I_n - \lambda W_n)$ need to be computed. But, the regression coefficient subvector β

in the log likelihood function of the ML approach can be easily concentrated out, and the resulting concentrated (or profile) likelihood function involves only the spatial effect parameter λ, while the joint GMM objective function does not have such a feature. The modified GMM approach based on the method of elimination and substitution has both computational advantage features.

In addition to computational issues, we investigate the relative efficiency of the modified and joint GMM estimators. The modified GMM estimator can be as efficient as the joint optimum GMM estimator under disturbances with a zero third order moment. Other cases depend on the design of IV matrices used for the moment functions in GMM estimation. We have also considered issues on selecting the best IV matrix for estimation.

The modified GMM approach and the asymptotic analysis in the paper have focused on the MRSAR model with a single spatial lag. The asymptotic analysis and the results derived in this paper may be generalized to high order spatial lags models. The nonlinear estimation will then focus on the several spatial effect parameters.[15] This may be, in particular, useful as the ML approach cannot easily carry out for spatial autoregression models with higher order lags.

Acknowledgments

I am grateful to the National Science Foundation, Economics Program for the financial support of my research under the Grant number SES-0111380. I appreciate receiving valuable comments and suggestions from I. Prucha and two anonymous referees.

Appendix A. Summary of notations used in the text and the proofs

$$A_{(\mathrm{p})} = A(A'A)^{-1}A'.$$

$$M_n = I_n - X_n(X_n'X_n)^{-1}X_n'.$$

$$Q_n = (Q_{n1}, X_n).$$

$$\mathscr{E}_n = (\varepsilon_1, \ldots, \varepsilon_n)'.$$

$$\varepsilon_{x,n}(\lambda) = M_n(I_n - \lambda W_n)Y_n.$$

$$\varepsilon_{x,n} = \varepsilon_{x,n}(\lambda_0).$$

$A^{\mathrm{s}} = A + A'$ where A is a square matrix.

$\mathrm{Diag}(A) = diag(a_{11}, \ldots, a_{nn})$ a diagonal matrix formed by the diagonal elements of a $n \times n$ matrix A.

[15]Many features of a spatial regression with higher spatial lags, however, have not been adequately understood or have not been studied. For example, there may be complicated restrictions on the parameter space of spatial lags coefficients.

$vec_D(A) = (a_{11}, \ldots, a_{nn})'$ a column vector formed by the diagonal elements of a $n \times n$ matrix A.

$$G_n = W_n(I_n - \lambda_0 W_n)^{-1}.$$

$$\omega_n = (vec_D(P_{1n}), \ldots, vec_D(P_{mn})).$$

$$C_{mn} = (\mathrm{tr}(P^{\mathrm{s}}_{1n}G_n), \ldots, \mathrm{tr}(P^{\mathrm{s}}_{mn}G_n)) = \frac{1}{2}vec'\left(\left(G_n - \frac{\mathrm{tr}(G_n)}{n}I_n\right)^{\mathrm{s}}\right)(vec(P^{\mathrm{s}}_{1n}), \ldots, vec(P^{\mathrm{s}}_{mn})).$$

$$D_n = (\sigma_0^2 C_{mn}, (G_n X_n \beta_0)' M_n Q_{n1})', \text{ where } \sigma_0^2 = E(\varepsilon^2).$$

$$D_{c,n} = \begin{pmatrix} D_n & (G_n X_n \beta_0)' X_n \\ 0 & X_n' X_n \end{pmatrix}'.$$

$$D^*_{c,n} = \begin{pmatrix} \sigma_0^2 C_{mn} & (G_n X_n \beta_0)' Q_n \\ 0 & X_n' Q_n \end{pmatrix}'.$$

$$\Delta_n = \begin{pmatrix} \mathrm{tr}(P_{1n}P^{\mathrm{s}}_{1n}) & \ldots & \mathrm{tr}(P_{1n}P^{\mathrm{s}}_{mn}) \\ \vdots & & \vdots \\ \mathrm{tr}(P_{1n}P^{\mathrm{s}}_{mn}) & \ldots & \mathrm{tr}(P_{mn}P^{\mathrm{s}}_{mn}) \end{pmatrix} = \tfrac{1}{2}(vec(P^{\mathrm{s}}_{1n}) \ldots vec(P^{\mathrm{s}}_{mn}))'(vec(P^{\mathrm{s}}_{1n}) \ldots vec(P^{\mathrm{s}}_{mn})).$$

$$\Omega_n = \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega_n'\omega_n + \sigma_0^4 \Delta_n & \mu_3 \omega_n' M_n Q_{n1} \\ \mu_3 Q_{n1}' M_n \omega_n & \sigma_0^2 Q_{n1}' M_n Q_{n1} \end{pmatrix}, \text{ where } \mu_l = E(\varepsilon^l) \text{ for } l = 3, 4.$$

$$\Omega_{c,n} = \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega_n'\omega_n + \sigma_0^4 \Delta_n & \mu_3 \omega_n' M_n Q_{n1} & \mu_3 \omega_n' X_n \\ \mu_3 Q_{n1}' M_n \omega_n & \sigma_0^2 Q_{n1}' M_n Q_{n1} & 0 \\ \mu_3 X_n' \omega_n & 0 & \sigma_0^2 X_n' X_n \end{pmatrix}.$$

$$\Omega_{\mathrm{s},n} = \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega_n'\omega_n + \sigma_0^4 \Delta_n & \mu_3 \omega_n' M_n Q_{n1} & 0 \\ \mu_3 Q_{n1}' M_n \omega_n & \sigma_0^2 Q_{n1}' M_n Q_{n1} & 0 \\ 0 & 0 & \sigma_0^2 X_n' X_n \end{pmatrix}.$$

$$\Omega^*_{c,n} = \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega_n'\omega_n + \sigma_0^4 \Delta_n & \mu_3 \omega_n' Q_n \\ \mu_3 Q_n' \omega_n & \sigma_0^2 Q_n' Q_n \end{pmatrix}.$$

$$\eta_n = (\mu_3 X_n' \omega_n, 0)'.$$

Appendix B. Some lemmas

Lemma B.1. *Suppose that the elements of the* $n \times k$ *matrices* X_n *are uniformly bounded for all* n; *and* $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ *exists and is nonsingular, then the projectors* $X_n(X_n'X_n)^{-1}X_n'$ *and* $I_n - X_n(X_n'X_n)^{-1}X_n'$ *are uniformly bounded in absolute value in both row and column sums. Furthermore, the elements of* $X_n(X_n'X_n)^{-1}X_n'$ *are of uniform order* $\mathrm{O}(\frac{1}{n})$.

Proof. Let $B_n = (\frac{1}{n}X_n'X_n)^{-1}$. From the assumptions, B_n converges to a finite limit. Therefore, there exists a constant c_b such that $|b_{n,ij}| \leqslant c_b$ for all n, where $b_{n,ij}$ is the (i,j)th element of B_n. By the uniform boundedness of X_n, there exists a constant c_x such that $|x_{n,ij}| \leqslant c_x$ for all i,j and n. Let $A_n = X_n(X_n'X_n)^{-1}X_n' = \frac{1}{n}\sum_{s=1}^{k}\sum_{r=1}^{k} b_{n,rs}x_{n,r}x_{n,s}'$, where $x_{n,r}$ is the rth column of X_n. It follows that $\sum_{j=1}^{n}|a_{n,ij}| = \frac{1}{n}\sum_{j=1}^{n}|\sum_{s=1}^{k}\sum_{r=1}^{k} b_{n,rs}x_{n,ir}x_{n,js}| \leqslant k^2 c_b c_x^2$, for all i and n, where $x_{n,ir}$ is the (i,r)th element of X_n. Similarly, $\sum_{i=1}^{n}|a_{n,ij}| = \frac{1}{n}\sum_{i=1}^{n}|\sum_{s=1}^{k}\sum_{r=1}^{k} b_{n,rs}x_{n,ir}x_{n,js}| \leqslant k^2 c_b c_x^2$ for all j and n. That is, $X_n(X_n'X_n)^{-1}X_n'$ are uniformly bounded in absolute value in both row and column sums. Consequently, $(I_n - X_n(X_n'X_n)^{-1}X_n')$ are also uniformly bounded in absolute value in both row and column sums.

