Efficient GMM Estimation of High Order Spatial Autoregressive Models with Autoregressive Disturbances*

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

Xiaodong Liu
Department of Economics
University of Colorado at Boulder
Boulder, CO 80309, USA
Email: xiaodong.liu@colorado.edu

First version, October 2006
Current revision, September 2008

*This version is an extension which generalizes and unifies two previous working papers with the titles "Efficient GMM Estimation of High Order Spatial Autoregressive Models" and "Efficient GMM Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances".

†We gratefully acknowledge financial support from the National Science Foundation Grant No. SES-0519204 and research assistantship support from the Department of Economics at the Ohio State University. We appreciate having received valuable comments and suggestions from three anonymous referees, an associate editor, and the editor to improve the presentation of our paper. Any remaining errors are solely ours.

Running Head: High Order Spatial Autoregressive Models

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

Abstract: In this paper, we extend the GMM framework for the estimation of the mixed-regressive spatial autoregressive model by Lee (2007a) to estimate a high order mixed-regressive spatial autoregressive model with spatial autoregressive disturbances. Identification of such a general model is considered. The GMM approach has computational advantage over the conventional ML method. The proposed GMM estimators are shown to be consistent and asymptotically normal. The best GMM estimator is derived, within the class of GMM estimators based on linear and quadratic moment conditions of the disturbances. The best GMM estimator is asymptotically as efficient as the ML estimator under normality, more efficient than the QML estimator otherwise, and is efficient relative to the G2SLS estimator.

Key Words: Spatial econometrics, spatial autoregressive models, regression, GMM, QML, efficiency.

JEL Classification: C13, C21, R15.

1 Introduction

The spatial autoregressive (SAR) model with high order spatial lags can characterize spatial interdependence based on different types of relationships (e.g. geographic distance, social relation) among cross-sectional units. In this paper, we consider the estimation of a general high order SAR model with SAR disturbances.

For the estimation of a SAR model with a first order spatial lag, the conventional estimation method would be the quasi-maximum likelihood (QML) (Anselin, 1988). In addition to that, alternative estimation methods have also been proposed. In the presence of exogenous variables, the SAR model is known as a mixed regressive, spatial autoregressive (MRSAR) model. For the first order MRSAR model with SAR disturbances, Kelejian and Prucha (1998) introduced a general spatial two-stage least squares (G2SLS) estimator that is consistent and asymptotically normal. Lee (2003) discussed the best instrumental variables (IVs) selection in the last step of the G2SLS procedure, and suggested the best one within the class of IV estimators. To further simplify the computation involved in the best G2SLS estimator, Kelejian et al. (2004) suggested a series-type best G2SLS estimator that is asymptotically equivalent to Lee's (2003) estimator. Kelejian and Prucha (2007a) considered the IV estimation of the first order MRSAR model allowing the disturbance process for general patterns of correlation and heteroskedasticity, and proposed a spatial heteroskedasticity and autocorrelation consistent (HAC) estimator for the variance–covariance (VC) matrix of the IV estimator.

The various IV or G2SLS estimators have the virtue of computational simplicity but they are inefficient relative to the ML estimator, when the disturbances are normally distributed so that the likelihood function is correctly specified. Also, as the IVs are functions of the spatial weights matrices and exogenous variables, the G2SLS method would not be applicable when all exogenous variables in a model are really irrelevant. Lee (2001; 2007a) proposed a systematic generalized method of moments (GMM) framework for the estimation of the first order SAR models, with or without exogenous regressors. The GMM approach combines the IV estimation with a generalization of the method of moments (MOM) in Kelejian and Prucha (1999) which has been proposed for the estimation of SAR disturbances in a regression model. That GMM approach is computationally more complicated than the G2SLS but is simpler than the QML. The GMM estimator is asymptotically efficient relative to the G2SLS estimator, and with proper moment equations, it can be asymptotically as efficient as the ML estimator with normally distributed disturbances.

In this paper, we extend the GMM approach to estimate the (MR)SAR model with general finite order spatial lags and SAR disturbances of a finite order. High order SAR models have been specified in Blommestein (1983; 1985), Huang (1984) and some others (see Anselin and Bera, 1998). The multiple spatial weights matrices may capture contiguity of units in various dimensions. For example, in Tao's (2005) strategic interaction model of local school expenditure, two spatial weights matrices are specified — one based on geographical contiguity and the other based on economic similarity. An alternative perspective stated in Anselin and Bera (1998, p.252) on the need for high order models is to consider them as alternatives of a poorly specified weights matrix rather than as a realistic data generating process. For this general model with high order spatial lags and disturbances, the QML approach is not practical and may be, in general, infeasible as the parameter space is quite complex and the Jacobian determinant in the log likelihood function can not be easily evaluated. The IV and G2SLS estimation approaches are still feasible. For instance, Kelejian and Prucha (2004) proposed the G2SLS estimation for the spatial simultaneous equation model, where a structural equation may have spatial lags of several endogenous variables on the right hand side. Also, Kelejian and Prucha (2007a) considered the G2SLS estimation of a structural equation with spatial lags and endogenous regressors where general patterns of spatial correlation and heteroskedasticity are allowed for the disturbance. The VC matrix of the G2SLS estimate can be consistently estimated with their proposed HAC estimator. With carefully designed quadratic moment equations, the GMM approach can be robust against unknown heteroskedasticity (see Kelejian and Prucha, 2007b; Lin and Lee, 2006). In this paper, we are interested in efficient estimators instead of robust ones. So we will focus on the model with homoskedastic disturbances. Under the homoskedasticity assumption, while the G2SLS estimation approach is feasible, it would not be asymptotically efficient. We study the identification of the model with homoskedastic disturbances and the asymptotic properties of the proposed GMM estimator. We discuss the selection of the best moment conditions without any specific distributional assumption, and suggest the best GMM (BGMM) estimator within the class of GMM estimators derived from linear and quadratic moment conditions.[1] As the GMM objective function is a polynomial of unknown parameters, constraints on parameters are not necessary and the BGMM is computationally tractable. Furthermore, the BGMM estimator is asymptotically as efficient as the ML estimator under normality, and more efficient than the QML estimator otherwise. It is also efficient relative to the best G2SLS estimator.

We conduct a Monte Carlo experiment to study the finite sample performance of the proposed GMM estimator. We find that the GMM estimator of the spatial effects have smaller bias and standard deviation than those of the G2SLS and B2SLS when the variation from the exogenous regressors relative to that of the disturbances is small. When the disturbances are asymmetrically distributed, the proposed BGMM improves upon the QML and B2SLS, and the improvement could be as large as 20% in terms of reduction in the standard deviation. The GMM estimators are also relatively robust to the misspecified order of spatial lags.

This paper is organized as follows. In Section 2, we introduce the high order MRSAR model with SAR disturbances. Section 3 discusses the existing estimators for this model. We establish identification of the model and propose a GMM estimation approach in Section 4. Section 5 investigates consistency and asymptotic distribution of the GMM estimators. Section 6 derives the best selection of moment functions and discusses the efficiency properties of the BGMM estimator. Section 7 provides some Monte Carlo results of finite sample properties of estimators. Section 8 concludes. All the proofs of the results are collected in the appendices.

2 The MRSAR Model with SAR Disturbances

We consider a general p-order MRSAR model with q-order SAR disturbances (for short, SARAR(p,q))

$$Y_n = \sum_{j=1}^{p} \lambda_j W_{jn} Y_n + X_n \beta + u_n, \qquad u_n = \sum_{k=1}^{q} \rho_k M_{kn} u_n + \epsilon_n, \tag{1}$$

where n is the total number of spatial units, X_n is an $n \times k_x$ dimensional matrix of nonstochastic exogenous variables, and the elements $\epsilon_{n1}, \cdots, \epsilon_{nn}$ of the n-dimensional vector ϵ_n are i.i.d. $(0, \sigma^2)$. $W_{1n}, \cdots, W_{pn}$ and $M_{1n}, \cdots, M_{qn}$ are $n \times n$ dimensional spatial weights matrices of known constants such that $W_{j_1 n} \neq W_{j_2 n}$ if $j_1 \neq j_2$ and $M_{k_1 n} \neq M_{k_2 n}$ if $k_1 \neq k_2$. However, W_{jn} and M_{kn} may or may not be the same for $j = 1, \cdots, p$ and $k = 1, \cdots, q$. The model (1) incorporates both high order spatial lags $W_{1n}Y_n, \cdots, W_{pn}Y_n$ and spatial correlated disturbances u_n.[2]

With a given contiguity-based spatial weights matrix W_n, it seems straightforward to define high order spatial lags operators as powers of W_n motivated as in time series. The corresponding SAR(p)

model would be $Y_n = \sum_{j=1}^{p} \lambda_j W_n^j Y_n + X_n\beta + \epsilon_n$. As emphasized in Blommestein (1985), powering W_n may result in the presence of circular and redundant routes. Proper high-order lag operators should have those circular and redundant routes eliminated. Algorithms have been introduced in Blommestein and Koper (1992) and Anselin and Smirnov (1996) to construct proper high-order lag operators. Such models can be regarded as special cases in our model framework. In general, our framework allows the several spatial matrices as (proper) high-order spatial lag operators generated from a contiguity-based spatial weights matrix but may not be so restricted.

Let $\rho = (\rho_1, \cdots, \rho_q)'$, $\lambda = (\lambda_1, \cdots, \lambda_p)'$, and $\theta = (\rho', \lambda', \beta')'$. In order to distinguish the true parameters from other possible values in the parameter space, $\theta_0 = (\rho_0', \lambda_0', \beta_0')'$ and σ_0^2 denote the true parameters. Denote $S_n(\lambda) = I_n - \sum_{j=1}^{p} \lambda_j W_{jn}$ and $R_n(\rho) = I_n - \sum_{k=1}^{q} \rho_k M_{kn}$. At θ_0, let $S_n = S_n(\lambda_0)$ and $R_n = R_n(\rho_0)$ for simplicity. (A list of special notations used for this paper has been collected in Appendix A for convenient reference.) This model is an equilibrium model so that S_n and R_n are invertible.[3] The reduced form equation of (1) is $Y_n = S_n^{-1}X_n\beta_0 + S_n^{-1}R_n^{-1}\epsilon_n$. Furthermore, let $G_{jn} = W_{jn}S_n^{-1}$, which provide the representations $W_{jn}Y_n = G_{jn}X_n\beta_0 + G_{jn}R_n^{-1}\epsilon_n$ for $j = 1, \cdots, p$. $W_{jn}Y_n$ is correlated with ϵ_n because, in general, $E((G_{jn}R_n^{-1}\epsilon_n)'\epsilon_n) = \sigma_0^2 tr(G_{jn}R_n^{-1}) \neq 0$. In most cases, these correlations rule out the ordinary least squares (OLS) for the estimation of (1).[4]

3 Existing Estimators

From (1), if ϵ_n is $N(0, \sigma_0^2 I_n)$, the log likelihood function of this model is

$$\begin{aligned} \ln L_n &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \\ &\quad -\frac{1}{2\sigma^2}[S_n(\lambda)Y_n - X_n\beta]'R_n'(\rho)R_n(\rho)[S_n(\lambda)Y_n - X_n\beta]. \end{aligned} \tag{2}$$

To guarantee the log likelihood function is well defined, we only consider the parameter space of λ and ρ such that the determinants of $S_n(\lambda)$ and $R_n(\rho)$ are strictly positive, i.e., $|S_n(\lambda)| > 0$ and $|R_n(\rho)| > 0$. Let $||\cdot||$ be any matrix norm. One has $||\sum_{j=1}^{p} \lambda_j W_{jn}|| \leq (\sum_{j=1}^{p} |\lambda_j|) \cdot \max_{j=1,\cdots,p} ||W_{jn}||$. Hence, when all W_{jn} are row-normalized such that $||W_{jn}||_\infty = 1$ for $j = 1, \cdots, p$,[5] a possible parameter space for λ can be one satisfying $\sum_{j=1}^{p} |\lambda_j| < 1$. In the event that the spatial weights matrices are not row-normalized, then the parameter space may be taken to be $\sum_{j=1}^{p} |\lambda_j| < (\max_{j=1,\cdots p} ||W_{jn}||)^{-1}$. The parameter space of ρ can be analogously obtained.

Even with the parameter space imposed, the ML method is still computationally cumbersome as $|S_n(\lambda)|$ and $|R_n(\rho)|$ are difficult to evaluate.[6] Therefore, it seems of interest to have available alternative efficient estimation methods which do not involve the complicated parameter space and computation of $|S_n(\lambda)|$ and $|R_n(\rho)|$. Towards this end, we develop the BGMM estimator in this paper.