The (i,j)th element of $X_n(X_n'X_n)^{-1}X_n'$ is $\frac{1}{n}\sum_{r=1}^{k}\sum_{s=1}^{k} b_{n,rs}x_{n,ir}x_{n,js}$. It follows that

$$\left|\frac{1}{n}\sum_{r=1}^{k}\sum_{s=1}^{k} b_{n,rs}x_{n,ir}x_{n,js}\right| \leqslant \frac{k^2 c_b c_x^2}{n} = \mathrm{O}\left(\frac{1}{n}\right). \qquad \square$$

Lemma B.2. *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* $\mathrm{O}(\frac{1}{n})$ (*resp.* $\mathrm{O}(\frac{1}{h_n})$) *uniformly in all i and j*; *and* $\{B_n\}$ *is a sequence of* $n \times n$ *matrices.*

(1) *If* $\{B_n\}$ *are uniformly bounded in absolute value in column sums, then the elements of* A_nB_n *have the uniform order* $\mathrm{O}(\frac{1}{n})$ (*resp.* $\mathrm{O}(\frac{1}{h_n})$).
(2) *If* $\{B_n\}$ *are uniformly bounded in absolute value in row sums, then the elements of* B_nA_n *have the uniform order* $\mathrm{O}(\frac{1}{n})$ (*resp.* $\mathrm{O}(\frac{1}{h_n})$).

For both cases (1) *and* (2), $|\mathrm{tr}(A_nB_n)| = |\mathrm{tr}(B_nA_n)| = \mathrm{O}(1)$ (*resp.* $\mathrm{O}(\frac{n}{h_n})$).

Proof. This is proved in Lee (2001a). $\square$

Lemma B.3. *Suppose that the elements of the two sequences of n-dimensional column vectors* $\{p_n\}$ *and* $\{q_n\}$ *are uniformly bounded. If* $\{A_n\}$ *is uniformly bounded in absolute value in either row or column sums, then* $|p_n'A_nq_n| = \mathrm{O}(n)$.

Proof. This is in Lee (2001b). It is a trivial result. $\square$

Lemma B.4. *Suppose that the sequence of* $n \times n$ *matrices* $\{A_n\}$ *are uniformly bounded in absolute value in both row and column sums. Elements of the* $n \times k$ *matrices* X_n *are uniformly bounded*; $\lim_{n\to\infty} \frac{X_n'X_n}{n}$ *exists and is nonsingular. Let* $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$. *Then*

(i) $\mathrm{tr}(M_nA_n) = \mathrm{tr}(A_n) + \mathrm{O}(1)$,
(ii) $\mathrm{tr}(A_n'M_nA_n) = \mathrm{tr}(A_n'A_n) + \mathrm{O}(1)$,

(iii) $\mathrm{tr}[(M_nA_n)^2] = \mathrm{tr}(A_n^2) + \mathrm{O}(1)$, *and*
(iv) $\mathrm{tr}[(A_n'M_nA_n)^2] = \mathrm{tr}[(M_nA_nA_n')^2] = \mathrm{tr}[(A_nA_n')^2] + \mathrm{O}(1)$.
Furthermore, if $A_{n,ij} = \mathrm{O}(\frac{1}{h_n})$ *for all i and j, where* h_n *is a rate not faster than the rate n, then*
(v) $\mathrm{tr}^2(M_nA_n) = \mathrm{tr}^2(A_n) + \mathrm{O}(\frac{n}{h_n})$,
(vi) $\sum_{i=1}^n [(M_nA_n)_{ii}]^2 = \sum_{i=1}^n [A_{n,ii}]^2 + \mathrm{O}(\frac{1}{h_n})$, *and*
(vii) $\sum_{i=1}^n [(A_nM_n)_{ii}]^2 = \sum_{i=1}^n [A_{n,ii}]^2 + \mathrm{O}(\frac{1}{h_n})$.

Proof. The assumptions imply that elements of the $k \times k$ matrix $(\frac{1}{n}X_n'X_n)^{-1}$, $\frac{1}{n}X_n'A_nX_n$, $\frac{1}{n}X_n'A_nA_n'X_n$ and $\frac{1}{n}X_n'A_n^2X_n$ are bounded for large enough n by Lemma B.3. It follows that

$$\mathrm{tr}(M_nA_n) = \mathrm{tr}(A_n) - \mathrm{tr}[(X_n'X_n)^{-1}X_n'A_nX_n] = \mathrm{tr}(A_n) + \mathrm{O}(1),$$

$$\mathrm{tr}(A_n'M_nA_n) = \mathrm{tr}(A_n'A_n) - \mathrm{tr}[(X_n'X_n)^{-1}X_n'A_nA_n'X_n] = \mathrm{tr}(A_n'A_n) + \mathrm{O}(1),$$

and

$$\mathrm{tr}[(M_nA_n)^2] = \mathrm{tr}(A_n^2) - 2\,\mathrm{tr}[(X_n'X_n)^{-1}X_n'A_n^2X_n] + \mathrm{tr}[(X_n'X_n)^{-1}X_n'B_nX_n],$$

where $B_n = A_nX_n(X_n'X_n)^{-1}X_n'A_n$. The B_n is uniformly bounded in absolute value in both row and column sums because both A_n and $X_n(X_n'X_n)^{-1}X_n'$ are. Hence, $\mathrm{tr}[(M_nA_n)^2] = \mathrm{tr}(A_n^2) + \mathrm{O}(1)$, which is (iii).

By (iii), $\mathrm{tr}[(A_n'M_nA_n)^2] = \mathrm{tr}[(M_nA_nA_n')^2] = \mathrm{tr}[(A_nA_n')^2] + \mathrm{O}(1)$ because A_nA_n' is uniformly bounded in absolute value in both row and column sums. The (i) implies that $\mathrm{tr}^2(M_nA_n) = (\mathrm{tr}(A_n) + \mathrm{O}(1))^2 = \mathrm{tr}^2(A_n) + 2\mathrm{tr}(A_n) \cdot \mathrm{O}(1) + \mathrm{O}(1) = \mathrm{tr}^2(A_n) + \mathrm{O}(n)$. Because A_n is uniformly bounded in absolute value in column sums and elements of X_n are uniformly bounded, $X_n'A_ne_{ni} = \mathrm{O}(1)$ for all i, where e_{ni} is the ith unit column vector of dimension n. By Lemma B.2, elements of $X_n(X_n'X_n)^{-1}X_n'A_n$ are of uniform order $\mathrm{O}(\frac{1}{n})$. Hence, $\sum_{i=1}^n [(M_nA_n)_{ii}]^2 = \sum_{i=1}^n (A_{n,ii} - e_{ni}'X_n(X_n'X_n)^{-1}X_n'A_ne_{ni})^2 = \sum_{i=1}^n (A_{n,ii} + \mathrm{O}(\frac{1}{n}))^2 = \sum_{i=1}^n\ [(A_{n,ii})^2 + 2A_{n,ii} \cdot \mathrm{O}(\frac{1}{n}) + \mathrm{O}(\frac{1}{n^2})] = \sum_{i=1}^n [A_{n,ii}]^2 + \mathrm{O}(\frac{1}{h_n})$ because $A_{n,ii} = \mathrm{O}(\frac{1}{h_n})$. Finally, $\sum_{i=1}^n\ [(A_nM_n)_{ii}]^2 = \sum_{i=1}^n (A_{n,ii} - e_{ni}'X_n(X_n'X_n)^{-1}X_n'e_{ni})^2 = \sum_{i=1}^n (A_{n,ii} + \mathrm{O}(\frac{1}{n}))^2 = \sum_{i=1}^n [(A_{n,ii})^2 + 2A_{n,ii} \cdot \mathrm{O}(\frac{1}{n}) + \mathrm{O}(\frac{1}{n^2})] = \sum_{i=1}^n (A_{n,ii})^2 + \mathrm{O}(\frac{1}{h_n})$. □

Lemma B.5. *Suppose that both* A_n *and* B_n *are uniformly bounded in absolute value in either row or column sums. Elements of the* $n \times k$ *matrices* X_n *are uniformly bounded*; $\lim_{n\to\infty} \frac{X_n'X_n}{n}$ *exists and is nonsingular. Then*

(1) $\mathrm{tr}(A_nM_nB_n) = \mathrm{tr}(A_nB_n) + \mathrm{O}(1)$, *and*
(2) $\mathrm{tr}(M_nA_nM_nB_n) = \mathrm{tr}(A_nB_n) + \mathrm{O}(1)$.