For the estimation of the SARAR(1,1), Kelejian and Prucha (1998) suggested a feasible G2SLS (FG2SLS) estimation method. With a consistent initial estimator $\hat{\rho}_n$ for ρ_0, the FG2SLS of δ_0 is defined as

$$\hat{\delta}_{g2sls,n} = [Z_n' R_n'(\hat{\rho}_n) Q_n (Q_n' Q_n)^{-1} Q_n' R_n(\hat{\rho}_n) Z_n]^{-1} Z_n' R_n'(\hat{\rho}_n) Q_n (Q_n' Q_n)^{-1} Q_n' R_n(\hat{\rho}_n) Y_n, \quad (3)$$

where $Z_n = (W_n Y_n, X_n)$ and Q_n is a matrix of IVs. Kelejian and Prucha (1998) suggested Q_n to be taken as a fixed subset of the linearly independent columns of $\{X_n, W_n X_n, W_n^2 X_n, \cdots, W_n^q X_n, M_n X_n, M_n W_n X_n, M_n W_n^2 X_n, \cdots, M_n W_n^q X_n\}$, where q is a pre-selected positive integer and the subset is required to contain at least the linearly independent columns of $\{X_n, M_n X_n\}$. The FG2SLS estimator has a closed form expression and is computationally the most simple. Lee (2003) discussed the selection of IVs and proposed the best FG2SLS estimator with $Q_n = R_n(\hat{\rho}_n)[G_n(\hat{\lambda}_n) X_n \hat{\beta}_n, X_n]$, where $G_n(\lambda) = W_n S_n^{-1}(\lambda)$.[7] As the various G2SLS estimators use functions of W_n and X_n as IVs, the G2SLS would not be applicable when all exogenous variables in X_n are really irrelevant. Another unsatisfactory feature of the G2SLS estimator is that the asymptotic distribution of $\hat{\delta}_{g2sls,n}$ does not depend on the asymptotic distribution of $\hat{\rho}_n$ (see Kelejian and Prucha, 1999; Lee, 2003).[8] In a time series model with lagged dependent variables and autoregressive disturbances, $y_t = \lambda_0 y_{t-1} + x_t \beta_0 + u_t$ with $u_t = \rho_0 u_{t-1} + \epsilon_t$, it is known that a feasible GLS estimation of λ_0 and β_0 based on the transformed equation $y_t - \hat{\rho}_n y_{t-1} = \lambda(y_{t-1} - \hat{\rho}_n y_{t-2}) + (x_t - \hat{\rho}_n x_{t-1})\beta + \hat{\epsilon}_t$ is not efficient (Maddala, 1971). The SARAR(1,1) includes this dynamic time series model as a special case.[9] With normal disturbances, the MLEs of ρ_0 and λ_0 are asymptotically correlated (e.g. Anselin and Bera, 1998), which suggests potential inefficiency of the G2SLS. We suggest the GMM approach which estimates ρ_0 and λ_0 simultaneously using quadratic moments in addition to the linear moments used in the G2SLS or the best G2SLS. With properly constructed moments, We show the GMM can be asymptotically more efficient than the best G2SLS.

4 GMM Estimation

The GMM method in its general setting is based on an $n \times k_{IV}$ IV matrix Q_n, and the IV functions $P_{in}\epsilon_n(\theta)$ where P_{in} is a $n \times n$ square (constant) matrix with $tr(P_{in}) = 0$ for $i = 1, \cdots, m$. Let $u_n(\delta) = S_n(\lambda)Y_n - X_n\beta$ and $\epsilon_n(\theta) = R_n(\rho)u_n(\delta)$, where $\delta = (\lambda', \beta')'$. The GMM estimation uses the following empirical moments[10]

$$\begin{aligned} g_n(\theta) &= [Q_n, P_{1n}\epsilon_n(\theta), \cdots, P_{mn}\epsilon_n(\theta)]'\epsilon_n(\theta) \\ &= [Q_n, P_{1n}R_n(\rho)u_n(\delta), \cdots, P_{mn}R_n(\rho)u_n(\delta)]'R_n(\rho)u_n(\delta), \end{aligned} \tag{4}$$

where $E(g_n(\theta_0)) = E[(Q_n, P_{1n}\epsilon_n, \cdots, P_{mn}\epsilon_n)'\epsilon_n] = 0$, because $E(Q_n'\epsilon_n) = Q_n'E(\epsilon_n) = 0$ and $E(\epsilon_n'P_{in}\epsilon_n) = \sigma_0^2 tr(P_{in}) = 0$ for $i = 1, \cdots, m$.[11] In a practical application, one has to select specific Q_n and P_{in}'s to implement the method. As a simple example, for the SARAR(1,1) model, Q_n may consist of X_n, W_nX_n and M_nX_n; and P_{1n} and P_{2n} are, respectively, W_n and M_n, where W_n and M_n have zero diagonals.[12] The general but arbitrary set of linear and quadratic moment conditions provides a framework to discuss the possible selection of best moment conditions.

There are two motivations to use quadratic moments in addition to linear moments for the GMM estimation. As will be shown below, one motivation is that the score vector of the likelihood function essentially consists of linear combinations of linear and quadratic moments functions. Another rationale is by the construction of IVs for the estimation of δ_0. Consider the MRSAR model $Y_n = \lambda_0 W_nY_n + X_n\beta_0 + \epsilon_n$ for an illustration. As $W_nY_n = G_nX_n\beta_0 + G_n\epsilon_n$, an IV for W_nY_n may be a function of exogenous variables that approximates $G_nX_n\beta_0$, the deterministic component of W_nY_n. This motivates the use of linear moments. The quadratic moments are motivated by using the instrumental function $P_n\epsilon_n$ which should be correlated with $G_n\epsilon_n$, the stochastic component of W_nY_n, but uncorrelated with ϵ_n.

5 Consistency and Asymptotic Distributions

To proceed, we follow the regularity assumptions in Lee (2007a) with proper modifications to fit in the current model.

Assumption 1 *The ϵ_{ni}'s are i.i.d. with zero mean, variance σ_0^2 and that a moment of order higher than the fourth exists.*

Assumption 2 *The elements of X_n are uniformly bounded constants, X_n has full column rank k_x, and $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular.*

Assumption 3 *The zero diagonal spatial weights matrices $\{W_{jn}\}$, $\{M_{kn}\}$ ($j = 1, \cdots, p$, $k = 1, \cdots, q$) and the corresponding $\{S_n^{-1}\}$, $\{R_n^{-1}\}$ are uniformly bounded in both row and column sums in absolute value.*[13]

Assumption 4 *The matrices P_{in}'s with $tr(P_{in}) = 0$, for $i = 1, \cdots, m$, are uniformly bounded in both row and column sums in absolute value, and elements of Q_n are uniformly bounded.*

The disturbances in Assumption 1 are in the form of triangular arrays for generality. It includes the case that $\epsilon_{ni} = \epsilon_i$, independent of the sample size n. The higher than the fourth moment condition in Assumption 1 is needed in order to apply the central limit theorem of Kelejian and Prucha (2001) for triangular arrays of random variables. The nonstochastic X_n and its uniform boundedness conditions in Assumption 2 are for analytical simplicity. The elements of X_n as well as those of W_{jn}'s and M_{kn}'s, in their generality, may depend on n too. Assumption 3 limits the spatial dependence among the units to a tractable degree and is originated by Kelejian and Prucha (1999). It rules out the unit root case (in time series as a special case). $||\sum_{j=1}^{p} \lambda_{0j} W_{jn}||_\infty < 1$ if $(\sum_{j=1}^{p} |\lambda_{0j}|) \max_{j=1,\cdots,p} \|W_{jn}\|_\infty < 1$. A sufficient condition for S_n^{-1} to be uniformly bounded in row sums in absolute value is that $\sum_{j=1}^{p} |\lambda_{0j}| < 1/\max_{j=1,\cdots,p} \|W_{jn}\|_\infty$, because $S_n^{-1} = I_n + (\sum_{j=1}^{p} \lambda_{0j} W_{jn}) + (\sum_{j=1}^{p} \lambda_{0j} W_{jn})^2 + \cdots$. Similarly, S_n^{-1} is uniformly bounded in column sums in absolute value if $\sum_{j=1}^{p} |\lambda_{0j}| < 1/\max_{j=1,\cdots,p} \|W_{jn}\|_1$. With an analogous argument, R_n^{-1} is uniformly bounded in both row and column sums in absolute value if $\sum_{k=1}^{q} |\rho_{0k}| < 1/\max_{k=1,\cdots,q}\{\|M_{kn}\|_1, \|M_{kn}\|_\infty\}$. The uniform boundedness assumptions of both S_n^{-1} and R_n^{-1} in Assumption 3 are assumed to be valid at λ_0 and ρ_0. But with the uniform boundedness of W_{jn}'s and M_{kn}'s, $S_n^{-1}(\lambda)$ and $R_n^{-1}(\rho)$ will also be uniformly bounded, uniformly in a neighbor of λ_0 and ρ_0, respectively (Lee, 2004). The spatial weights matrices are assumed to have zero diagonals to facilitate the interpretation of a spatial effect and exclude self-influence. For analytical tractability, in Assumption 4, P_{in}'s are assumed to have the uniformly boundedness properties as the spatial weights matrices.

For any feasible θ, the model (1) implies that

$$E\left(g_n(\theta)\right) = \begin{pmatrix} Q_n' R_n(\rho) d_n(\delta) \\ d_n'(\delta) R_n'(\rho) P_{1n} R_n(\rho) d_n(\delta) + \sigma_0^2 tr[F_n^{'-1} F_n'(\rho,\lambda) P_{1n} F_n(\rho,\lambda) F_n^{-1}] \\ \vdots \\ d_n'(\delta) R_n'(\rho) P_{mn} R_n(\rho) d_n(\delta) + \sigma_0^2 tr[F_n^{'-1} F_n'(\rho,\lambda) P_{mn} F_n(\rho,\lambda) F_n^{-1}] \end{pmatrix}, \tag{5}$$

where $d_n(\delta) = \sum_{j=1}^{p} (\lambda_{0j} - \lambda_j) G_{jn} X_n \beta_0 + X_n(\beta_0 - \beta)$, $F_n(\rho,\lambda) = R_n(\rho) S_n(\lambda)$ and $F_n = F_n(\rho_0, \lambda_0)$.[14] Let $\Lambda_n = (G_{1n} X_n \beta_0, \cdots, G_{pn} X_n \beta_0, X_n)$.

Assumption 5 *Either (i)* $\lim_{n\to\infty} \frac{1}{n} Q_n' R_n(\rho) \Lambda_n$ *has full rank* $(p + k_x)$ *for each possible* ρ *in its parameter space, and the moment equations*

$$tr[R_n^{'-1} R_n'(\rho) P_{in} R_n(\rho) R_n^{-1}] = 0, \tag{6}$$

for $i = 1, \cdots, m$, *have the unique solution at* ρ_0, *or (ii)* $\lim_{n\to\infty} \frac{1}{n} Q_n' R_n(\rho) \Lambda_n$ *has column rank* $(p + k_x - p_0)$ *for some* $1 \le p_0 \le p$ *for each possible* ρ *in its parameter space, and the moment equations*

$$tr[F_n^{'-1} F_n'(\rho,\lambda) P_{in} F_n(\rho,\lambda) F_n^{-1}] = 0, \tag{7}$$

for $i = 1, \cdots, m$, *have the unique solution at the true parameter values.*

Assumption 5 summarizes some sufficient conditions for the identification of θ_0. We provide identification conditions for the moment equations (6) and (7) in Propositions 7 and 8 in Appendix B.

Proposition 1 *Under Assumptions 1-5,* $E\left(g_n(\theta)\right) = 0$ *has a unique solution at* $\theta = \theta_0$.

The moment conditions (7) correspond to those of a pure SARAR(p, q) process,

$$Y_n = \sum\nolimits_{j=1}^{p} \lambda_{0j} W_{jn} Y_n + u_n, \qquad u_n = \sum\nolimits_{k=1}^{q} \rho_{0k} M_{kn} u_n + \epsilon_n. \tag{8}$$

For this process, $\epsilon_n(\theta) = F_n(\rho,\lambda) F_n^{-1} \epsilon_n$ and, hence, $E[\epsilon_n'(\theta) P_{in} \epsilon_n(\theta)] = \sigma_0^2 tr[F_n^{'-1} F_n'(\rho,\lambda) P_{in} F_n(\rho,\lambda) F_n^{-1}]$ for $i = 1, \cdots, m$. This pure SAR process implies the transformed process $Y_n = \sum_{k=1}^{q} \rho_{0k} M_{kn} Y_n + \sum_{j=1}^{p} \lambda_{0j} W_{jn} Y_n - \sum_{j=1}^{p} \sum_{k=1}^{q} \rho_{0k} \lambda_{0j} M_{kn} W_{jn} Y_n + \epsilon_n$. For the pure SAR process with $p = q$, identification of ρ_0 and λ_0 separately would not be possible if $W_{jn} = M_{jn}$ for $j = 1, \cdots, p$.

This is because the transformed equation would be reduced to $Y_n = \sum_{j=1}^{p}(\rho_{0j} + \lambda_{0j})W_{jn}Y_n - \sum_{j=1}^{p}\sum_{k=1}^{p}\rho_{0k}\lambda_{0j}W_{kn}W_{jn}Y_n + \epsilon_n$, and, hence, ρ_0 and λ_0 would not be distinguished from each other.[15]

Let $\Omega_n = var\left(g_n\left(\theta_0\right)\right)$. Ω_n involves variances and covariances of linear and quadratic forms of ϵ_n. For any square matrix A, $vec_D(A) = (a_{11}, \cdots, a_{nn})'$ is the column vector formed with the diagonal elements of A, and $A^s = A + A'$. It follows from Lee (2007a) that

$$\Omega_n = \begin{pmatrix} 0_{k_{IV}\times k_{IV}} & \mu_3 Q_n' \omega_{nm} \\ \mu_3 \omega_{nm}' Q_n & (\mu_4 - 3\sigma_0^4)\omega_{nm}'\omega_{nm} \end{pmatrix} + V_n, \tag{9}$$

with $\omega_{nm} = [vec_D(P_{1n}), \cdots, vec_D(P_{mn})]$, and

$$V_n = \sigma_0^4 \begin{pmatrix} \frac{1}{\sigma_0^2}Q_n'Q_n & 0_{k_{IV}\times 1} & \cdots & 0_{k_{IV}\times 1} \\ 0_{1\times k_{IV}} & tr(P_{1n}^s P_{1n}) & \cdots & tr(P_{1n}^s P_{mn}) \\ \vdots & \vdots & \ddots & \vdots \\ 0_{1\times k_{IV}} & tr(P_{mn}^s P_{1n}) & \cdots & tr(P_{mn}^s P_{mn}) \end{pmatrix} = \sigma_0^4 \begin{pmatrix} \frac{1}{\sigma_0^2}Q_n'Q_n & 0_{k_{IV}\times m} \\ 0_{m\times k_{IV}} & \Delta_{mn} \end{pmatrix}, \tag{10}$$

where $0_{k\times l}$ denote the zero matrix of dimension $k \times l$, μ_3 and μ_4 are, respectively, the third and fourth moments of ϵ_{ni}, and $\Delta_{mn} = [vec(P_{1n}^s), \cdots, vec(P_{mn}^s)]'[vec(P_{1n}), \cdots, vec(P_{mn})]$. When ϵ_n is normally distributed, Ω_n is simplified to V_n because $\mu_3 = 0$ and $\mu_4 = 3\sigma_0^4$. In general, Ω_n is nonsingular if and only if both matrices $(vec(P_{1n}), \cdots, vec(P_{mn}))$ and Q_n have full column ranks. As elements of P_{in}'s and Q_n are uniformly bounded by Assumption 4, and $P_{in}^s P_{jn}$ is bounded in row or column sums, $\frac{1}{n}\Omega_n = O(1)$. It is thus meaningful to impose the following conventional regularity condition on the limit of $\frac{1}{n}\Omega_n$:

Assumption 6 *The limit of $\frac{1}{n}\Omega_n$ exists and is a nonsingular matrix.*

The asymptotic analysis in this paper assumes each unit has only a finite (bounded) number of neighbors which does not increase as n increases. The spatial weights matrices may be sparse. Assumption 6 and parts of Assumption 5 provide the regular conditions for estimators to have the usual $\sqrt{n}$-rate of convergence.[16]

The following proposition provides the asymptotic distribution of a GMM estimator with a linear transformation of the moment equations, $a_n g_n(\theta)$, where a_n is a matrix with full row rank greater than or equal to the number of unknown parameters $(k_x + p + q)$. The a_n is assumed to converge to a constant matrix a_0 which has also full row rank. This corresponds to the Hansen's GMM setting, which illustrates the optimal weighting issue. As usual for nonlinear estimation, the parameter space Θ of θ will be taken to be a bounded set with θ_0 in its interior.[17]

Assumption 7 *The θ_0 is in the interior of the parameter space Θ, which is a bounded subset of R^{k_x+p+q}.*

Let

$$D_n = \frac{\partial E(g_n(\theta_0))}{\partial \theta'} \tag{11}$$
$$= -\begin{pmatrix} 0_{k_{IV}\times 1} & \cdots & 0_{k_{IV}\times 1} & Q_n'\bar{G}_{1n}R_nX_n\beta_0 & \cdots & Q_n'\bar{G}_{pn}R_nX_n\beta_0 & Q_n'R_nX_n \\ \sigma_0^2 tr(P_{1n}^s H_{1n}) & \cdots & \sigma_0^2 tr(P_{1n}^s H_{qn}) & \sigma_0^2 tr(P_{1n}^s \bar{G}_{1n}) & \cdots & \sigma_0^2 tr(P_{1n}^s \bar{G}_{pn}) & 0_{1\times k_x} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_0^2 tr(P_{mn}^s H_{1n}) & \cdots & \sigma_0^2 tr(P_{mn}^s H_{qn}) & \sigma_0^2 tr(P_{mn}^s \bar{G}_{1n}) & \cdots & \sigma_0^2 tr(P_{mn}^s \bar{G}_{pn}) & 0_{1\times k_x} \end{pmatrix},$$

where $\bar{G}_{jn} = R_nG_{jn}R_n^{-1}$ and $H_{kn} = M_{kn}R_n^{-1}$ for $j = 1, \cdots, p$ and $k = 1, \cdots, q$.[18]

Proposition 2 *Under Assumptions 1-7, suppose $g_n(\theta)$ is given by (4) so that $\lim_{n\to\infty} a_nE(g_n(\theta)) = 0$ has a unique root at θ_0 in Θ. Then, the GMM estimator $\hat{\theta}_n$ derived from $\min_{\theta\in\Theta} g_n'(\theta)a_n'a_ng_n(\theta)$ is a consistent estimator of θ_0, and $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \Sigma)$, where*

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

with D_n in (11) under the assumption that $\lim_{n\to\infty} \frac{1}{n}a_nD_n$ exists and has the full rank $(k_x + p + q)$.

From Proposition 2, with $g_n(\theta)$ in (4), the optimal choice of a weighting matrix $a_n'a_n$ is Ω_n^{-1} by the generalized Schwartz inequality. As Ω_n involves unknown parameters σ_0^2, μ_3 and μ_4, the optimal GMM objective function will be formulated with a two-step feasible approach by estimating consistently σ_0^2, as well as μ_3 and μ_4 in the first step. That can be done by using estimated residuals of ϵ_n from an initial consistent estimate of θ_0.[19] The Ω_n can then be consistently estimated as $\hat{\Omega}_n$. The following proposition shows that the feasible optimum GMM estimator with a consistently estimated $\hat{\Omega}_n$ has the same limiting distribution of the optimum GMM estimator based on Ω_n. With the optimum GMM objective function, an overidentification test is available, which can be used as a goodness-of-fit test for the selection of the order of spatial lags.

Proposition 3 *Under Assumptions 1-7, suppose that* $(\frac{\hat{\Omega}_n}{n})^{-1} - (\frac{\Omega_n}{n})^{-1} = o_p(1)$, *then the feasible optimal GMM estimator* $\hat{\theta}_{fo,n}$ *derived from* $\min_{\theta \in \Theta} g_n'(\theta)\hat{\Omega}_n^{-1} g_n(\theta)$ *based on* $g_n(\theta)$ *in (4) has the asymptotic distribution*

$$\sqrt{n}(\hat{\theta}_{fo,n} - \theta_0) \xrightarrow{D} N(0, (\lim_{n \to \infty} \frac{1}{n} D_n' \Omega_n^{-1} D_n)^{-1}). \tag{12}$$

Furthermore, $g_n'(\hat{\theta}_n)\hat{\Omega}_n^{-1} g_n(\hat{\theta}_n) \xrightarrow{D} \chi^2((m + k_{IV}) - (k_x + p + q))$, *where* $(m + k_{IV}) > (k_x + p + q)$.

6 Efficiency and the BGMM estimator

Consider now the issue of selecting the best P_{in}'s and the best IV matrix Q_n. By transforming u_n into ϵ_n free of spatial correlation, the model (1) implies that a SAR(p) process for the transformed variables $\bar{Y}_n = R_n Y_n$ and $\bar{X}_n = R_n X_n$,

$$\bar{Y}_n = \sum_{j=1}^{p} \lambda_{0j} R_n W_{jn} Y_n + R_n X_n \beta_0 + \epsilon_n = \sum_{j=1}^{p} \lambda_{0j} \bar{W}_{jn} \bar{Y}_n + \bar{X}_n \beta_0 + \epsilon_n, \tag{13}$$

where $\bar{W}_{jn} = R_n W_{jn} R_n^{-1}$.

First, consider the case that ϵ_n is normally distributed. Under normality, $\mu_4 = 3\sigma_0^4$ and $\mu_3 = 0$. Hence, the VC matrix $\Omega_n = V_n$ in (10) is a block diagonal matrix. This VC matrix and the derivative matrix in (11) together imply the asymptotic precision matrix (the inverse of the asymptotic VC matrix of an estimator, see p.101, Davidson and MacKinnon, 2004) of $\hat{\theta}_{fo,n}$ as

$$D_n' \Omega_n^{-1} D_n = \begin{pmatrix} A_n B_n^{-1} A_n' & 0_{(p+q) \times k_x} \\ 0_{k_x \times (p+q)} & 0_{k_x \times k_x} \end{pmatrix} + \begin{pmatrix} 0_{q \times q} & 0_{q \times (p+k_x)} \\ 0_{(p+k_x) \times q} & \frac{1}{\sigma_0^2} C_n \end{pmatrix}, \tag{14}$$

where

$$A_n = \begin{pmatrix} tr(P_{1n}^s H_{1n}) & \cdots & tr(P_{mn}^s H_{1n}) \\ \vdots & \ddots & \vdots \\ tr(P_{1n}^s H_{qn}) & \cdots & tr(P_{mn}^s H_{qn}) \\ tr(P_{1n}^s \bar{G}_{1n}) & \cdots & tr(P_{mn}^s \bar{G}_{1n}) \\ \vdots & \ddots & \vdots \\ tr(P_{1n}^s \bar{G}_{pn}) & \cdots & tr(P_{mn}^s \bar{G}_{pn}) \end{pmatrix}, B_n = \frac{1}{2} \begin{pmatrix} tr(P_{1n}^s P_{1n}^s) & \cdots & tr(P_{1n}^s P_{mn}^s) \\ \vdots & \ddots & \vdots \\ tr(P_{mn}^s P_{1n}^s) & \cdots & tr(P_{mn}^s P_{mn}^s) \end{pmatrix},$$

and $C_n = (\bar{G}_{1n}\bar{X}_n\beta_0, \cdots, \bar{G}_{pn}\bar{X}_n\beta_0, \bar{X}_n)'Q_n(Q_n'Q_n)^{-1}Q_n'(\bar{G}_{1n}\bar{X}_n\beta_0, \cdots, \bar{G}_{pn}\bar{X}_n\beta_0, \bar{X}_n)$. With the asymptotic precision matrix in (14), it follows from the generalized Schwartz inequality that the best selection of Q_n is $(\bar{G}_{1n}\bar{X}_n\beta_0, \cdots, \bar{G}_{pn}\bar{X}_n\beta_0, \bar{X}_n)$, and the best selection of P_n's are $\bar{G}_{jn} - \frac{tr(\bar{G}_{jn})}{n}I_n$ and $H_{kn} - \frac{tr(H_{kn})}{n}I_n$ for $j = 1, \cdots, p$ and $k = 1, \cdots, q$.

Let $\mathcal{P}_{1n}$ denote the class of P_n's satisfying Assumption 4. The subclass $\mathcal{P}_{2n}$ of $\mathcal{P}_{1n}$ consisting of P_n's with zero diagonals is also interesting. The corresponding GMM estimator with P_n's from $\mathcal{P}_{2n}$ is robust against distributional assumptions, because, when $vec_D(P_{in}) = 0$ for $i = 1, \cdots, m$, $\Omega_n = V_n$ regardless of the values of μ_3 and $\mu_4 - 3\sigma_0^4$.[20] Based on the Schwartz inequality, the best selection of IV matrix Q_n is still $(\bar{G}_{1n}\bar{X}_n\beta_0, \cdots, \bar{G}_n\bar{X}_n\beta_0, \bar{X}_n)$ but the best P_n's from $\mathcal{P}_{2n}$ are $\bar{G}_{jn} - D(\bar{G}_{jn})$ and $H_{kn} - D(H_{kn})$, for $j = 1, \cdots, p$ and $k = 1, \cdots, q$, under homoskedasticity. $D(A)$ denotes a diagonal matrix with diagonal elements being those of A if A is a vector, or the diagonal elements of A if A is a square matrix.

When the distribution of ϵ_n is unknown, the following proposition provides the best linear and quadratic moments for the estimation of the SARAR(p, q) model via selecting the best P_n's and Q_n.[21] If an intercept appears in $\bar{X}_n$, define $\bar{X}_n^*$ as the submatrix of $\bar{X}_n$ with the intercept column deleted. Thus, $\bar{X}_n = [\bar{X}_n^*, c(\rho_0)l_n]$, where $c(\rho_0)$ is a scalar function of ρ_0 and l_n is an n-dimensional vector of ones.[22] Otherwise $\bar{X}_n^* \equiv \bar{X}_n$. Suppose there are k_x^* columns in $\bar{X}_n^*$. Let $\bar{X}_{nj}$ be the jth column of $\bar{X}_n$, and $\bar{X}_{nj}^*$ be the jth column of $\bar{X}_n^*$. Denote $\bar{X}_{nj}^{*d} = \bar{X}_{nj}^* - \frac{1}{n}l_n l_n'\bar{X}_{nj}^*$, the deviation of $\bar{X}_{nj}^*$ from its sample mean. Let $\bar{G}_{jn}^* = \bar{G}_{jn} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2}D(\bar{G}_{jn}) - \frac{\eta_3}{\sigma_0[(\eta_4-1)-\eta_3^2]}D(\bar{G}_{jn}\bar{X}_n\beta_0)$ and $H_{kn}^* = H_{kn} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2}D(H_{kn})$, for $j = 1, \cdots, p$ and $k = 1, \cdots, q$, where $\eta_3 = \mu_3/\sigma_0^3$ being the skewness of the disturbance, and $\eta_4 = \mu_4/\sigma_0^4$ being the kurtosis of the disturbance. And denote $\mathcal{M}_n = \{\hat{\theta}_{o,n}\}$ the class of optimal GMM estimators derived from linear and quadratic moment conditions (4), with P_n's and Q_n satisfying Assumption 4.