Proof. The assumptions in this Lemma imply that B_nA_n is uniformly bounded in absolute value in either row or column sums, and the elements of $\frac{1}{n}X_n'B_nA_nX_n$ are uniformly bounded. Therefore,

$$\mathrm{tr}(A_nM_nB_n) = \mathrm{tr}(A_nB_n) - \mathrm{tr}[(X_n'X_n)^{-1}X_n'B_nA_nX_n] = \mathrm{tr}(A_nB_n) + \mathrm{O}(1).$$

Then (2) follows from (1). Let $C_n = M_n A_n$. If A_n is uniformly bounded in absolute value in row (column) sums, C_n is uniformly bounded in absolute value in row (column) sums. Hence, $\mathrm{tr}(C_n M_n B_n) = \mathrm{tr}(C_n B_n) + \mathrm{O}(1) = \mathrm{tr}(M_n A_n B_n) + \mathrm{O}(1) = \mathrm{tr}(A_n B_n) + \mathrm{O}(1)$ by (1). □

Lemma B.6. *Suppose that A_n is a square matrix with its column sums being uniformly bounded in absolute value and elements of the $n \times k$ matrix C_n are uniformly bounded. The ε_i's in $\mathscr{E}_n = (\varepsilon_1, \ldots, \varepsilon_n)$ are i.i.d. $(0, \sigma_0^2)$. Then, $\frac{1}{\sqrt{n}} C_n' A_n \mathscr{E}_n = \mathrm{O_P}(1)$. Furthermore, if the limit of $\frac{1}{n} C_n' A_n A_n' C_n$ exists and is positive definite, then $\frac{1}{\sqrt{n}} C_n' A_n \mathscr{E}_n \xrightarrow{D} \mathrm{N}(0, \sigma_0^2 \lim_{n\to\infty} \frac{1}{n} C_n' A_n A_n' C_n)$.*

Proof. The first result follows from the Chebyshev inequality, and the second one follows from the Lindeberg–Feller central limit theorem. □

Lemma B.7. *Suppose that A_n is a constant $n \times n$ matrix uniformly bounded in absolute value in both row and column sums, and the ε_i's in $\mathscr{E}_n = (\varepsilon_1, \ldots, \varepsilon_n)$ are i.i.d. $(0, \sigma_0^2)$. Let c_n be a column vector of constants. If $\frac{h_n}{n} c_n' c_n = \mathrm{o}(1)$, then $\sqrt{\frac{h_n}{n}} c_n' A_n \mathscr{E}_n = \mathrm{o_P}(1)$. On the other hand, if $\frac{h_n}{n} c_n' c_n = \mathrm{O}(1)$, then $\sqrt{\frac{h_n}{n}} c_n' A_n \mathscr{E}_n = \mathrm{O_P}(1)$.*

Proof. The first result follows from Chebyshev's inequality if $var(\sqrt{\frac{h_n}{n}} c_n' A_n \mathscr{E}_n) = \sigma_0^2 \frac{h_n}{n} c_n' A_n A_n' c_n$ goes to zero. Let Λ_n be the diagonal matrix of eigenvalues of $A_n A_n'$ and Γ_n be the orthonormal matrix of eigenvectors. As eigenvalues in absolute values are bounded by any norm of the matrix, eigenvalues in Λ_n in absolute value are uniformly bounded because $\|A_n\|_\infty$ (or $\|A_n\|_1$) are uniformly bounded. Hence, $|\frac{h_n}{n} c_n' A_n A_n' c_n| \leqslant \frac{h_n}{n} c_n' \Gamma_n \Gamma_n' c_n \cdot |\lambda_{n,\max}| = \frac{h_n}{n} c_n' c_n |\lambda_{n,\max}| = \mathrm{o}(1)$, where $\lambda_{n,\max}$ is the eigenvalue of $A_n A_n'$ with the largest absolute value.

When $\frac{h_n}{n} c_n' c_n = \mathrm{O}(1)$, $\frac{h_n}{n} c_n' A_n A_n' c_n \leqslant \frac{h_n}{n} c_n' c_n |\lambda_{n,\max}| = \mathrm{O}(1)$. Hence,

$$\mathrm{var}\left(\sqrt{\frac{h_n}{n}} c_n' A_n \mathscr{E}_n\right) = \sigma_0^2 \frac{h_n}{n} c_n' A_n A_n' c_n = \mathrm{O}(1).$$

Therefore, $\sqrt{\frac{h_n}{n}} c_n' A_n \mathscr{E}_n = \mathrm{O_P}(1)$. □

Lemma B.8. *Suppose that $\{A_n\}$ are uniformly bounded in absolute value in either row or column sums, and the elements $a_{n,ij}$ of A_n have the order $\mathrm{O}(\frac{1}{h_n})$ uniformly in i and j. The ε_i's in $\mathscr{E}_n = (\varepsilon_1, \ldots, \varepsilon_n)$ are i.i.d. with zero mean and its fourth order moment exists.*

Then,

(i) $E(\mathscr{E}_n' A_n \mathscr{E}_n) = \mathrm{O}(\frac{n}{h_n})$ *and* $\mathrm{var}(\mathscr{E}_n' A_n \mathscr{E}_n) = \mathrm{O}(\frac{n}{h_n})$,
(ii) *furthermore, if* $\lim_{n\to\infty} \frac{h_n}{n} = 0$, *then* $\frac{h_n}{n}(\mathscr{E}_n' A_n \mathscr{E}_n - E(\mathscr{E}_n' A_n \mathscr{E}_n)) = \mathrm{o_P}(1)$.

Proof. This is in Lee (2001a). □

Lemma B.9. *Suppose that $\{A_n\}$ is a sequence of symmetric matrices uniformly bounded in absolute value in row and column sums and $\{b_n\}$ is a sequence of constant vectors with its elements uniformly bounded. The ε_i's in $\mathscr{E}_n = (\varepsilon_1, \ldots, \varepsilon_n)$ are i.i.d. with zero mean and its fourth order moment exists. Let $\sigma_{q_n}^2$ be the variance of q_n where $q_n = b_n' \mathscr{E}_n + \mathscr{E}_n' A_n \mathscr{E}_n - \sigma^2 \mathrm{tr}(A_n)$. Assume that the variance $\sigma_{q_n}^2$ is $\mathrm{O}(n/h_n)$ with $\{(h_n/n)\sigma_{q_n}^2\}$ bounded away* *from zero, the elements of* A_n *are of uniform order* $O(1/h_n)$ *and the elements of* b_n *are of uniform order* $O(1/\sqrt{h_n})$.

If either (i) $\{h_n\}$ *is bounded from above and is bounded away from zero, or* (ii) *when* $h_n \to \infty$, $\lim_{n\to\infty} \frac{h_n^{1+2/\delta}}{n} = 0$ *for some* $\delta > 0$ *such that* $E(|\varepsilon|^{4+2\delta})$ *exists, then* $\frac{q_n}{\sigma_{q_n}} \xrightarrow{d} N(0,1)$.

Proof. The result of (i) follows from the central limit theorem in Kelejian and Prucha (2001). Its generalization in (ii) is in Lee (2004). □

Appendix C. Proofs

Proof of Theorem 1. It shall be shown that $(\frac{1}{n}F_n g_n(\lambda))'(\frac{1}{n}F_n g_n(\lambda)) - E(\frac{1}{n}F_n g_n(\lambda))' E(\frac{1}{n}F_n g_n(\lambda)) = o_P(1)$ uniformly in λ in Λ. As

$$\left(\frac{1}{n}F_n g_n(\lambda)\right)'\left(\frac{1}{n}F_n g_n(\lambda)\right) - E\left(\frac{1}{n}F_n g_n(\lambda)\right)' E\left(\frac{1}{n}F_n g_n(\lambda)\right)$$
$$= \left[\frac{1}{n}F_n g_n(\lambda) - E\left(\frac{1}{n}F_n g_n(\lambda)\right)\right]'\left[\frac{1}{n}F_n g_n(\lambda) - E\left(\frac{1}{n}F_n g_n(\lambda)\right)\right]$$
$$+ 2\left[\frac{1}{n}F_n g_n(\lambda) - E\left(\frac{1}{n}F_n g_n(\lambda)\right)\right]' E\left(\frac{1}{n}F_n g_n(\lambda)\right),$$

it is sufficient to show that $\frac{1}{n}F_n g_n(\lambda) - E(\frac{1}{n}F_n g_n(\lambda)) = o_P(1)$ uniformly in λ in Λ, and $E(\frac{1}{n}F_n g_n(\lambda))$ is uniformly bounded in Λ.