Proposition 4 *Let* $P_{jn}^* = \bar{G}_{jn}^* - \frac{1}{n}tr(\bar{G}_{jn}^*)I_n$ *for* $j = 1, \cdots, p$, $P_{p+k,n}^* = H_{kn}^* - \frac{1}{n}tr(H_{kn}^*)I_n$ *for* $k = 1, \cdots, q$, *and* $P_{p+q+l,n}^* = D(\bar{X}_{nl}^{*d})$ *for* $l = 1, \cdots, k_x^*$. *Let* $Q_n^* = (Q_{1n}^*, Q_{2n}^*, Q_{3n}^*)$ *with* $Q_{2n}^* = (Q_{2n1}^*, \cdots, Q_{2np}^*)$ *and* $Q_{3n}^* = (Q_{3n1}^*, \cdots, Q_{3nq}^*)$ *such that* $Q_{1n}^* = \bar{X}_n + \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}(\bar{X}_n - \frac{1}{n}l_n l_n'\bar{X}_n)$, $Q_{2nj}^* = \bar{G}_{jn}\bar{X}_n\beta_0 + \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}(\bar{G}_{jn}\bar{X}_n\beta_0 - \frac{1}{n}l_n l_n'\bar{G}_{jn}\bar{X}_n\beta_0) - \frac{2\sigma_0\eta_3}{(\eta_4-1)-\eta_3^2}[vec_D(\bar{G}_{jn}) - \frac{1}{n}tr(\bar{G}_{jn})l_n]$, *for* $j = 1, \cdots, p$, *and* $Q_{3nk}^* = vec_D(H_{kn}) - \frac{1}{n}tr(H_{kn})l_n$, *for* $k = 1, \cdots, q$. *Within the class of optimal GMM estimators* $\mathcal{M}_n$, *under Assumptions 1-7, the estimator* $\hat{\theta}_{b,n}$ *derived from* $\min_{\theta\in\Theta} g_n^{*\prime}(\theta)\,\Omega_n^{*-1}g_n^*(\theta)$, *where* $\Omega_n^* = var(g_n^*(\theta_0))$ *and* $g_n^*(\theta) = [Q_n^*, P_{1n}^*\epsilon_n(\theta), \cdots, P_{p+q+k_x^*,n}^*\epsilon_n(\theta)]'\epsilon_n(\theta)$, *is the BGMM estimator with the limiting distribution* $\sqrt{n}(\hat{\theta}_{b,n}-\theta_0)\xrightarrow{D}N(0,\Sigma_b^{-1})$, *where*

$$\Sigma_b=\lim_{n\to\infty}\frac{1}{n}\begin{pmatrix}\Sigma_{11} & \Sigma_{12} & -\frac{2\eta_3}{\sigma_0[(\eta_4-1)-\eta_3^2]}Q_{3n}^{*\prime}\bar{X}_n\\ \Sigma_{12}^{\prime} & \Sigma_{22} & \sigma_0^{-2}Q_{2n}^{*\prime}\bar{X}_n\\ -\frac{2\eta_3}{\sigma_0[(\eta_4-1)-\eta_3^2]}\bar{X}_n^{\prime}Q_{3n}^{*} & \sigma_0^{-2}\bar{X}_n^{\prime}Q_{2n}^{*} & \sigma_0^{-2}\bar{X}_n^{\prime}Q_{1n}^{*}\end{pmatrix},$$

and

$$\Sigma_{11}=\begin{pmatrix}tr(P_{p+1,n}^{*s}H_{1n}) & \cdots & tr(P_{p+q,n}^{*s}H_{1n})\\ \vdots & \ddots & \vdots\\ tr(P_{p+1,n}^{*s}H_{qn}) & \cdots & tr(P_{p+q,n}^{*s}H_{qn})\end{pmatrix},\ \Sigma_{12}=\begin{pmatrix}tr(P_{1n}^{*s}H_{1n}) & \cdots & tr(P_{pn}^{*s}H_{1n})\\ \vdots & \ddots & \vdots\\ tr(P_{1n}^{*s}H_{qn}) & \cdots & tr(P_{pn}^{*s}H_{qn})\end{pmatrix},$$

$$\Sigma_{22}=\begin{pmatrix}\sigma_0^{-2}\left(\bar{G}_{1n}\bar{X}_n\beta_0\right)^{\prime}Q_{2n1}^{*}+tr\left(P_{1n}^{*s}\bar{G}_{1n}\right) & \cdots & \sigma_0^{-2}\left(\bar{G}_{1n}\bar{X}_n\beta_0\right)^{\prime}Q_{2np}^{*}+tr\left(P_{pn}^{*s}\bar{G}_{1n}\right)\\ \vdots & \ddots & \vdots\\ \sigma_0^{-2}\left(\bar{G}_{pn}\bar{X}_n\beta_0\right)^{\prime}Q_{2n1}^{*}+tr\left(P_{1n}^{*s}\bar{G}_{pn}\right) & \cdots & \sigma_0^{-2}\left(\bar{G}_{pn}\bar{X}_n\beta_0\right)^{\prime}Q_{2np}^{*}+tr\left(P_{pn}^{*s}\bar{G}_{pn}\right)\end{pmatrix}.$$

The moment functions $[P_{p+1,n}^{*}\epsilon_n(\theta),\cdots,P_{p+q,n}^{*}\epsilon_n(\theta)]^{\prime}\epsilon_n(\theta)$ are apparently designed for the estimation of ρ_0 in $u_n=\sum_{k=1}^{q}\rho_{0k}M_{kn}u_n+\epsilon_n$. Due to the correlation between linear and quadratic moment functions, it is more involved than the best moment function for estimating the (pure) SAR(q) process $Y_n=\sum_{k=1}^{q}\rho_{0k}M_{kn}Y_n+\epsilon_n$.[23] And the selection of $(P_{1n}^{*},\cdots,P_{pn}^{*},P_{p+q+1,n}^{*},\cdots,P_{p+q+k_x^*,n}^{*})$ and $(Q_{1n}^{*},Q_{2n}^{*},Q_{3n}^{*})$ corresponds to the selection of the best quadratic moment functions and the best IV matrix for the estimation of the transformed MRSAR model (13). These two sets of moment functions estimate ρ_0 and δ_0 simultaneously. The best selections of P_n's and Q_n from $\mathcal{P}_{1n}$ under normality assumption are special cases of P_n^{*}'s and Q_n^{*} given in Proposition 4. When ϵ_n is normally distributed, $\bar{G}_{jn}^{*}$ and H_{kn}^{*} reduce to $\bar{G}_{jn}$ and H_{kn}, respectively, for $j=1,\cdots,p$ and $k=1,\cdots,q$. Hence, it follows that $P_{jn}^{*}=\bar{G}_{jn}-\frac{tr(\bar{G}_{jn})}{n}I_n$, $P_{p+k,n}^{*}=H_{kn}-\frac{tr(H_{kn})}{n}I_n$, $Q_{1n}^{*}=\bar{X}_n$, and $Q_{2n}^{*}=(\bar{G}_{1n}\bar{X}_n\beta_0,\cdots,\bar{G}_{pn}\bar{X}_n\beta_0)$ as $\eta_3=0$, for $j=1,\cdots,p$ and $k=1,\cdots,q$. And it follows arguments in Breusch et al. (1999) that moment functions $[Q_{3n}^{*},D(X_{n1}^{*d})\epsilon_n(\theta),\cdots,D(X_{nk_x^*}^{*d})\epsilon_n(\theta)]^{\prime}\epsilon_n(\theta)$ are redundant given $[Q_{1n}^{*},Q_{2n}^{*},P_{1n}^{*}\epsilon_n(\theta),\cdots,P_{p+q,n}^{*}\epsilon_n(\theta)]^{\prime}\epsilon_n(\theta)$ under normality.[24]

The moment function $g_n^{*}(\theta)$ of the BGMM and its VC matrix Ω_n^{*} involve the unknown parameters $\theta_0,\sigma_0^2,\mu_3$ and μ_4. In practice, with initial $\sqrt{n}$-consistent estimators $\hat{\theta}_n,\hat{\sigma}_n^2,\hat{\mu}_{3n}$ and $\hat{\mu}_{4n}$, P_{in}^{*} and Q_n^{*} in $g_n^{*}(\theta)$ can be replaced by their estimated counterparts $\hat{P}_{in}^{*}$ and $\hat{Q}_n^{*}$, for $i=1,\cdots,k_x^{*}+p+q$, and Ω_n^{*} can be estimated accordingly as $\hat{\Omega}_n^{*}$. The following proposition shows that the feasible BGMM estimator has the same limiting distribution as the BGMM estimator in Proposition 4.

Proposition 5 *Let $\hat{P}_{in}^*$, $\hat{Q}_n^*$ and $\hat{\Omega}_n^*$ be the estimated counterparts of P_{in}^*, Q_n^* and Ω_n^*, for $i = 1, \cdots, k_x^* + p + q$, with the unknown parameters replaced by their $\sqrt{n}$-consistent estimators $\hat{\theta}_n, \hat{\sigma}_n^2, \hat{\mu}_{3n}$ and $\hat{\mu}_{4n}$. Then, under Assumptions 1-7, the estimator $\hat{\theta}_{fb,n}$ from $\min_{\theta \in \Theta} \hat{g}_n^{*\prime}(\theta) \hat{\Omega}_n^{*-1} \hat{g}_n^*(\theta)$ with $\hat{g}_n^*(\theta) = [\hat{Q}_n^*, \hat{P}_{1n}^* \epsilon_n(\theta), \cdots, \hat{P}_{k_x^*+p+q,n}^* \epsilon_n(\theta)]' \epsilon_n(\theta)$ has the same limiting distribution of $\hat{\theta}_{b,n}$ derived from $\min_{\theta \in \Theta} g_n^{*\prime}(\theta) \Omega_n^{*-1} g_n^*(\theta)$.*

Lastly, we compare the asymptotic efficiency of the BGMM estimator with that of the conventional QML estimator and the best G2SLS estimator in Lee (2003). As the first order conditions of the log likelihood function (2) are asymptotically equivalent to some linear and quadratic moment conditions in the sense that their consistent roots have the same limiting distribution, the QML estimator is asymptotically equivalent to some GMM estimator based on those linear and quadratic moment conditions. The BGMM estimator is asymptotically as efficient as the ML estimator when ϵ_{ni}'s are i.i.d. normally distributed. When ϵ_{ni}'s are i.i.d. non-normal errors, the extremum estimator based on the normal likelihood function is a QML estimator. The BGMM estimator improves the efficiency of such a QML estimator by using the best linear and quadratic moment conditions via the selection of P_n^*'s and Q_n^*, and by using the optimal weighting matrix Ω_n^*. On the other hand, the BGMM estimator improves the best G2SLS estimator via joint estimation of ρ_0 and δ_0 using the quadratic moment conditions in addition to the linear moment conditions used in the G2SLS. The additional quadratic moment conditions provide the additional information on the correlation structure of the reduced form disturbances for the estimation. The result is summarized in the following proposition.

Proposition 6 *Under Assumptions 1-7, the BGMM estimator is asymptotically efficient relative to the QML estimator and the best G2SLS estimator.*

7 Monte Carlo Study

In the Monte Carlo study, we first consider the SARAR(1,1) model specified as $Y_n = \lambda W_n Y_n + X_{n1}\beta_1 + X_{n2}\beta_2 + u_n$, where $u_n = \rho W_n u_n + \epsilon_n$, $X_{n1} = (x_{11}, \cdots, x_{n1})'$ and $X_{n2} = (x_{12}, \cdots, x_{n2})'$. x_{i1} and x_{i2} are independently generated standard normal variables for all i, and ϵ_{ni}'s are independently generated from the following 3 distributions, all of which are scaled to have mean 0 and variance

2: (a) normal, $\epsilon_{ni} \sim N(0,2)$, (b) symmetric bimodal mixture normal, $\epsilon_{ni} = \sqrt{2/17}u$ where $u \sim .5N(-4,1) + .5N(4,1)$, and (c) gamma, $\epsilon_{ni} = u - 2$ where $u \sim gamma(2,1)$. The skewness (η_3) and kurtosis (η_4) of these distributions are correspondingly: (a) $\eta_3 = 0$, $\eta_4 = 3$; (b)$\eta_3 = 0$, $\eta_4 \approx 1.228$; and (c) $\eta_3 = \sqrt{2}$, $\eta_4 = 6$. Normal distribution is the basis for comparison. Symmetric bimodal mixture normal distribution and gamma distribution will study the effects of skewness and kurtosis excess on the finite sample performance of various estimators. Asymptotically, the feasible BGMM estimator proposed in Proposition 5 is as efficient as the MLE under (a), and is more efficient than the QML estimator under (b) and (c).

Let W_A denote the weights matrix for the study of crimes across 49 districts in Columbus, Ohio in Anselin (1988). For moderate sample sizes of $n = 245$ and 490, the corresponding spatial weights matrices in the Monte Carlo study are given by $I_5 \otimes W_A$ and $I_{10} \otimes W_A$ respectively, where $\otimes$ denotes the Kronecker product operator. The true λ_0 and ρ_0 are set to be 0.4 in the data generating process. We use different β_0 in different experiments.

The estimation methods considered are: (1) the G2SLS and B2SLS: the G2SLS approach in Kelejian and Prucha (1998) and the best G2SLS method in Lee (2003);[25] (2) the QML: the quasi maximum likelihood method;[26] (3) the GMM1: the feasible best optimal GMM in the class of $\mathcal{P}_2$; (4) the GMM2: the feasible best optimal GMM under the normality assumption; and (5) the BGMM: the general feasible best GMM described in Proposition 5.

The number of repetitions is $1,000$ for each case in the Monte Carlo experiment. The regressors are randomly redrawn for each repetition. In each case, we report the mean 'Mean' and standard deviation 'SD' of the empirical distributions of the estimates. To facilitate the comparison of various estimators, their root mean square errors 'RMSE' are also reported.

Computationally, the G2SLS is the most simple. The best G2SLS involves $S_n^{-1}(\lambda)$, and the GMM1, GMM2, and BGMM involve both $S_n^{-1}(\lambda)$ and $R_n^{-1}(\lambda)$, hence they are more complicated than the G2SLS but much simpler than the conventional QML because they do not need the computation of $|S_n(\lambda)|$ and $|R_n(\rho)|$, and $S_n^{-1}(\lambda)$ and $R_n^{-1}(\lambda)$ are evaluated only once at an initial consistent estimate.

[Tables 1-3 approximately here]

Tables 1-3 report the results of the case that $\beta_{01} = 1$ and $\beta_{02} = -1$, which will be referred to as the case with strong x. The ratio of the variance of $x_{i1}\beta_{10} + x_{i2}\beta_{20}$ over the sum of the variances of $x_{i1}\beta_{10} + x_{i2}\beta_{20}$ and ϵ_i is 0.5. In this case, we use the G2SLS estimate as the initial estimate to implement the B2SLS and the various feasible optimal GMM.[27] For sample size $n = 245$, the G2SLS estimates of ρ_0 are biased downwards by about 12%, under all disturbance specifications. As the sample size increases to $n = 490$, biases in the G2SLS estimates of ρ_0 reduce to $5 \sim 7\%$. When $n = 245$, the G2SLS estimates of λ_0 are slightly biased upwards, and the B2SLS and various GMM estimates of λ_0 as well as the QML estimates of λ_0 and ρ_0 are slightly biased downwards. All the estimates of β_{01} and β_{02} are essentially unbiased for both sample sizes considered. In terms of SD and RMSE, the G2SLS estimates are almost as good as those of the QML, GMM1 and GMM2, under all disturbance specifications. The B2SLS estimates of λ_0 have slightly larger SDs than those of the G2SLS estimates for $n = 245$. Other than that, the B2SLS and the G2SLS estimates are similar for both sample sizes considered. The good finite sample performance of the G2SLS similar to that of the QML has been noted in Kelejian et al. (2004) when X's have strong effects. When the disturbances are normally distributed, for sample size $n = 245$, the QML, GMM1 and GMM2 estimates of λ_0 and ρ_0 are better than the BGMM estimates in terms of smaller SD and RMSE. The performance of the BGMM estimates is as good as the others when $n = 490$. When the disturbances are symmetric and platykurtic, the BGMM estimates of β_0 are a little better than the others. When the disturbances follow gamma distribution that has $\eta_3 \neq 0$, the BGMM estimators have smaller SD and RMSE than the other estimates for both sample sizes considered. For example, when $n = 490$, the percentage reduction in SD of the BGMM estimates of λ_0, ρ_0, β_{01} and β_{02} relative to the QML estimates is, respectively, 23%, 16%, 23% and 24%.[28]

[Figures 1-3 approximately here]

To illustrate whether the finite sample distributions of the estimates can be approximated by the normal distribution in the experiment, we report quantile-quantile plots from the computer package S-Plus with the BGMM estimates for samples size 490 in Figures 1-3. The quantile-quantile plots have similar features for other estimators. As the plotted lines mostly lie on straight lines, the normal approximations seem adequate (Chambers et al., 1983).

[Table 4 approximately here]

Table 4 reports the results of the case that $\beta_{01} = 0.4$ and $\beta_{02} = -0.4$, which will be referred to as the case with weak x. The ratio of the variance of $x_{i1}\beta_{10} + x_{i2}\beta_{20}$ over the sum of the variances of $x_{i1}\beta_{10} + x_{i2}\beta_{20}$ and ϵ_i is about 0.14. Hence λ_0 may be difficult to estimate by the G2SLS. As the feasible B2SLS and GMM estimators may be sensitive to initial consistent estimates, we use the unweighted GMM with $Q_n = (X_n, W_nX_n, W_n^2X_n)$ for linear moments, $P_{1n} = W_n$ and $P_{2n} = W_n^2 - \frac{1}{n}tr(W_n^2)I_n$ for quadratic moments and I_n as the weighting matrix to get initial estimates.[29] The G2SLS estimates of λ_0 are biased upwards and those of ρ_0 are biased downwards. For instance, when $n = 490$ and the disturbances follow the gamma distribution, the G2SLS estimator of λ_0 is upward biased by 21% and that of ρ_0 is downward biased by 37%. The biases of the QML estimates of λ_0 and ρ_0 are in the same direction as those of the G2SLS estimates but smaller in magnitude. The B2SLS and various GMM estimates of λ_0 are downward biased and the B2SLS estimates of ρ_0 are upward biased. When $n = 490$, the biases in the GMM estimators are less than 15% for the normal error and less than 10% for the other error distributions considered. The other estimates are essentially unbiased. The GMM1 and GMM2 estimates of λ_0 and ρ_0 have the smallest SDs for all error distributions considered. For instance, when $n = 490$ and the disturbances follows the normal distribution, the percentage reduction in SD of the GMM2 (the best GMM under normality assumption) estimates of λ_0 and ρ_0 relative to the B2SLS estimates is, respectively, 31% and 18%. On the other hand, when the disturbances are asymmetrically distributed, the BGMM estimates of β_0 have smaller SD and RMSE than the other estimates, as in the case with strong x.[30]

To study the properties of the estimators when the order of the spatial lags is misspecified, we consider a SARAR(2,1) specified as $Y_n = \lambda_1 W_{1n}Y_n + \lambda_2 W_{2n}Y_n + X_{n1}\beta_1 + X_{n2}\beta_2 + u_n$, where $u_n = \rho W_{1n}u_n + \epsilon_n$. W_{1n} and W_{2n} correspond to the row-normalized weights matrices for the study of local school expenditure across 612 urban school districts in Ohio in Tao (2005). Before row normalization, W_{1n} is based on neighbors with common borders: $w_{1ij} = 1$ if i and j share a border and $w_{1ij} = 0$ otherwise. W_{2n} has weights based on the inverse of income differences: $w_{2ij} = 1/|INCOME_i - INCOME_j|$, with $INCOME_i$ being median per capita income in district i over the sample period, for all urban school districts j within the same metropolitan area as i. In the data generating process, we use $\lambda_{01} = 0.4, \lambda_{02} = 0.2, \rho_0 = 0.4, \beta_{01} = 1$ and $\beta_{02} = -1$. The misspecified model has mistakenly excluded $W_{2n}Y_n$ in the estimation. The estimation results are reported in Tables 5-7.

[Tables 5-7 approximately here]

To facilitate the comparison, we report the various estimates of the correctly specified model in the upper panels of Tables 5-7.[31] We use the G2SLS estimate as the initial estimate for the various feasible estimators. Except that the B2SLS estimates of λ_{02} is distorted by outliers, we observe a similar pattern as the results reported in Tables 1-3. We also estimate the misspecified model, i.e., under the exclusion restriction of $\lambda_{02} = 0$, and the results are reported in the lower panels of Tables 5-7. The omitted economic interaction effect represented by $W_{2n}Y_n$ is partly captured by the effect of $W_{1n}Y_n$, but not much. For the misspecified model, the G2SLS estimates of λ_{01} are biased upwards by about 7% and those of ρ_0 are biased downwards by about 5%. The QML and various GMM estimates of λ_{01} are slightly upward biased. The other estimates are essentially unbiased. The estimates of λ_0 and ρ_0 in the misspecified model also have slightly larger SDs. Overall, the exclusion of a spatial lag seems to have small effects on the estimates of the remaining parameters.

In summary, the GMM approaches with both linear and quadratic moments can improve upon the G2SLS and B2SLS in the finite sample when the variation from the exogenous regressors relative to that of the innovations is small. The proposed BGMM improves upon the QML and B2SLS when disturbances are asymmetrically distributed, and the improvement could be as large as 20% in terms of reduction in SD. Furthermore, the GMM estimators are relatively robust to the misspecified order of spatial lags.

8 Conclusion

In this paper, we consider the GMM estimation of high order MRSAR models with SAR disturbances. The proposed GMM approach improves upon the G2SLS in Kelejian and Prucha (1998) and the best G2SLS in Lee (2003) in asymptotic efficiency. Among the optimal GMM estimators, we show the existence of the BGMM estimator that is asymptotically as efficient as MLE under normality, and more efficient than the QML estimator when the disturbances are not normally distributed. Some evidence from Monte Carlo experiments confirms that the proposed GMM may improve upon the finite sample performance of the conventional QML and the best G2SLS approaches.

Notes

[1] The best GMM estimator is the optimal GMM estimator with the best linear and quadratic moment conditions. It is called the "best" because it is the most efficient one within the class of GMM estimators derived from linear and quadratic moment conditions.

[2]The feature of (1) is that the λ's and ρ's are unknown parameters. If the spatial lag components have a form like $\lambda\sum_{j=1}^{p}\omega_j W_{jn}Y_n = \lambda W_n^* Y_n$, where the weight parameters ω_j's are known and determined outside the model, then such an alternative model is technically a SAR model of the first order as analyzed in Lee (2007a).

[3]As the values of the dependent variable are determined by the model with X_n and ϵ_n, the model is, therefore, an equilibrium one. This feature differs from a time series autoregressive model where there is an initial value problem.

[4]Lee (2002) has identified a subclass of models for which the OLS estimator can be consistent.

[5]For any $n\times n$ matrix $A_n = [a_{n,ij}]$, the row sum matrix norm is defined by $||A_n||_\infty = \max_{i=1,\cdots,n}\sum_{j=1}^{n}|a_{n,ij}|$, and the column sum matrix norm is defined by $||A_n||_1 = \max_{j=1,\cdots,n}\sum_{i=1}^{n}|a_{n,ij}|$.

[6]With a single weights matrix W_n, the Ord device (Ord, 1975), explained as follows, can simplify the evaluation of $|I_n - \lambda W_n|$. When W_n is diagonalizable, we have $W_n = R_n D_n R_n^{-1}$, where D_n is a diagonal matrix of eigenvalues and R_n is the corresponding eigenvector matrix. It follows that $|I_n - \lambda W_n| = |I_n - \lambda R_n D_n R_n^{-1}| = |I_n - \lambda D_n| = \prod_{i=1}^{n}(1-\lambda d_{ni})$, where d_{ni}'s are diagonal elements in D_n. The Ord device is to compute the eigenvalues of the spatial weights matrix once and then use them to evaluate the determinant at different values of λ. However, the Ord device will not be applicable to the current model with a few exceptions. For a simple illustration, consider two matrices W_{1n} and W_{2n} that are both diagonalizable, i.e., $W_{jn} = R_{jn}D_{jn}R_{jn}^{-1}$, $j = 1, 2$. Unless $R_{1n} = R_{2n}$, they can not be canceled out in $|I_n - \lambda_1 R_{1n}D_{1n}R_{1n}^{-1} - \lambda_2 R_{2n}D_{2n}R_{2n}^{-1}|$. The R_{1n} might be equal to R_{2n} in some special situations. A well-known case is when both W_{1n} and W_{2n} can be simultaneously diagonalizable. However, to be simultaneously diagonalizable, the sufficient conditions are that both W_{1n} and W_{2n} are symmetric and commutative, i.e., $W_{1n}W_{2n} = W_{2n}W_{1n}$ (see Dhrymes, 1978). Another case is the high order spatial lags model with $W_{jn} = W_n^j$, $j = 1,\cdots,p$, generated as powers of a diagonalizable W_n. In this case, $|I_n - \sum_{j=1}^{p}\lambda_j W_n^j| = |I_n - \sum_{j=1}^{p}\lambda_j R_n D_n^j R_n^{-1}| = |I_n - \sum_{j=1}^{p}\lambda_j D_n^j| = \prod_{i=1}^{n}(1-\sum_{j=1}^{p}\lambda_j d_{in}^j)$. However, the Ord device would not be applicable if redundant and circular routes of the high order spatial operators are eliminated (Blommestein, 1985). The ML method may be practically tractable when all the spatial weights matrices are sparse such that $S_n(\lambda)$ and $R_n(\rho)$ can be effectively evaluated by sparse matrix techniques.

[7]To simplify the computation involved in the best FG2SLS estimator, Kelejian et al. (2004) suggested the best series FG2SLS estimator that is also an asymptotically efficient estimator within the class of IV estimators, with $Q_n = R_n(\hat{\rho}_n)[\sum_{k=0}^{r_n}\hat{\lambda}_n^k W_n^{k+1}X_n\hat{\beta}_n, X_n]$ and r_n is some sequence of natural numbers going to infinite.

[8]In the regression model with SAR disturbances, as all the explanatory variables in the main equation are exogenous variables, the asymptotic distribution of $\hat{\rho}_n$ in Kelejian and Prucha (1999) via the least squares residual does not depend on the asymptotic distribution of the least squares estimator of β_0. For the SARAR(1,1) model, as the second step estimator, the asymptotic distribution of $\hat{\rho}_n$ depends on the asymptotic distribution of the first step estimator of δ_0 via the estimated residual $\hat{u}_n$ in the presence of the spatial lag W_nY_n in the main equation (Kelejian and Prucha, 2007b).

[9]It has a special spatial weights matrix of a single neighbor for each spatial unit and $y_0 = 0$.

[10]These moments have been designed to focus on the estimation of θ. If we are also interested in the estimation of σ^2. It can be estimated by the empirical second moment with estimated residuals of ϵ_n. In Liu et al. (2006), we show that this approach will not lose asymptotic efficiency by focusing on θ.

[11]However, the zero trace assumption of P_n's is not sufficient for consistency of the GMM estimator in the presence of the heteroskedasticity of unknown form. Under heteroskedasticity, we need to use P_n's with zero diagonals to

ensure consistency. (Lin and Lee, 2006)

[12]For the SARAR(1,1) model, Kelejian et al. (2004) suggested the use of the linearly independent columns of $[X_n, W_nX_n, M_nX_n, M_nW_nX_n]$ for Q_n in the G2SLS procedure for estimating the main equation, and use M_n and $M_n'M_n$ to set up moments via estimated residuals of the first stage to estimate the disturbance process.

[13]A sequence of square matrices $\{A_n\}$, where $A_n = [a_{n,ij}]$, is said to be uniformly bounded in row sums (column sums) in absolute value if the sequence of row sum matrix norm $||A_n||_\infty$ (column sum matrix norm $||A_n||_1$) are bounded. (Horn and Johnson, 1985)

[14]Derivation of (5) is given in the Lemma C.9.

[15]It is noted that when the identification of the MRSAR model via linear moments is possible, W_n is not required to be distinct from M_n.

When $X_n = l_n$ (i.e., only intercept) and $M_n = W_n$ is row normalized, Λ_n will not have a full column rank. In this case, the parameters can not be identifiable. When $X_n = l_n$ and W_n is row-normalized, $G_nX_n\beta_0 = X_nc$ where $c = \beta_0/(1 - \lambda_0)$. Thus, $\Lambda_n = (G_nX_n\beta_0, X_n)$ does not have the full column rank.

In practice, if there is a need to specify an M_n for the error process, which should be different from W_n, a possible thinking is, while the spatial weight matrix W_n for the main equation may be designed to capture reactions of economic competitors, there might still be autocorrelation in variables not crucial to the model. Autocorrelated disturbances might then be considered to capture such correlations. This interpretation has been offered, e.g., in Benirschka and Binkley (1994) for a model of agricultural land values. In that case, the correlation of disturbances may be captured by the specification of a spatial correlated process with M_n representing geographic proximity.