As $(I_n - \lambda W_n)(I_n - \lambda_0 W_n)^{-1} = I_n + (\lambda_0 - \lambda)G_n$,

$$\varepsilon_{x,n}(\lambda) = M_n(I_n - \lambda W_n)Y_n = (\lambda_0 - \lambda)M_n G_n X_n \beta_0 + M_n(I_n + (\lambda_0 - \lambda)G_n)\mathscr{E}_n.$$

It follows that

$$F_n g_n(\lambda) = \sum_{l=1}^{m} F_{nl}\varepsilon'_{x,n}(\lambda)P_{ln}\varepsilon_{x,n}(\lambda) + F_{nx}Q'_{n1}\varepsilon_{x,n}(\lambda)$$
$$= (\lambda_0 - \lambda)^2 \sum_{l=1}^{m} F_{nl}(M_n G_n X_n \beta_0)' P_{ln}(M_n G_n X_n \beta_0) + (\lambda_0 - \lambda)l_{1n}(\lambda)$$
$$+ q_n(\lambda) + (\lambda_0 - \lambda)F_{nx}Q'_{n1}M_n G_n X_n \beta_0 + l_{2n}(\lambda),$$

where

$$l_{1n}(\lambda) = \sum_{l=1}^{m} F_{nl}(M_n G_n X_n \beta_0)' P^s_{ln} M_n(I_n + (\lambda_0 - \lambda)G_n)\mathscr{E}_n,$$

$$l_{2n}(\lambda) = F_{nx}Q'_{n1}M_n(I_n + (\lambda_0 - \lambda)G_n)\mathscr{E}_n$$

and

$$q_n(\lambda) = \sum_{l=1}^{m} F_{nl}\mathscr{E}'_n(I_n + (\lambda_0 - \lambda)G_n)' M_n P_{ln} M_n(I_n + (\lambda_0 - \lambda)G_n)\mathscr{E}_n.$$

Lemma B.6 implies that $\frac{1}{n}l_{1n}(\lambda)$ and $\frac{1}{n}l_{2n}(\lambda)$ are $o_P(1)$ uniformly in λ in Λ. The uniform convergence holds because λ appears linearly and Λ is a bounded set. Lemma B.8 implies that $\frac{1}{n}q_n(\lambda) - E(\frac{1}{n}q_n(\lambda)) = o_P(1)$ uniformly in λ in Λ. Therefore, $\frac{1}{n}F_n g_n(\lambda) =$

$\frac{1}{n}E(F_n g_n(\lambda)) + o_P(1)$. The equicontinuity of $\frac{1}{n}E(F_n g_n(\lambda))$ on Λ is apparent from its expression:

$$E(F_n g_n(\lambda)) = (\lambda_0 - \lambda)F_{nx}Q'_{n1}M_nG_nX_n\beta_0 + \sum_{l=1}^{m} F_{nl}(M_nG_nX_n\beta_0)'P_{ln}(M_nG_nX_n\beta_0) \times(\lambda_0-\lambda)^2 + E(q_n(\lambda)),$$

where $E(q_n(\lambda)) = \sigma_0^2\sum_{l=1}^{m}F_{nl}\{\text{tr}[M_nP_{ln}] + (\lambda_0-\lambda)\,\text{tr}[G'_nM_nP^{\text{s}}_{ln}M_n] + (\lambda_0-\lambda)^2\,\text{tr}[G'_nM_nP_{ln}M_nG_n]\}$.

Consider the identification uniqueness of λ_0 (White, 1994). Suppose the identification uniqueness condition that $\liminf_{n\to\infty}\min_{\lambda\in\bar{N}_\varepsilon(\lambda_0)}E(\frac{1}{n}F_ng_n(\lambda))'E(\frac{1}{n}F_ng_n(\lambda))>0$ does not hold for some $\varepsilon>0$, where $\bar{N}_\varepsilon(\lambda_0)$ is the complement of the open ε-neighborhood of λ_0 in Λ. Then, there exists a sequence $\{\lambda_n\}$ convergent to $\lambda_+\in\Lambda$ where $\lambda_+\neq\lambda_0$ such that $F_0\lim_{n\to\infty}E(\frac{1}{n}g_n(\lambda_n))=0$. By the equicontinuity of $E(\frac{1}{n}g_n(\lambda))$, $\lim_{n\to\infty}[E(\frac{1}{n}g_n(\lambda_n)) - E(\frac{1}{n}g_n(\lambda_+))]=0$. This implies, in turn, that $F_0\lim_{n\to\infty}E(\frac{1}{n}g_n(\lambda_+))=0$, a contradiction to the identification assumption. Hence, the identification uniqueness of λ_0 holds. The consistency of $\hat{\lambda}_n$ follows from the uniform convergence and the identification uniqueness condition (White, 1994, Theorem 3.4).

By expansion,

$$\begin{aligned}
&\frac{1}{n}\varepsilon'_{x,n}(\lambda)P^{\text{s}}_nM_nG_n(X_n\beta_0+\mathscr{E}_n)\\
&\quad= (\lambda_0-\lambda)\frac{1}{n}[M_nG_n(X_n\beta_0+\mathscr{E}_n)]'P^{\text{s}}_nM_nG_n(X_n\beta_0+\mathscr{E}_n) + \frac{1}{n}\mathscr{E}'_nM_nP^{\text{s}}_nM_nG_n(X_n\beta_0+\mathscr{E}_n)\\
&\quad= (\lambda_0-\lambda)\frac{1}{n}[(M_nG_nX_n\beta_0)'P^{\text{s}}_nM_nG_nX_n\beta_0 + \sigma_0^2\,\text{tr}(P^{\text{s}}_nG_nG'_n)] + \frac{\sigma_0^2}{n}\,\text{tr}(P^{\text{s}}_nG_n) + o_P(1),
\end{aligned}$$

by Lemmas B.4 and B.5. Hence, $\frac{1}{n}\frac{\partial g_n(\bar{\lambda}_n)}{\partial\lambda} = -\frac{1}{n}(P^{\text{s}}_{1n}\varepsilon_{x,n}(\bar{\lambda}_n),\ldots,P^{\text{s}}_{mn}\varepsilon_{x,n}(\bar{\lambda}_n),Q_{n1})'M_nG_n(X_n\beta_0+\mathscr{E}_n) = -\frac{1}{n}D_n + o_P(1)$, for any consistent estimate $\bar{\lambda}_n$ of λ_0.

The $\frac{1}{\sqrt{n}}\mathscr{E}'_nP^{\text{s}}_nX_n(X'_nX_n)^{-1}X'_n\mathscr{E}_n$ and $\frac{1}{\sqrt{n}}\mathscr{E}'_nX_n(X'_nX_n)^{-1}X'_nP_nX_n(X'_nX_n)^{-1}X'_n\mathscr{E}_n$ are both of $o_P(1)$. It follows that $\frac{1}{\sqrt{n}}\mathscr{E}'_nM_nP_nM_n\mathscr{E}_n = \frac{1}{\sqrt{n}}\mathscr{E}'_nP_n\mathscr{E}_n + o_P(1)$. By the central limit theorem of linear-quadratic form in Kelejian and Prucha (2001),

$$\frac{1}{\sqrt{n}}g_n(\lambda_0) = \frac{1}{\sqrt{n}}(\mathscr{E}'_nP_{1n}\mathscr{E}_n,\ldots,\mathscr{E}'_nP_{mn}\mathscr{E}_n,\mathscr{E}'_nM_nQ_{n1})' + o_P(1)\xrightarrow{d} N\left(0,\lim_{n\to\infty}\frac{1}{n}\Omega_n\right).$$

The asymptotic distribution of $\hat{\lambda}_n$ follows from the Taylor expansion of the first order condition of the GMM minimization and

$$\sqrt{n}(\hat{\lambda}_n-\lambda_0) = \left(\frac{1}{n}\frac{\partial g'_n(\hat{\lambda}_n)}{\partial\lambda}F'_nF_n\frac{1}{n}\frac{\partial g_n(\bar{\lambda}_n)}{\partial\lambda}\right)^{-1}\frac{1}{n}\frac{\partial g'_n(\hat{\lambda}_n)}{\partial\lambda}F'_nF_n\frac{1}{\sqrt{n}}g_n(\lambda_0)\xrightarrow{d} N(0,\Sigma_\lambda). \qquad \square$$

Proof of Lemma 1. Because P_n and $P_nM_nP^{\text{s}}_n$ are uniformly bounded in absolute value in both row and column sums, (1) and (2) follow from Lemmas B.1 and B.4.

Note that $M_nP_n = P_n - X_n(X'_nX_n)^{-1}X'_nP_n$ and

$$\begin{aligned}
M_nP_nM_n &= P_n - X_n(X'_nX_n)^{-1}X'_nP_n - P_nX_n(X'_nX_n)^{-1}X'_n\\
&\quad + X_n(X'_nX_n)^{-1}X'_nP_nX_n(X'_nX_n)^{-1}X'_n.
\end{aligned}$$

As the elements of $X_n(X_n'X_n)^{-1}X_n'P_n$ are of uniform $O(\frac{1}{n})$ by Lemmas B.1 and B.2, and elements of q_n are uniformly bounded, $vec_D'(X_n(X_n'X_n)^{-1}X_n'P_n)q_n = n \cdot O(\frac{1}{n}) = O(1)$. Similarly, $vec_D'(P_nX_n(X_n'X_n)^{-1}X_n')q_n$ and $vec_D'(X_n(X_n'X_n)^{-1}X_n'P_nX_n(X_n'X_n)^{-1}X_n')q_n$ are of O(1). Therefore, $vec_D'(M_nP_n)q_n = vec_D'(P_n)q_n + O(1)$ and $vec_D'(M_nP_nM_n)q_n = vec_D'(P_n)q_n +$ O(1), which is (3). The result in (4) follows from (3). The (vi) and (vii) of Lemma B.4 imply that

$$\mathrm{tr}(Diag^2(M_nP_nM_n)) = \sum_{i=1}^{n}[(M_nP_nM_n)_{ii}]^2 = \sum_{i=1}^{n}[(P_nM_n)_{ii}]^2 + O\left(\frac{1}{h_n}\right)$$
$$= \sum_{i=1}^{n}(P_{n,ii})^2 + O\left(\frac{1}{h_n}\right),$$

which is (5). □

Proof of Theorem 2. The results follow from Theorem 1 and (4.14) as $F_n'F_n = (\frac{1}{n}\Omega_n)^{-1}$. □

Proof of Theorem 3. Because $\min_\beta(S_n(\lambda)Y_n - X_n\beta)'X_n(X_n'X_n)^{-1}X_n'(S_n(\lambda)Y_n - X_n\beta) = 0$,

$$\min_\theta f_n'(\theta)\Omega_{s,n}^{-1}f_n(\theta)$$
$$= \min_\lambda[g_n'(\lambda)\Omega_n^{-1}g_n(\lambda) + \frac{1}{\sigma_0^2}\min_\beta(S_n(\lambda)Y_n - X_n\beta)'X_n(X_n'X_n)^{-1}X_n'(S_n(\lambda)Y_n - X_n\beta)]$$
$$= \min_\lambda g_n'(\lambda)\Omega_n^{-1}g_n(\lambda).$$

Thus, the modified estimation corresponds to a GMM with the recursive moment functions $f_n(\theta)$ in (2.9) and the distance matrix $\Omega_{s,n}^{-1}$. □

Proof of Corollary 1. When $\mu_3\omega_n'X_n$ is not zero, $\Omega_{s,n} \neq \Omega_{c,n}$. When $\mu_3 = 0$ or P_n's have zero diagonals, $\omega_n = 0$ and $\Omega_{s,n} = \Omega_{c,n}$. The modified estimator is as efficient as the joint optimum GMM with $f_n(\theta)$. □

Proof of Theorem 4. Let $A = (\mu_4 - 3\sigma_0^4)\omega_n'\omega_n + \sigma_0^4\Delta_n$ and $D = \sigma_0^2C_{mn}$. The matrices $D_{c,n}$ and $\Omega_{c,n}$ can be rewritten into block matrix forms

$$D_{c,n}' = \begin{pmatrix} D & E_1 \\ 0 & E_2 \end{pmatrix}, \quad \Omega_{c,n} = \begin{pmatrix} A & B \\ B' & C \end{pmatrix},$$

where $E_1 = (G_nX_n\beta_0)'(M_nQ_{n1}, X_n)$, $E_2 = (0, X_n'X_n)$, $B = \mu_3\omega_n'(M_nQ_{n1}, X_n)$, and

$$C = \begin{pmatrix} \sigma_0^2Q_{n1}'M_nQ_{n1} & 0 \\ 0 & \sigma_0^2X_n'X_n \end{pmatrix}.$$

Correspondingly, the $D_{c,n}^*$ and $\Omega_{c,n}^*$ can be partitioned into

$$D_{c,n}^{*'} = \begin{pmatrix} D & E_1^* \\ 0 & E_2^* \end{pmatrix}, \quad \Omega_{c,n}^* = \begin{pmatrix} A & B^* \\ B^{*'} & C^* \end{pmatrix},$$

where $E_1^* = (G_nX_n\beta_0)'Q_n$, $E_2^* = X_n'Q_n$, $B^* = \mu_3\omega_n'Q_n$, and $C^* = \sigma_0^2Q_n'Q_n$.

For the inverses of $\Omega_{c,n}$ and $\Omega_{c,n}^*$, the inversion formula for partitioned matrix is useful. The first diagonal block of $\Omega_{c,n}^{-1}$ is $(A - BC^{-1}B')^{-1}$, and that of $\Omega_{c,n}^{*-1}$ is $(A - B^*C^{*-1}B^{*'})^{-1}$.

Explicitly, $A-BC^{-1}B' = A - \frac{\mu_3^2}{\sigma_0^2}\omega_n'[(M_nQ_{n1})_{(p)} + X_n(X_n'X_n)^{-1}X_n']\omega_n$, and $A - B^*C^{*-1}B^{*'} = A - \frac{\mu_3^2}{\sigma_0^2}\omega_n'(Q_n)_{(p)}\omega_n$. Because of the identity $(Q_n)_{(p)} - X_n(X_n'X_n)^{-1}X_n' = (M_nQ_{n1})_{(p)}$ when $Q_n = (Q_{n1}, X_n)$ (see, e.g., Ruud, 2000),

$$A - BC^{-1}B' = A - B^*C^{*-1}B^{*'}.$$

Denote this common matrix by H.

The inversion formula of partitioned matrix implies

$$\Omega_{c,n}^{-1} = \begin{pmatrix} H^{-1} & -H^{-1}BC^{-1} \\ -C'^{-1}B'H^{-1} & C^{-1} + C^{-1}B'H^{-1}BC^{-1} \end{pmatrix},$$

and, therefore, $D_{c,n}'\Omega_{c,n}^{-1}D_{c,n} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$ with

$$V_{11} = D'H^{-1}D' - E_1C^{-1}B'H^{-1}D' - DH^{-1}BC^{-1}E_1' + E_1C^{-1}E_1' + E_1C^{-1}B'H^{-1}BC^{-1}E_1',$$

$$V_{12} = -DH^{-1}BC^{-1}E_2' + E_1C^{-1}E_2' + E_1C^{-1}B'H^{-1}BC^{-1}E_2'$$

and

$$V_{22} = E_2C^{-1}E_2' + E_2C^{-1}B'H^{-1}BC^{-1}E_2'.$$

These expressions can be simplified. Because

$$BC^{-1}E_1' = \frac{\mu_3}{\sigma_0^2}\omega_n'[(M_nQ_{n1})_{(p)} + X_n(X_n'X_n)^{-1}X_n']G_nX_n\beta_0 = \frac{\mu_3}{\sigma_0^2}\omega_n'(Q_n)_{(p)}G_nX_n\beta_0$$

and $E_1C^{-1}E_1' = \frac{1}{\sigma_0^2}(G_nX_n\beta_0)'(Q_n)_{(p)}G_nX_n\beta_0$, it follows that

$$V_{11} = DH^{-1}D' - \frac{\mu_3}{\sigma_0^2}(G_nX_n\beta_0)'(Q_n)_{(p)}\omega_nH^{-1}D' - \frac{\mu_3}{\sigma_0^2}DH^{-1}\omega_n'(Q_n)_{(p)}G_nX_n\beta_0 + \frac{1}{\sigma_0^2}(G_nX_n\beta_0)'(Q_n)_{(p)}G_nX_n\beta_0 + \left(\frac{\mu_3}{\sigma_0^2}\right)^2(G_nX_n\beta_0)'(Q_n)_{(p)}\omega_nH^{-1}\omega_n'(Q_n)_{(p)}G_nX_n\beta_0.$$

Because $BC^{-1}E_2' = \frac{\mu_3}{\sigma_0^2}\omega_n'X_n$ and $E_1C^{-1}E_2' = \frac{1}{\sigma_0^2}(G_nX_n\beta_0)'X_n$,

$$V_{12} = -\frac{\mu_3}{\sigma_0^2}DH^{-1}\omega_n'X_n + \frac{1}{\sigma_0^2}(G_nX_n\beta_0)'X_n + \left(\frac{\mu_3}{\sigma_0^2}\right)^2(G_nX_n\beta_0)'(Q_n)_{(p)}\omega_nH^{-1}\omega_n'X_n.$$

Furthermore, because $E_2C^{-1}E_2' = \frac{1}{\sigma_0^2}X_n'X_n$,

$$V_{22} = \frac{1}{\sigma_0^2}X_n'X_n + \left(\frac{\mu_3}{\sigma_0^2}\right)^2X_n'\omega_nH^{-1}\omega_n'X_n.$$