[16]There are scenarios where the number of neighbors increases as n increases. Those are large group interaction scenarios, which are relevant for in-filling asymptotics. In Lee (2004), it is shown that such scenarios might imply estimates to have lower than the usual $\sqrt{n}$-rate of convergence. The analysis in this paper can be extended to incorporate the large group interaction scenarios but will involve much complicated notations. For additional and related analyses for GMM estimation with large group interactions, see Lee (2007b).

[17]Note that it is unnecessary to require that, for each θ in Θ, $|S_n(\lambda)|$ is positive. The property of such a determinant does not play a role in the GMM estimation. In theory, any bounded set in Assumption 7 will do as long as θ_0 is in the interior of the parameter space and other assumptions are satisfied at $\theta = \theta_0$. The boundedness (or compactness) assumption of the parameter space is needed for asymptotic analysis in proving the uniform convergence in probability of the GMM objective function. In this regard, the G2SLS estimation has the theoretical advantage as restrictions on the parameter space is not explicitly needed even though there are implicit restrictions due to the uniform boundedness of $S_n(\lambda_0)$ and $R_n(\rho_0)$. The disadvantage of the G2SLS may simply be due to inefficiency, in particular, when exogenous variables in X_n have small effects (relative to disturbances) on the outcomes.

[18]The derivation of (11) is given in the Lemma C.10.

[19]The detailed proof is straightforward and is omitted here.

[20]It is also robust against unknown heteroskedasticity (Lin and Lee, 2006).

[21]When the disturbances are normally distributed, it is quite easy to identify the best moments via the generalized Schwartz matrix as shown above. Without normality, it is not so. In general, the key ingredient is to incorporate proper diagonal elements of G_n in the construction of additional moment conditions. The final derivation of the best moments is based on such an insight and trial by errors.

[22]When M_n is row-normalized, $M_n l_n = l_n$ and $(I_n - \rho_0 M_n)^{-1} l_n = (1-\rho_0)^{-1} l_n$. Hence, $R_n l_n = M_n(I_n - \rho_0 M_n)^{-1} l_n = (I_n - \rho_0 M_n)^{-1} M_n l_n = (1-\rho_0)^{-1} l_n$. In this case, $c_n(\rho_0) = (1-\rho_0)^{-1}$.

[23]For the pure SAR(q) process, the BGMM estimator uses the quadratic moment conditions with $P_{kn}^* = H_{kn}^t - \frac{\eta_4 - 3}{\eta_4 - 1} D(H_{kn}^t)$ for $k = 1, \cdots, q$ (Liu et al., 2006), where $A^t = A - \frac{1}{n} tr(A) I_n$ for an $n \times n$ matrix A.

[24]We note that the quadratic moments with P_n's from $\mathcal{P}_{1n}$ but not $\mathcal{P}_{2n}$ will not be robust when ϵ_{ni}'s have heteroskedastic variances (Lin and Lee, 2006). The quadratic moments with P_n^*'s given in Proposition 4 can improve asymptotic efficiency only under the homoskedasticity assumption.

[25]To estimate the SARAR(1,1) model, we use $Q_n = (X_n, W_n X_n, W_n^2 X_n)$ as the IV matrix for the G2SLS. In general, a valid IV matrix could be $(X_n, W_n X_n, \cdots, W_n^q X_n)$ for some $q \geq 1$. We have tried different values of q. We found that as more spatial lags of X_n are included as IVs, the SD of the estimated (λ_0, ρ_0) will decrease slightly while the bias will increase a lot. To balance the tradeoff between SD and bias, we picked the Q_n according to the RMSE for illustration. To estimate the SARAR(2,1) model presented later, we use $Q_n = (X_n, W_{1n} X_n, W_{2n} X_n, W_{1n}^2 X_n, W_{2n}^2 X_n, W_{1n} W_{2n} X_n, W_{2n} W_{1n} X_n)$ as the IV matrix for the G2SLS.

[26]The QML estimator is calculated using sac.m in Econometrics Toolbox (version 7) by James P. Lesage. Function option $info.lflag = 0$ for full computation (instead of approximation), and other options are set to the default values.

[27]The G2SLS estimates of (λ_0, ρ_0) lie in $(-1, 1)^2$ for all replications.

[28]We also estimated the model by the iterated G2SLS and B2SLS. In the 1000 repetitions, only about 650 repetitions generated convergent estimates. Also the convergent iterated estimates of ρ_0 are severely downward biased. To save space, the Monte Carlo results of the iterated G2SLS and B2SLS estimators are not reported in this paper.

[29]We impose a restricted parameter space on the simple unweighted GMM, so that the estimated $(\hat{\lambda}_n, \hat{\rho}_n)$ lie in $(-1, 1)^2$. There are a few divergent cases. For $n = 490$, the numbers of divergent cases are from 15 to 17 with different error specifications. Additional replications are generated to have a total of 1000 convergent cases for the reported results.

[30]Additional Monte Carlo results can be found in our previous two working papers. We considered alternative disturbance distributions (t distribution and asymmetric bimodal mixture normal distribution) and weights matrix for the SARAR(1,1) model. The general conclusions are similar. We also considered smaller values of β_0 for the case with weak x. We found that as the variation from the exogenous regressors relative to that of the disturbances becomes smaller (than 0.14), the biases and SDs in the G2SLS and B2SLS estimates of (λ_0, ρ_0) dramatically increases, while the various GMM estimators are still reasonably good. Also, there are additional results on the SARAR(2,0) model.

[31]As the QML approach is hard to implement for high order MRSAR models, we did not report the QML estimates for the correctly specified model.

References

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

Anselin, L. and Bera, A. (1998). Spatial dependence in linear regression models, with an introduction to spatial econometrics, *in* A. Ullah and D. E. A. Giles (eds), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York.

Anselin, L. and Smirnov, O. (1996). Efficient algorithms for constructing proper higher order spatial lag operators, *Journal of Regional Science* **36**: 67–89.

Benirschka, M. and Binkley, J. K. (1994). Land price volatility in a geographically dispersed market, *American Journal of Agricultural Economics* **76**: 185–195.

Blommestein, H. J. (1983). Specification and estimation of spatial econometric models: a discussion of alternative strategies for spatial economic modelling, *Regional Science and Urban Economics* **13**: 250–271.

Blommestein, H. J. (1985). Elimination of circular routes in spatial dynamic regression equations, *Regional Science and Urban Economics* **15**: 121–130.

Blommestein, H. J. and Koper, N. A. (1992). Recursive algorithms for the elimination of redundant paths in spatial lag operators, *Journal of Regional Science* **32**: 91–111.

Breusch, T., Qian, H., Schmidt, P. and Wyhowski, D. (1999). Redundancy of moment conditions, *Journal of Econometrics* **91**: 89–111.

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. (1983). *Graphical Methods for Data Analysis*, Wadsworth and Brooks/Cole Pub., Pacific Grove, CA.

Davidson, R. and MacKinnon, J. (2004). *Econometric Theory and Methods*, Oxford University Press, New York.

Dhrymes, P. J. (1978). *Mathematics for Econometrics*, Springer-Verlag, New York.

Horn, R. and Johnson, C. (1985). *Matrix Analysis*, Cambridge Univsersity Press, Cambridge.

Huang, J. S. (1984). The autoregressive moving average model for spatial analysis, *Australian Journal of Statistics* **26**: 169–178.

Kelejian, H. H. and 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. and 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. and Prucha, I. R. (2001). On the asymptotic distribution of the moran i test statistic with applications, *Journal of Econometrics* **104**: 219–257.

Kelejian, H. H. and Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations, *Journal of Econometrics* **118**: 27–50.

Kelejian, H. H. and Prucha, I. R. (2007a). Hac estimation in a spatial framework, *Journal of Econometrics* **140**: 131–154.

Kelejian, H. H. and Prucha, I. R. (2007b). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Working paper, Department of Economics, University of Maryland.

Kelejian, H. H., Prucha, I. R. and Yuzefovich, E. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results, *in* J. LeSage and K. Pace (eds), *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*, Elsevier, New York.

Lee, L. F. (2001). Generalized method of moments estimation of spatial autoregressive processes. Manuscript, Department of Economics, Ohio State University.

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 econometric models, *Econometrica* **72**: 1899–1926.

Lee, L. F. (2007a). Gmm and 2sls estimation of mixed regressive, spatial autoregressive models, *Journal of Econometrics* **137**: 489–514.

Lee, L. F. (2007b). The method of elimination and substitution in the gmm estimation of mixed regressive, spatial autoregressive models, *Journal of Econometrics* **140**: 155–189.

Lin, X. and Lee, L. F. (2006). Gmm estimation of spatial autoregressive models with unknown heteroskedasticity. Working paper, Department of Economics, Ohio State University.

Liu, X., Lee, L. F. and Bollinger, C. (2006). Improved efficient quasi maximum likelihood estimator of spatial autoregressive models. Working paper, Department of Economics, Ohio State University.

Maddala, G. S. (1971). Generalized least squares with an estimated covariance matrix, *Econometrica* **39**: 23–33.

Ord, J. (1975). Estimation methods for models of spatial interaction, *Journal of the American Statistical Association* **70**: 120–126.

Tao, J. (2005). Spatial econometrics: models, methods and applications. PhD thesis, Department of Economics, Ohio State University.

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

APPENDICES

A Summary of Notations

$D(A) = Diag(A)$ is a diagonal matrix with diagonal elements being those of A if A is a vector, or diagonal elements of A if A is a square matrix.

$vec_D(A)$ is a column vector formed by the diagonal elements of a square matrix A.

$A^s = A + A'$ where A is a square matrix.

$A^t = A - \frac{1}{n} tr(A) I_n$ where A is an $n \times n$ matrix.

A^L is a linearly transformed square matrix of A which preserves the uniform boundedness property.

$\delta = (\lambda', \beta')'$; $\theta = (\rho', \delta')'$.

$S_n(\lambda) = I_n - \sum_{j=1}^{p} \lambda_j W_{jn}$; $S_n = S_n(\lambda_0)$; $R_n(\rho) = I_n - \sum_{k=1}^{q} \rho_k M_{kn}$; $R_n = R_n(\rho_0)$.

$G_{jn}(\lambda) = W_{jn} S_n^{-1}(\lambda)$; $G_{jn} = G_{jn}(\lambda_0)$; $H_{kn}(\rho) = M_{kn} R_n^{-1}(\rho)$; $H_{kn} = H_{kn}(\rho_0)$.

$F_n(\rho, \lambda) = R_n(\rho) S_n(\lambda)$; $F_n = R_n S_n$.

$u_n(\delta) = S_n(\lambda) Y_n - X_n \beta$; $\epsilon_n(\theta) = R_n(\rho) u_n(\delta)$.

$\bar{Y}_n(\rho) = R_n(\rho) Y_n$; $\bar{X}_n(\rho) = R_n(\rho) X_n$; $\bar{W}_{jn}(\rho) = R_n(\rho) W_{jn} R_n^{-1}(\rho)$.

$\bar{S}_n(\rho, \lambda) = R_n(\rho) S_n(\lambda) R_n^{-1}(\rho)$; $\bar{G}_{jn}(\rho, \lambda) = \bar{W}_{jn}(\rho) \bar{S}_n^{-1}(\rho, \lambda)$.

$\bar{Y}_n = \bar{Y}_n(\rho_0)$; $\bar{X}_n = \bar{X}_n(\rho_0)$; $\bar{W}_{jn} = \bar{W}_{jn}(\rho_0)$; $\bar{S}_{jn} = \bar{S}_{jn}(\rho_0, \lambda_0)$; $\bar{G}_{jn} = \bar{G}_{jn}(\rho_0, \lambda_0)$.

l_n is an $n \times 1$ vector of ones.

e_{kj} is the jth unit column vector in R^k.

If an intercept appears in $\bar{X}_n$ such that $\bar{X}_n = [\bar{X}_n^*, c(\rho_0)l_n]$, where $c(\rho_0)$ is a scalar function of ρ_0, $\bar{X}_n^*$ is the submatrix of $\bar{X}_n$ with the intercept term removed. Otherwise $\bar{X}_n^* \equiv \bar{X}_n$.

$\bar{X}_{nj}^{*d} = \bar{X}_{nj}^* - \frac{1}{n} l_n l_n' \bar{X}_{nj}^*$ is the deviation of observation $\bar{X}_{nj}^*$ from its sample mean.

$\bar{G}_{jn}^* = \bar{G}_{jn} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2} D(\bar{G}_{jn}) - \frac{\eta_3}{\sigma_0[(\eta_4-1)-\eta_3^2]} D(\bar{G}_{jn}\bar{X}_n\beta_0)$, where $\eta_3 = \mu_3/\sigma_0^3$ and $\eta_4 = \mu_4/\sigma_0^4$.

$H_{kn}^* = H_{kn} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2} D(H_{kn})$.

B Identification

In this appendix, we first estabish the identication of the GMM. Then, we provide identification conditions for the moment equations (6) and (7) in Assumption 5 in the following two subsections.

Proof of Proposition 1. From (5), $Q_n' R_n(\rho) d_n(\delta) = 0$ is explicitly $Q_n' R_n(\rho)\Lambda_n(\delta_0 - \delta) = 0$, which has the unique solution δ_0 if $Q_n' R_n(\rho)\Lambda_n$, where $\Lambda_n = (G_{1n}X_n\beta_0, \cdots, G_{pn}X_n\beta_0, X_n)$, has full column rank $(k_x + p)$ for each possible ρ in its parameter space. With δ_0 identified, because $F_n(\rho, \lambda_0)F_n^{-1} = R_n(\rho)R_n^{-1}$ and $d_n(\delta_0) = 0$, the remaining moment equations in (5) become (6). The identification of ρ_0 via these moment conditions is the same as that of the pure SAR process $u_n = \sum_{k=1}^q \rho_k M_{kn} u_n + \epsilon_n$ via the moments $[P_{1n}R_n(\rho)u_n, \cdots, P_{mn}R_n(\rho)u_n]' R_n(\rho)u_n$ as if u_n were observable. The necessary and sufficient condition, as well as some other sufficient conditions, for the identification of ρ_0 via (6) is given in Proposition 7.