For Ω_c^*, its inverse is

$$\Omega_c^{*-1} = \begin{pmatrix} H^{-1} & -H^{-1}B^*C^{*-1} \\ -C^{*-1}B^{*'}H^{-1} & C^{*-1} + C^{*-1}B^{*'}H^{-1}B^*C^{*-1} \end{pmatrix},$$

and $D_c^{*'}\Omega_c^{*-1}D_c^* = \begin{pmatrix} V_{11}^* & V_{12}^* \\ V_{21}^* & V_{22}^* \end{pmatrix}$, where

$$V_{11}^* = D'H^{-1}D - E_1^*C^{*-1}B^{*'}H^{-1}D - DH^{-1}B^*C^{*-1}E_1^{*'} + E_1^*C^{*-1}E_1^{*'} + E_1^*C^{*-1}B^{*'}H^{-1}B^*C^{*-1}E_1^{*'},$$

$$V_{12}^* = -DH^{-1}B^*C^{*-1}E_2^{*'} + E_1^*C^{*-1}E_2^{*'} + E_1^*C^{*-1}B^{*'}H^{-1}B^*C^{*-1}E_2^{*'}$$

and

$$V_{22}^* = E_2^*C^{*-1}E_2^{*'} + E_2^*C^{*-1}B^{*'}H^{-1}B^*C^{*-1}E_2^{*'}.$$

These expressions can be simplified as $E_1^*C^{*-1}B^{*'} = \frac{\mu_3}{\sigma_0^2}(G_nX_n\beta_0)'(Q_n)_{(\mathrm{p})}\omega_n$,

$$E_1^*C^{*-1}E_1^{*'} = \frac{1}{\sigma_0^2}(G_nX_n\beta_0)'(Q_n)_{(\mathrm{p})}G_nX_n\beta_0, \quad B^*C^{*-1}E_2^{*'} = \frac{\mu_3}{\sigma_0^2}\omega_n'(Q_n)_{(\mathrm{p})}X_n = \frac{\mu_3}{\sigma_0^2}\omega_n'X_n$$

and $E_1^*C^{*-1}E_2^{*'} = \frac{1}{\sigma_0^2}(G_nX_n\beta_0)'(Q_n)_{(\mathrm{p})}X_n = \frac{1}{\sigma_0^2}(G_nX_n\beta_0)'X_n$, where the last two equalities hold because X_n is in the column space of Q_n. Hence, it follows that $V_{11}^* = V_{11}$, $V_{12}^* = V_{12}$, and $V_{22}^* = V_{22}$.

In conclusion, when $Q_n = (Q_{n1}, X_n)$, $D_{c,n}'\Omega_{c,n}^{-1}D_{c,n} = D_{c,n}^{*'}\Omega_{c,n}^{*-1}D_{c,n}^*$. □

Proof of Theorem 5. The proof of this theorem is similar to that of Theorem 1 by taking into account the situation in Assumptions 8′ and 10 under $\lim_{n\to\infty} h_n = \infty$.

It shall be shown that $\frac{h_n}{n}g_n(\lambda) - E(\frac{h_n}{n}g_n(\lambda)) = \frac{h_n}{n}[(\lambda_0-\lambda)l_{1n}(\lambda) + l_{2n}(\lambda) + q_n(\lambda)] = o_P(1)$ uniformly in Λ, where l_{1n}, l_{2n} and q_n are defined in the proof of Theorem 1. Lemma B.7 implies that $\frac{h_n}{n}l_{1n}(\lambda)$ and $\frac{h_n}{n}l_{2n}(\lambda)$ are of $O_P(\sqrt{\frac{h_n}{n}})$ uniformly in Λ under Assumptions 8′ and 9. Lemma B.2 implies that elements of $(I_n + (\lambda_0-\lambda)G_n)'M_nP_{ln}M_n(I_n + (\lambda_0-\lambda)G_n)$ are of order $O(\frac{1}{h_n})$ because elements of P_{ln} has order $O(\frac{1}{h_n})$. Therefore, Lemma B.8 implies that $\frac{h_n}{n}[q_n(\lambda) - E(q_n(\lambda))] = o_P(1)$ uniformly in Λ.

Note that

$$\begin{aligned}\frac{h_n}{n}&E(F_ng_n(\lambda)) \\ &= (\lambda_0-\lambda)^2\sum_{l=1}^m F_{nl}\frac{h_n}{n}(G_nX_n\beta_0)'M_nP_{ln}M_nG_nX_n\beta_0 \\ &\quad + (\lambda_0-\lambda)\frac{h_n}{n}F_{nx}Q_{n1}'M_nG_nX_n\beta_0 + \frac{h_n}{n}E(q_n(\lambda)),\end{aligned}$$

where $E(q_n(\lambda)) = \sigma_0^2[(\lambda_0-\lambda)\sum_{l=1}^m F_{nl}\,\mathrm{tr}(G_n'M_nP_{ln}^{\mathrm{s}}M_n) + (\lambda_0-\lambda)^2\sum_{l=1}^m F_{nl}\,\mathrm{tr}(M_nP_{ln}M_nG_nG_n')]$. As $\frac{1}{2}P_n^{\mathrm{s}}$ is uniformly bounded in absolute value in either column or row sums and all its eigenvalues must be less than $\|P_n\|_\infty$ and $\|P_n\|_1$, those eigenvalues are uniformly bounded. Therefore,

$$\begin{aligned}\frac{h_n}{n}\Big|(G_nX_n\beta_0)'M_nP_nM_nG_nX_n\beta_0| &= \frac{1}{2}\frac{h_n}{n}|(G_nX_n\beta_0)'M_nP_n^{\mathrm{s}}M_nG_nX_n\beta_0| \\ &\leqslant |\lambda_{n,\max}|\frac{h_n}{n}(G_nX_n\beta_0)'M_nG_nX_n\beta_0 = O(1),\end{aligned}$$

where $\lambda_{n,\max}$ is the largest eigenvalue of $\frac{P_n^s}{2}$ in absolute value. Under Assumption 8′ and 9, the Cauchy inequality implies that $\frac{h_n}{n}|F_{j,nx}Q_n' M_n G_n X_n \beta_0| \leqslant (\frac{h_n}{n} F_{j,nx} Q_n' M_n Q_n F_{j,nx}')^{1/2} (\frac{h_n}{n}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0))^{1/2} = O(1)$, where $F_{j,nx}$ is the jth row of F_{nx}, $j = 1, \ldots, s$. Lemma B.2 implies that $\frac{h_n}{n}\operatorname{tr}(G_n' M_n P_n^s M_n)$ and $\frac{h_n}{n}\operatorname{tr}(M_n P_n M_n G_n G_n')$ are of order O(1). Hence, $\frac{h_n}{n} E(F_n g_n(\lambda))$ are equicontinuous on Λ.

The identification uniqueness condition that $\lim_{n\to\infty} \min_{\lambda \in \bar{N}_\varepsilon(\lambda_0)} E(\frac{h_n}{n} F_n g_n(\lambda))' E(\frac{h_n}{n} F_n g_n(\lambda)) > 0$ is satisfied under Assumption 8′ and the property of F_0 in the theorem. The consistency of $\hat{\lambda}_n$ follows.

For the asymptotic distribution,

$$\frac{h_n}{n}\frac{\partial g_n(\lambda)}{\partial \lambda} = -\frac{h_n}{n}(P_{1n}^s \varepsilon_{x,n}(\lambda), \ldots, P_{mn}^s \varepsilon_{x,n}(\lambda), Q_{n1})' M_n (G_n X_n \beta_0 + G_n \mathscr{E}_n).$$

Lemma B.7 implies that terms $\frac{h_n}{n}(G_n X_n \beta_0)' M_n P_n^s M_n \mathscr{E}_n$, $\frac{h_n}{n}(G_n X_n \beta_0)' M_n P_n^s M_n G_n \mathscr{E}_n$ and $\frac{h_n}{n} Q_{n1}' M_n G_n \mathscr{E}_n$ are of order $O_P(\sqrt{\frac{h_n}{n}})$. The $\frac{h_n}{n}[\mathscr{E}_n' M_n P_n^s M_n G_n \mathscr{E}_n - \sigma_0^2 \operatorname{tr}(P_n^s G_n)]$ and $\frac{h_n}{n}[\mathscr{E}_n' G_n' P_n^s M_n G_n \mathscr{E}_n - \sigma_0^2 \operatorname{tr}(P_n^s G_n G_n')]$ are $o_P(1)$ by Lemmas B.8 and B.5. Therefore,

$$\begin{aligned}
&\frac{h_n}{n}\varepsilon_{x,n}'(\lambda) P_n^s M_n G_n (X_n \beta_0 + \mathscr{E}_n) \\
&= (\lambda_0 - \lambda)\frac{h_n}{n}(X_n\beta_0 + \mathscr{E}_n)' G_n' M_n P_n^s M_n G_n (X_n \beta_0 + \mathscr{E}_n) + \frac{h_n}{n}\mathscr{E}_n' M_n P_n^s M_n G_n (X_n\beta_0 + \mathscr{E}_n) \\
&= (\lambda_0 - \lambda)\left[\frac{h_n}{n}(G_n X_n \beta_0)' M_n P_n^s M_n G_n X_n \beta_0 \right.\\
&\quad \left. + \sigma_0^2 \frac{h_n}{n}\operatorname{tr}(P_n^s G_n G_n')\right] + \sigma_0^2 \frac{h_n}{n}\operatorname{tr}(P_n^s G_n) + o_P(1),
\end{aligned}$$

uniformly in Λ. As $\frac{h_n}{n}(G_n X_n \beta_0)' M_n P_n^s M_n G_n X_n \beta_0$ and $\frac{h_n}{n}\operatorname{tr}(P_n^s G_n G_n')$ are of O(1), for any consistent estimate $\bar{\lambda}_n$ of λ_0, $\frac{h_n}{n}\frac{\partial g_n(\bar{\lambda}_n)}{\partial \lambda} = -\frac{h_n}{n} D_n + o_P(1)$.

Note that $F_n g_n(\lambda_0) = \sum_{l=1}^m F_{nl} \varepsilon_n' M_n P_{ln} M_n \varepsilon_n + F_{nx} Q_{n1}' M_n \varepsilon_n$. By expansion, $\varepsilon_n' M_n P_n M_n \varepsilon_n = \varepsilon_n' P_n \varepsilon_n - \varepsilon_n' X_n (X_n' X_n)^{-1} X_n' P_n^s \varepsilon_n + \varepsilon_n' X_n (X_n' X_n)^{-1} X_n' P_n X_n (X_n' X_n)^{-1} X_n' \varepsilon_n = \varepsilon_n' P_n \varepsilon_n + O_P(1)$. Therefore, by the CLT of quadratic and linear functions,

$$\sqrt{\frac{h_n}{n}} F_n g_n(\lambda_0) = \sqrt{\frac{h_n}{n}}\left(\varepsilon_n' \sum_{l=1}^m F_{nl} P_{nl} \varepsilon_n + F_{nx} Q_{n1}' M_n \varepsilon_n\right) + O_P(\sqrt{\frac{h_n}{n}}) \xrightarrow{d} N\left(0, \lim_{n\to\infty}\frac{h_n}{n} F_0 \Omega_n F_0'\right).$$

The asymptotic distribution of $\hat{\lambda}_n$ follows from

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = \left(\frac{h_n}{n}\frac{\partial g_n'(\bar{\lambda}_n)}{\partial \lambda} F_n' F_n \frac{h_n}{n}\frac{\partial g_n'(\bar{\lambda}_n)}{\partial \lambda}\right)^{-1} \frac{h_n}{n}\frac{\partial g_n'(\bar{\lambda}_n)}{\partial \lambda} F_n' \sqrt{\frac{h_n}{n}} F_n g_n(\lambda_0) \xrightarrow{d} N(0, \Sigma_\lambda). \qquad \square$$

Proof of Theorem 6. It shall be shown that $\frac{h_n}{n} F_n g_n(\lambda) - E(\frac{h_n}{n} F_n g_n(\lambda)) = o_P(1)$ and $E(\frac{h_n}{n} F_n g_n(\lambda))$ is uniformly bounded, uniformly in λ in Λ.

As $\varepsilon_{x,n}(\lambda) = (\lambda_0 - \lambda) M_n G_n X_n \beta_0 + M_n (I_n + (\lambda_0 - \lambda) G_n) \mathscr{E}_n$,

$$F_n g_n(\lambda) = (\lambda_0 - \lambda)^2 \sum_{l=1}^m F_{nl} (M_n G_n X_n \beta_0)' P_{ln} M_n G_n X_n \beta_0 + (\lambda_0 - \lambda) l_n(\lambda) + q_n(\lambda),$$

where $F_n = (F_{n1}, \ldots, F_{nm})$, $l_n(\lambda) = \sum_{l=1}^{m} F_{nl}(M_n G_n X_n \beta_0)' P^s_{ln} M_n (I_n + (\lambda_0 - \lambda) G_n) \mathscr{E}_n$ and

$$q_n(\lambda) = \sum_{l=1}^{m} F_{nl} \mathscr{E}'_n (I_n + (\lambda_0 - \lambda) G_n)' M_n P_{ln} M_n (I_n + (\lambda_0 - \lambda) G_n) \mathscr{E}_n.$$

Under Assumption 8″, Lemma B.7 implies that $\sqrt{\frac{h_n}{n}} l_n(\lambda) = o_P(1)$ uniformly in $\lambda \in \Lambda$. Lemma B.8 implies that $\frac{h_n}{n} q_n(\lambda) - E(\frac{h_n}{n} q_n(\lambda)) = o_P(1)$ uniformly in λ in Λ. Hence, $\frac{h_n}{n} F_n g_n(\lambda) - E(\frac{h_n}{n} F_n g_n(\lambda)) = o_P(1)$ uniformly in Λ. The equicontinuity of $\frac{h_n}{n} E(F_n g_n(\lambda))$ on Λ is apparent from its expression.

Note that

$$\begin{aligned} E(q_n(\lambda)) &= \sigma_0^2 \sum_{l=1}^{m} F_{nl} \{ \mathrm{tr}[M_n P_{ln}] + (\lambda_0 - \lambda) \mathrm{tr}[G'_n M_n P^s_{ln} M_n] + (\lambda_0 - \lambda)^2 \mathrm{tr}[G'_n M_n P_{ln} M_n G_n] \} \\ &= \sigma_0^2 \left\{ (\lambda_0 - \lambda) \sum_{l=1}^{m} F_{nl} \mathrm{tr}(G'_n P^s_{ln}) + (\lambda_0 - \lambda)^2 \sum_{l=1}^{m} F_{nl} \mathrm{tr}(P_{ln} G_n G'_n) \right\} + O(1), \end{aligned}$$

by Lemmas B.4 and B.5 and $\mathrm{tr}(P_{ln}) = 0$. Assumption 8″ implies also $\frac{h_n}{n}(M_n G_n X_n \beta_0)' P_n M_n G_n X_n \beta_0 = 2\frac{h_n}{n}(M_n G_n X_n \beta_0)' P^s_n M_n G_n X_n \beta_0 = o(1)$ because the eigenvalues of P^s_n are uniformly bounded as P^s_n is uniformly bounded in absolute value in row and column sums. Thus, $\frac{h_n}{n} E(F_n g_n(\lambda)) = \frac{h_n}{n} E(F_n q_n(\lambda)) + o(1)$.

The identification uniqueness condition of λ_0 in this case is

$$\liminf_{n \to \infty} \min_{\lambda \in \bar{N}_\varepsilon(\lambda_0)} E\left(\frac{h_n}{n} F_n q_n(\lambda) \right)' E\left(\frac{h_n}{n} F_n q_n(\lambda) \right) > 0,$$

which follows from the identification conditions in Assumption 10. The consistency follows from uniform convergence and identification uniqueness (White, 1994, Theorem 3.4).

By Lemmas B.7 and B.8 and $\frac{h_n}{n}(M_n G_n X_n \beta_0)' P^s_n M_n G_n X_n \beta_0 = o(1)$,

$$\begin{aligned} &\frac{h_n}{n} \varepsilon'_{x,n}(\lambda) P^s_n M_n G_n (X_n \beta_0 + \mathscr{E}_n) \\ &\quad = (\lambda_0 - \lambda) \frac{h_n}{n} [M_n G_n (X_n \beta_0 + \mathscr{E}_n)]' P^s_n M_n G_n (X_n \beta_0 + \mathscr{E}_n) + \frac{h_n}{n} \mathscr{E}'_n M_n P^s_n M_n G_n (X_n \beta_0 + \mathscr{E}_n) \\ &\quad = (\lambda_0 - \lambda) \frac{h_n}{n} \sigma_0^2 \mathrm{tr}(P^s_n G_n G'_n) + \sigma_0^2 \frac{h_n}{n} \mathrm{tr}(P^s_n G_n) + o_P(1). \end{aligned}$$

Hence, $\frac{h_n}{n} \frac{\partial g_n(\bar{\lambda}_n)}{\partial \lambda} = -\frac{h_n}{n} (P^s_{1n} \varepsilon_{x,n}(\bar{\lambda}_n), \ldots, P^s_{mn} \varepsilon_{x,n}(\bar{\lambda}_n), Q_{n1})' M_n G_n (X_n \beta_0 + \mathscr{E}_n) = -\sigma_0^2 \frac{h_n}{n} C_{mn} + o_P(1)$, for any consistent estimate $\bar{\lambda}_n$ of λ_0.