On the other hand, if Λ_n does not have a full column rank $(k_x + p)$, then $d_n(\delta) = 0$ alone could not identify δ_0. Suppose X_n has full column rank k_x. Without loss of generality, assume that $(G_{p_0+1,n}X_n\beta_0, \cdots, G_{pn}X_n\beta_0, X_n)$ has full rank $(p + k_x - p_0)$, for some $1 \le p_0 \le p$, and there exist constant vectors a_j and constants c_{jl} such that $G_{jn}X_n\beta_0 = \sum_{l=p_0+1}^p G_{ln}X_n\beta_0 c_{jl} + X_n a_j$ for $j = 1, \cdots, p_0$. Hence, the linear moment equations $Q_n' R_n(\rho) d(\delta) = 0$ from (5) reduce to

$$Q_n' R_n(\rho)\Big\{ \sum_{l=p_0+1}^{p} G_{ln}X_n\beta_0[\sum_{j=1}^{p_0} (\lambda_{0j} - \lambda_j)\, c_{jl} + (\lambda_{0l} - \lambda_l)] + X_n[\sum_{j=1}^{p_0} (\lambda_{0j} - \lambda_j)\, a_j + (\beta_0 - \beta)]\Big\} = 0,$$

which have all its solutions satisfying

$$\lambda_l = \lambda_{0l} + \sum\nolimits_{j=1}^{p_0} (\lambda_{0j} - \lambda_j)\, c_{jl}, \quad \beta = \sum\nolimits_{j=1}^{p_0} (\lambda_{0j} - \lambda_j)\, a_j + \beta_0, \tag{15}$$

for $l = p_0 + 1, \cdots, p$. From (15), β_0 and λ_{0l} $(l = p_0 + 1, \cdots, p)$ are identifiable once $\lambda_{01}, \cdots, \lambda_{0p_0}$ are identified. With $d_n(\delta) = 0$, the identification of $\lambda_{01}, \cdots, \lambda_{0p_0}$ based on (5) will reduce to (7). Let $v_n = F_n^{-1}\epsilon_n$ be the disturbance vector of that equation. The reduced form equation becomes $Y_n = X_n[\beta_0 + \sum_{j=1}^{p_0} \lambda_{0j} a_j] + \sum_{l=p_0+1}^{p} G_{ln} X_n \beta_0 [\lambda_{0l} + \sum_{j=1}^{p_0} \lambda_{0j} c_{jl}] + v_n$. The moment equations (7) correspond to a pure SARAR(p, q) process,

$$v_n = \sum_{j=1}^{p} \lambda_{0j} W_{jn} v_n + u_n, \qquad u_n = \sum_{k=1}^{q} \rho_{0k} M_{kn} u_n + \epsilon_n. \tag{16}$$

We provide the necessary and sufficient condition for the identification of the moment equations (7) in Proposition 8, and we also discuss some weaker sufficient conditions for identification. With $(\rho_0', \lambda_{01}, \cdots, \lambda_{0p_0})$ identified, as shown above, the remaining parameters can be identified from the linear moment conditions. ■

B.1 Identification of a Pure SAR(q) Process

In this subsection, we discuss the identification of the pure SAR process $u_n = \sum_{k=1}^{q} \rho_k M_{kn} u_n + \epsilon_n$ via the quadratic moment equations (6). Let φ_k and φ_{jk} be m-dimensional vectors with the ith element being, respectively, $\varphi_{k,i} = tr(P_{in}^s H_{kn})$ and $\varphi_{jk,i} = tr(H_{jn}' P_{in} H_{kn})$.

Proposition 7 *The necessary and sufficient condition for (6) to have the unique solution at ρ_0 is that the vectors φ_k's and φ_{jk}'s do not have a linear combination with nonlinear coefficients in the form that*

$$\sum_{k=1}^{q} \delta_k \varphi_k + \sum_{j=1}^{q} \sum_{k=1}^{q} \delta_j \delta_k \varphi_{jk} = 0, \tag{17}$$

for some nonzero constants $\delta_1, \cdots, \delta_q$.

Proof. As $R_n(\rho) R_n^{-1} = I_n + \sum_{k=1}^{q} (\rho_{0k} - \rho_k) H_{kn}$, $tr[R_n'^{-1} R_n'(\rho) P_{in} R_n(\rho) R_n^{-1}] = \sum_{k=1}^{q} (\rho_{0k} - \rho_k) \varphi_{k,i} + \sum_{j=1}^{q} \sum_{k=1}^{q} (\rho_{0j} - \rho_j)(\rho_{0k} - \rho_k) \varphi_{jk,i}$ for $i = 1, \cdots, m$. It is apparent that ρ_0 is a common solution of these m moment equations. The desired result follows. ■

A sufficient identification condition for the pure SAR(q) model is that the φ's are linearly independent. Weaker sufficient conditions are available. If there exists a solution ρ_1 not equal to ρ_{01}, one has $\delta_1 \neq 0$ in (17). This will imply that either φ_1 or φ_{11} will be linearly dependent on all the other φ's. So it is sufficient to identify ρ_1 if each of φ_1 and φ_{11} is linearly independent of the other φ's.

With ρ_1 being identified, (17) becomes $\sum_{k=2}^{q}\delta_k\varphi_k+\sum_{j=2}^{q}\sum_{k=2}^{q}\delta_j\delta_k\varphi_{jk}=0$. Similar arguments apply to the identification of ρ_{02}, and so on.

B.2 Identification of a Pure SARAR(p,q) Process

When Λ_n does not have full column rank, the identification of the original model (1) reduces to the identification of a pure SARAR(p,q) process (16), as shown in the proof of Proposition 1. The identification conditions of (16) can be derived by investigating some characteristics of the moment equations (7). Let $h_{jn}=\bar{G}_{jn}-\sum_{l=p_0+1}^{p}c_{jl}\bar{G}_{ln}$, $\alpha_{\rho_k,i}=tr(P_{in}^sH_{kn})$, $\alpha_{\lambda_j,i}=tr(P_{in}^sh_{jn})$, $\alpha_{\rho_{k_1k_2},i}=tr(H_{k_1n}'P_{in}H_{k_2n})$, $\alpha_{\lambda_{j_1j_2},i}=tr(h_{jn}'P_{in}h_{jn})$, $\alpha_{\rho_k\lambda_j,i}=tr(P_{in}^sH_{kn}h_{jn}+H_{kn}'P_{in}^sh_{jn})$, $\alpha_{\rho_{k_1k_2}\lambda_j,i}=tr(H_{k_1n}'P_{in}^sH_{k_2n}h_{jn})$, $\alpha_{\rho_k\lambda_{j_1j_2},i}=tr(h_{j_1n}'P_{in}^sH_{kn}h_{j_2n})$, and $\alpha_{\rho_{k_1k_2}\lambda_{j_1j_2},i}=tr(h_{j_1n}'H_{k_1n}'P_{in}H_{k_2n}h_{j_2n})$. Let α_{ρ_k} be the m-dimensional vector with $\alpha_{\rho_k,i}$ as its ith element. Similarly, α_{λ_j}, etc., are defined.

Proposition 8 *Suppose Λ_n has column rank $(p+k_x-p_0)$, for some $1\leq p_0\leq p$. The necessary and sufficient condition for (7) and (15) to have the unique solution at $(\rho_0',\lambda_{01},\cdots,\lambda_{0p_0})$ is that the vectors α's do not have a linear combination with nonlinear coefficients in the form that*

$$\sum_{k=1}^{q}\alpha_{\rho_k}\delta_k+\sum_{j=1}^{p_0}\alpha_{\lambda_j}\gamma_j+\sum_{k_1,k_2=1}^{q}\alpha_{\rho_{k_1k_2}}\delta_{k_1}\delta_{k_2}+\sum_{j_1,j_2=1}^{p_0}\alpha_{\lambda_{j_1j_2}}\gamma_{j_1}\gamma_{j_2}$$
$$+\sum_{j=1}^{p_0}\sum_{k=1}^{q}\alpha_{\rho_k\lambda_j}\delta_k\gamma_j+\sum_{j=1}^{p_0}\sum_{k_1,k_2=1}^{q}\alpha_{\rho_{k_1k_2}\lambda_j}\delta_{k_1}\delta_{k_2}\gamma_j+$$
$$\sum_{j_1,j_2=1}^{p_0}\sum_{k=1}^{q}\alpha_{\rho_k\lambda_{j_1j_2}}\delta_k\gamma_{j_1}\gamma_{j_2}+\sum_{j_1,j_2=1}^{p_0}\sum_{k_1,k_2=1}^{q}\alpha_{\rho_{k_1k_2}\lambda_{j_1j_2}}\delta_{k_1}\delta_{k_2}\gamma_{j_1}\gamma_{j_2}=0, \tag{18}$$

for some nonzero constants $\delta_1,\cdots,\delta_q$ or $\gamma_1,\cdots,\gamma_{p_0}$.

Proof. For the identification of the pure SARAR(p,q) process in (16), as $F_n(\rho,\lambda)=F_n+\sum_{k=1}^{q}(\rho_{0k}-\rho_k)M_{kn}S_n+\sum_{j=1}^{p}(\lambda_{0j}-\lambda_j)R_nW_{jn}+\sum_{j=1}^{p}\sum_{k=1}^{q}(\rho_{0k}-\rho_k)(\lambda_{0j}-\lambda_j)M_{kn}W_{jn}$, it implies that $F_n(\rho,\lambda)F_n^{-1}=I_n+\sum_{k=1}^{q}(\rho_{0k}-\rho_k)H_{kn}+\sum_{j=1}^{p}(\lambda_{0j}-\lambda_j)\bar{G}_{jn}+\sum_{j=1}^{p}\sum_{k=1}^{q}(\rho_{0k}-\rho_k)(\lambda_{0j}-$

$\lambda_j)H_{kn}\bar{G}_{jn}$. It follows that

$$
\begin{aligned}
& tr(F_n^{'-1}F_n'(\rho,\lambda)P_{in}F_n(\rho,\lambda)F_n^{-1}) \\
= & \sum_{k=1}^{q}(\rho_{0k}-\rho_k)tr(P_{in}^sH_{kn}) + \sum_{j=1}^{p}(\lambda_{0j}-\lambda_j)tr(P_{in}^s\bar{G}_{jn}) \\
& + \sum_{k_1,k_2=1}^{q}(\rho_{0k_1}-\rho_{k_1})(\rho_{0k_2}-\rho_{k_2})tr(H_{k_1n}'P_{in}H_{k_2n}) \\
& + \sum_{j_1,j_2=1}^{p}(\lambda_{0j_1}-\lambda_{j_1})(\lambda_{0j_2}-\lambda_{j_2})tr(\bar{G}_{j_1n}'P_{in}\bar{G}_{j_2n}) \\
& + \sum_{j=1}^{p}\sum_{k=1}^{q}(\rho_{0k}-\rho_k)(\lambda_{0j}-\lambda_j)tr(P_{in}^sH_{kn}\bar{G}_{jn}+H_{kn}'P_{in}^s\bar{G}_{jn}) \\
& + \sum_{j=1}^{p}\sum_{k_1,k_2=1}^{q}(\rho_{0k_1}-\rho_{k_1})(\rho_{0k_2}-\rho_{k_2})(\lambda_{0j}-\lambda_j)tr(H_{k_1n}'P_{in}^sH_{k_2n}\bar{G}_{jn}) \\
& + \sum_{j_1,j_2=1}^{p}\sum_{k=1}^{q}(\rho_{0k}-\rho_k)(\lambda_{0j_1}-\lambda_{j_1})(\lambda_{0j_2}-\lambda_{j_2})tr(\bar{G}_{j_1n}'P_{in}^sH_{kn}\bar{G}_{j_2n}) \\
& + \sum_{j_1,j_2=1}^{p}\sum_{k_1,k_2=1}^{q}(\rho_{0k_1}-\rho_{k_1})(\rho_{0k_2}-\rho_{k_2})(\lambda_{0j_1}-\lambda_{j_1})(\lambda_{0j_2}-\lambda_{j_2})tr(\bar{G}_{j_1n}'H_{k_1n}'P_{in}H_{k_2n}\bar{G}_{j_2n}),
\end{aligned}
$$

for $i=1,\cdots,m$. Substitution of (15) gives

$$
\begin{aligned}
& tr(F_n^{'-1}F_n'(\rho,\lambda)P_{in}F_n(\rho,\lambda)F_n^{-1}) \\
= & \sum_{k=1}^{q}(\rho_{0k}-\rho_k)\alpha_{\rho_k,i} + \sum_{j=1}^{p_0}(\lambda_{0j}-\lambda_j)\alpha_{\lambda_j,i} \\
& + \sum_{k_1,k_2=1}^{q}(\rho_{0k_1}-\rho_{k_1})(\rho_{0k_2}-\rho_{k_2})\alpha_{\rho_{k_1k_2},i} + \sum_{j_1,j_2=1}^{p_0}(\lambda_{0j_1}-\lambda_{j_1})(\lambda_{0j_2}-\lambda_{j_2})\alpha_{\lambda_{j_1j_2},i} \\
& + \sum_{j=1}^{p_0}\sum_{k=1}^{q}(\rho_{0k}-\rho_k)(\lambda_{0j}-\lambda_j)\alpha_{\rho_k\lambda_j,i} \\
& + \sum_{j=1}^{p_0}\sum_{k_1,k_2=1}^{q}(\rho_{0k_1}-\rho_{k_1})(\rho_{0k_2}-\rho_{k_2})(\lambda_{0j}-\lambda_j)\alpha_{\rho_{k_1k_2}\lambda_j,i} \\
& + \sum_{j_1,j_2=1}^{p_0}\sum_{k=1}^{q}(\rho_{0k}-\rho_k)(\lambda_{0j_1}-\lambda_{j_1})(\lambda_{0j_2}-\lambda_{j_2})\alpha_{\rho_k\lambda_{j_1j_2},i} \\
& + \sum_{j_1,j_2=1}^{p_0}\sum_{k_1,k_2=1}^{q}(\rho_{0k_1}-\rho_{k_1})(\rho_{0k_2}-\rho_{k_2})(\lambda_{0j_1}-\lambda_{j_1})(\lambda_{0j_2}-\lambda_{j_2})\alpha_{\rho_{k_1k_2}\lambda_{j_1j_2},i}.
\end{aligned}
$$