The $\sqrt{\frac{h_n}{n}} \mathscr{E}'_n P^s_n X_n (X'_n X_n)^{-1} X'_n \mathscr{E}_n$ and $\sqrt{\frac{h_n}{n}} \mathscr{E}'_n X_n (X'_n X_n)^{-1} X'_n P_n X_n (X'_n X_n)^{-1} X'_n \mathscr{E}_n$ are of $O_P(\sqrt{\frac{h_n}{n}})$. Hence, $\sqrt{\frac{h_n}{n}} \mathscr{E}'_n M_n P_n M_n \mathscr{E}_n = \sqrt{\frac{h_n}{n}} \mathscr{E}'_n P_n \mathscr{E}_n + o_P(1)$. By the central limit theorem in Lemma B.9,

$$\sqrt{\frac{h_n}{n}} g_n(\lambda_0) = \sqrt{\frac{h_n}{n}} (\mathscr{E}'_n P_{1n} \mathscr{E}_n, \ldots, \mathscr{E}'_n P_{mn} \mathscr{E}_n)' + o_P(1) \xrightarrow{d} N\left(0, \lim_{n \to \infty} \frac{h_n}{n} \Omega_n \right),$$

where Ω_n is in (6.3). The asymptotic distribution of $\hat{\lambda}_n$ follows from the Taylor expansion:

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = \left(\frac{h_n}{n}\frac{\partial g_n'(\hat{\lambda}_n)}{\partial \lambda}F_n'F_n\frac{h_n}{n}\frac{\partial g_n(\bar{\lambda}_n)}{\partial \lambda}\right)^{-1}\frac{h_n}{n}\frac{\partial g_n'(\hat{\lambda}_n)}{\partial \lambda}F_n'F_n \times \sqrt{\frac{h_n}{n}}g_n(\lambda_0) \xrightarrow{d} \mathrm{N}(0, \Sigma_\lambda). \qquad \square$$

Appendix D. Sequential two-stage least squares estimation

The 2SLS approach in (2.3) with IV matrix $Q_n = (Q_{n1}, X_n)$ can be regarded as a special case of the optimum GMM estimation with the moments $Q_n'(Y_n - Z_n\lambda - X_n\beta)$, where $Z_n = W_nY_n$, and the distance matrix $(\sigma_0^2 Q_n'Q_n)^{-1}$ (or simply $(Q_n'Q_n)^{-1}$).

Correspondingly, λ and β can also be estimated by the method of elimination and substitution. This method solves $\hat{\beta}(\lambda)$ from $X_n'(y_n - Z_n\lambda - X_n\hat{\beta}(\lambda)) = 0$ for any λ. By substitution, the remaining moments are $Q_{n1}'(y_n - Z_n\lambda - X_n\hat{\beta}(\lambda)) = Q_{n1}'M_n(y_n - Z_n\lambda)$. At $\lambda = \lambda_0$,

$$Q_{n1}'M_n(Y_n - Z_n\lambda_0) = Q_{n1}'M_n(Y_n - Z_n\lambda_0 - X_n\beta_0) = Q_{n1}'M_n\mathscr{E}_n,$$

which has zero mean and its variance matrix is $\sigma_0^2 Q_{n1}'M_nQ_{n1}$. Thus, the optimum distance in the GMM estimation using $Q_{n1}'M_n(y_n - Z_n\lambda)$ shall be $(Q_{n1}'M_nQ_{n1})^{-1}$. The modified optimum GMM estimation of λ is

$$\min_{\lambda}(Y_n - Z_n\lambda)'(M_nQ_{n1})_{(\mathrm{p})}(Y_n - Z_n\lambda). \tag{D.1}$$

The modified estimator $\tilde{\lambda}_n$ from (D.1) is

$$\tilde{\lambda}_n = (Z_n'(M_nQ_{n1})_{(\mathrm{p})}Z_n)^{-1}Z_n'(M_nQ_{n1})_{(\mathrm{p})}Y_n, \tag{D.2}$$

and the corresponding estimator of β is

$$\tilde{\beta}_n = (X_n'X_n)^{-1}X_n'[I_n - Z_n(Z_n'(M_nQ_{n1})_{(\mathrm{p})}Z_n)^{-1}Z_n'(M_nQ_{n1})_{(\mathrm{p})}]Y_n. \tag{D.3}$$

When $Q_n = (Q_{n1}, X_n)$, $(Q_n)_{(\mathrm{p})}X_n = X_n$ and, hence, the joint 2SLS estimator in (2.3) can be rewritten as

$$\begin{pmatrix} \hat{\lambda}_n \\ \hat{\beta}_n \end{pmatrix} = \begin{pmatrix} Z_n'(Q_n)_{(\mathrm{p})}Z_n & Z_n'X_n \\ X_n'Z_n & X_n'X_n \end{pmatrix}^{-1}\begin{pmatrix} Z_n'(Q_n)_{(\mathrm{p})}Y_n \\ X_n'Y_n \end{pmatrix}.$$

Let $R_n = Z_n'(Q_n)_{(\mathrm{p})}Z_n - Z_n'X_n(X_n'X_n)^{-1}X_n'Z_n = Z_n'((Q_n)_{(\mathrm{p})} - X_n(X_n'X_n)^{-1}X_n')Z_n$. By the inverse formula of a partitioned matrix,

$$\begin{pmatrix} Z_n'(Q_n)_{(\mathrm{p})}Z_n & Z_n'X_n \\ X_n'Z_n & X_n'X_n \end{pmatrix}^{-1} = \begin{pmatrix} R_n^{-1} & -R_n^{-1}Z_n'X_n(X_n'X_n)^{-1} \\ -(X_n'X_n)^{-1}X_n'Z_nR_n^{-1} & (X_n'X_n)^{-1} + (X_n'X_n)^{-1}X_n'Z_nR_n^{-1}Z_n'X_n(X_n'X_n)^{-1} \end{pmatrix}.$$

As $Q_n = (X_n, Q_{n1})$, $(Q_n)_{(p)} - X_n(X_n'X_n)^{-1}X_n' = (M_nQ_{n1})_{(p)}$ and, hence, $R_n = Z_n'(M_nQ_{n1})_{(p)}Z_n$. It follows that

$$\hat{\lambda}_n = R_n^{-1}(Z_n'(Q_n)_{(p)}Y_n - Z_n'X_n(X_n'X_n)^{-1}X_n'Y_n) = R_n^{-1}Z_n'(M_nQ_{n1})_{(p)}Y_n$$

and

$$\begin{aligned}\hat{\beta}_n = & -(X_n'X_n)^{-1}X_n'Z_nR_n^{-1}Z_n'(Q_n)_{(p)}Y_n + (X_n'X_n)^{-1}X_n'Y_n \\ & + (X_n'X_n)^{-1}X_n'Z_nR_n^{-1}Z_n'X_n(X_n'X_n)^{-1}X_n'Y_n \\ = & (X_n'X_n)^{-1}X_n'[I_n - Z_nR_n^{-1}Z_n'(M_nQ_{n1})_{(p)}]Y_n,\end{aligned}$$

which are, respectively, numerically identical to the sequential estimators in (D.2) and (D.3).

References

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, Massachusetts.
Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, The Netherlands.
Case, A.C., 1991. Spatial patterns in household demand. Econometrica 59, 953–965.
Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. Pion Limited, London.
Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.
Horn, R., Johnson, C., 1985. Matrix Analysis. Cambridge University Press, New York.
Kelejian, H.H., Prucha, I.R., 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., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.
Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219–257.
Lee, L.F., 2001a. Generalized method of moments estimation of spatial autoregressive processes. Manuscript, Department of Economics, Ohio State University, Columbus, OH.
Lee, L.F., 2001b. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Manuscript, Department of Economics, Ohio State University, Columbus, OH.
Lee, L.F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252–277.
Lee, L.F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22, 307–335.
Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1926.
Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam (Chapter 36).
Ogaki, M., 1993. Generalized method of moments: econometric applications. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Econometrics Handbook of Statistics, vol. 11. North-Holland, Amsterdam (Chapter 17).
Ord, J., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120–126.
Pace, R.K., Barry, R., 1997. Sparse spatial autoregressions. Statistics and Probability Letters 33, 291–297.
Ruud, P.A., 2000. An Introduction to Classical Econometric Theory. Oxford University Press, Oxford.
Smirnov, O., Anselin, L., 2001. Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. Computational Statistics and Data Analysis 35, 301–319.
White, H., 1994. Estimation Inference and Specification Analysis. Cambridge University Press, New York.