It is apparent that $(\rho_0', \lambda_{01}, \cdots, \lambda_{0p_0})$ is a common solution of these m moment equations. The desired result follows. ■

A sufficient identification condition is that the α's are linearly independent. Weaker sufficient conditions are available. If there exists a solution ρ_1 not equal to ρ_{01}, one has $\delta_1 \neq 0$ in (18). This will imply that either α_{ρ_1} or $\alpha_{\rho_{11}}$ will be linearly dependent on all the other α's. So it is sufficient to identify ρ_{01} if both α_{ρ_1} and $\alpha_{\rho_{11}}$ are linearly independent of all the other α's. With ρ_{01} being identified, (18) becomes

$$\sum_{k=2}^{q} \alpha_{\rho_k} \delta_k + \sum_{j=1}^{p_0} \alpha_{\lambda_j} \gamma_j + \sum_{k_1,k_2=2}^{q} \alpha_{\rho_{k_1 k_2}} \delta_{k_1} \delta_{k_2} + \sum_{j_1,j_2=1}^{p_0} \alpha_{\lambda_{j_1 j_2}} \gamma_{j_1} \gamma_{j_2} + \sum_{j=1}^{p_0} \sum_{k=2}^{q} \alpha_{\rho_k \lambda_j} \delta_k \gamma_j + \sum_{j=1}^{p_0} \sum_{k_1,k_2=2}^{q} \alpha_{\rho_{k_1 k_2} \lambda_j} \delta_{k_1} \delta_{k_2} \gamma_j + \sum_{j_1,j_2=1}^{p_0} \sum_{k=2}^{q} \alpha_{\rho_k \lambda_{j_1 j_2}} \delta_k \gamma_{j_1} \gamma_{j_2} + \sum_{j_1,j_2=1}^{p_0} \sum_{k_1,k_2=2}^{q} \alpha_{\rho_{k_1 k_2} \lambda_{j_1 j_2}} \delta_{k_1} \delta_{k_2} \gamma_{j_1} \gamma_{j_2} = 0, \tag{19}$$

Then similar arguments apply to the identification of ρ_{02}, and so on. With ρ_0 being identified, (19) further reduces to

$$\sum_{j=1}^{p_0} \alpha_{\lambda_j} \gamma_j + \sum_{j_1,j_2=1}^{p_0} \alpha_{\lambda_{j_1 j_2}} \gamma_{j_1} \gamma_{j_2} = 0. \tag{20}$$

So it is sufficient to identify λ_{01} if both α_{λ_1} and $\alpha_{\lambda_{11}}$ are linearly independent of all the other α's in (20). Then similar arguments apply to the identification of λ_{02}, and so on. By symmetric arguments, a similar set of sufficient conditions can be stated for the identification of λ_0 first and then the identification of ρ_0. As a general principle, the true (ρ_0, λ_0) may be identifiable when sufficient distinct moment equations are used and their solution sets intersect only at the true parameter vector. As the GMM estimation with those moment functions can be rewritten in a nonlinear least squares estimation framework with nonlinearity only in parameters, sufficient identification condition can also be derived from the corresponding nonlinear regression equation.

C Some Useful Lemmas

In this appendix, we list some lemmas which are useful for the proofs of the results in the text. First, we state some basic properties. The central limit theorem C.5 is in Kelejian and Prucha (2001). The other properties in C.1-C.8 are either trivial or can be found in Lee (2004; 2007a).

C.1 *Suppose that the elements of the sequences of n-dimensional column vectors $\{z_{1n}\}$ and $\{z_{2n}\}$ are uniformly bounded. If the $n \times n$ dimensional matrices $\{A_n\}$ are uniformly bounded in either row or column sums in absolute value, then $|z'_{1n} A_n z_{2n}| = O(n)$.*

C.2 *Suppose that $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. random variables with zero mean and finite variance σ^2 and finite fourth moment μ_4. Then, for any two $n \times n$ matrices A_n and B_n, $E\left(\epsilon'_n A_n \epsilon_n \cdot \epsilon'_n B_n \epsilon_n\right) = \left(\mu_4 - 3\sigma^4\right) vec'_D\left(A_n\right) vec_D\left(B_n\right) + \sigma^4\left[tr\left(A_n\right) tr\left(B_n\right) + tr\left(A_n B_n^s\right)\right]$, where $B_n^s = B_n + B'_n$.*

C.3 *Suppose that $\{A_n\}$ are uniformly bounded in both row and column sums in absolute value. $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. with zero mean and finite fourth moment. Then, $E(\epsilon_n' A_n \epsilon_n) = O(n)$, $var(\epsilon_n' A_n \epsilon_n) = O(n)$, $\epsilon_n' A_n \epsilon_n = O_p(n)$, and $\frac{1}{n}\epsilon_n' A_n \epsilon_n - \frac{1}{n}E(\epsilon_n' A_n \epsilon_n) = o_p(1)$.*

C.4 *Suppose that A_n is an $n \times n$ matrix with its column sums being uniformly bounded in absolute value, elements of the $n \times k_x$ matrix C_n are uniformly bounded, and $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. with zero mean and finite variance σ^2. Then, $\frac{1}{\sqrt{n}} C_n' A_n \epsilon_n = O_p(1)$ and $\frac{1}{n} C_n' A_n \epsilon_n = o_p(1)$. Furthermore, if the limit of $\frac{1}{n} C_n' A_n A_n' C_n$ exists and is positive definite, then $\frac{1}{\sqrt{n}} C_n' A_n \epsilon_n \xrightarrow{D} N(0, \sigma^2 \lim_{n\to\infty} \frac{1}{n} C_n' A_n A_n' C_n)$.*

C.5 *Suppose that $\{A_n\}$ is a sequence of symmetric $n \times n$ matrices with row and column sums uniformly bounded in absolute value and $b_n = (b_{n1}, \cdots, b_{nn})'$ is an n-dimensional vector such that $\sup_n \frac{1}{n}\sum_{i=1}^n |b_{ni}|^{2+\eta_1} < \infty$ for some $\eta_1 > 0$. $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. random variables with zero mean and finite variance σ^2, and its moment $E(|\epsilon|^{4+2\delta})$ for some $\delta > 0$ exists. Let $\sigma^2_{Q_n}$ be the variance of Q_n where $Q_n = \epsilon_n' A_n \epsilon_n + b_n' \epsilon_n - \sigma^2 tr(A_n)$. Assume that the variance $\sigma^2_{Q_n}$ is bounded away from zero at the rate n. Then, $\frac{Q_n}{\sigma_{Q_n}} \xrightarrow{D} N(0,1)$.*

C.6 *Let $\hat{\theta}_n$ and $\hat{\theta}_n^*$ be, respectively, the minimizers of $\digamma_n(\theta)$ and $\digamma_n^*(\theta)$ in Θ. Suppose that $\frac{1}{n}(\digamma_n(\theta) - \bar{\digamma}_n(\theta))$ converges in probability to zero uniformly in $\theta \in \Theta$, where θ_0 is in the interior of Θ, and $\left\{\frac{1}{n}\bar{\digamma}_n(\theta)\right\}$ satisfies the uniqueness identification condition at θ_0. If $\frac{1}{n}(\digamma_n^*(\theta) - \digamma_n(\theta)) = o_p(1)$ uniformly in $\theta \in \Theta$, then both $\hat{\theta}_n$ and $\hat{\theta}_n^*$ converge in probability to θ_0.*

In addition, suppose that $\frac{1}{n}\frac{\partial^2 \digamma_n(\theta)}{\partial\theta\partial\theta'}$ converges in probability to a well defined limiting matrix, uniformly in $\theta \in \Theta$, which is nonsingular at θ_0, and $\frac{1}{\sqrt{n}}\frac{\partial \digamma_n(\theta_0)}{\partial\theta} = O_p(1)$. If $\frac{1}{n}(\frac{\partial^2 \digamma_n^*(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \digamma_n(\theta)}{\partial\theta\partial\theta'}) = o_p(1)$ uniformly in $\theta \in \Theta$ and $\frac{1}{\sqrt{n}}(\frac{\partial \digamma_n^*(\theta_0)}{\partial\theta} - \frac{\partial \digamma_n(\theta_0)}{\partial\theta}) = o_p(1)$, then $\sqrt{n}(\hat{\theta}_n^* - \theta_0)$ and $\sqrt{n}(\hat{\theta}_n - \theta_0)$ have the same limiting distribution.

C.7 *Under Assumption 2, the projectors $X_n(X_n'X_n)^{-1}X_n'$ and $I_n - X_n(X_n'X_n)^{-1}X_n'$ are uniformly bounded in both row and column sums in absolute value.*

C.8 *Suppose that $\{||W_{jn}||\}$, $\{||M_{kn}||\}$, $\{||S_n^{-1}||\}$, and $\{||R_n^{-1}||\}$, where $||\cdot||$ is a matrix norm, are bounded for $j = 1, \cdots, p$ and $k = 1, \cdots, q$. Then $\{||S_n(\lambda)^{-1}||\}$ and $\{||R_n(\rho)^{-1}||\}$ are uniformly bounded in a neighborhood of λ_0 and ρ_0 respectively.*

The following are two facts for the model (1).

C.9 *For the model (1),* $\epsilon_n(\theta) = R_n(\rho)d_n(\delta) + F_n(\rho,\lambda)F_n^{-1}\epsilon_n$*, where* $d_n(\delta) = \sum_{j=1}^{p}(\lambda_{0j}-\lambda_j)G_{jn}X_n\beta_0 + X_n(\beta_0 - \beta)$*.*

Proof. As $S_n(\lambda)S_n^{-1} = \sum_{j=1}^{p}(\lambda_{0j}-\lambda_j)G_{jn}+I_n$, $\epsilon_n(\theta) = R_n(\rho)[S_n(\lambda)Y_n - X_n\beta] = R_n(\rho)[S_n(\lambda)(S_n^{-1}X_n\beta_0 + S_n^{-1}R_n^{-1}\epsilon_n) - X_n\beta] = R_n(\rho)[\sum_{j=1}^{p}(\lambda_{0j}-\lambda_j)G_{jn}X_n\beta_0 + X_n(\beta_0-\beta)] + R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n$. ■

C.10 *For the model (1),* $D_n = \frac{\partial}{\partial\theta'}E(g_n(\theta_0))$ *is given by (11).*

Proof. The derivatives of $g_n(\theta)$ in (4) with respect to ρ_k, λ_j, and β are $\frac{\partial g_n(\theta)}{\partial \rho_k} = -[Q_n, P_{1n}^s R_n(\rho)u_n(\delta), \cdots, P_{mn}^s R_n(\rho)u_n(\delta)]' M_{kn}u_n(\delta)$, $\frac{\partial g_n(\theta)}{\partial \lambda_j} = -[Q_n, P_{1n}^s R_n(\rho)u_n(\delta), \cdots, P_{mn}^s R_n(\rho)u_n(\delta)]' R_n(\rho)W_{jn}Y_n$, and $\frac{\partial g_n(\theta)}{\partial \beta'} = -[Q_n, P_{1n}^s R_n(\rho)u_n(\delta), \cdots, P_{mn}^s R_n(\rho)u_n(\delta)]' R_n(\rho)X_n$, for $j = 1, \cdots, p$ and $k = 1, \cdots, q$. At θ_0, as $u_n = R_n^{-1}\epsilon_n$ and $W_{jn}Y_n = G_{jn}X_n\beta_0 + G_{jn}R_n^{-1}\epsilon_n$, (11) follows from Assumption 1. ■

The following properties are specific to the model in this paper. C.11 is a trivial extension of Liu et al. (2006). The proofs of C.12 and C.13 will be presented subsequently.

C.11 *Suppose that* z_{1n} *and* z_{2n} *are* n*-dimensional column vectors of constants of which elements are uniformly bounded, the* $n\times n$ *constant matrix* A_n *is uniformly bounded in column sums in absolute value, the* $n\times n$ *constant matrices* B_{1n} *and* B_{2n} *are uniformly bounded in both row and column sums in absolute value, and* $\epsilon_{n1}, \cdots, \epsilon_{nn}$ *are i.i.d. random variables with zero mean and finite second moment.* $\sqrt{n}(\hat\alpha_n - \alpha_0) = O_p(1)$ *where* α_0 *is a* r*-dimensional vector in the interior of its parameter space. The* $n\times n$ *matrix* $C_n(\hat\alpha_n)$ *has the expansion that*

$$C_n(\hat\alpha_n) - C_n(\alpha_0) = \sum_{i=1}^{m-1}\sum_{j_1=1}^{r}\cdots\sum_{j_i=1}^{r}(\hat\alpha_{nj_1}-\alpha_{0j_1})\cdots(\hat\alpha_{nj_i}-\alpha_{0j_i})K_{in}(\alpha_0) + \sum_{j_1=1}^{r}\cdots\sum_{j_m=1}^{r}(\hat\alpha_{nj_1}-\alpha_{0j_1})\cdots(\hat\alpha_{nj_m}-\alpha_{0j_m})K_{mn}(\hat\alpha_n), \quad (21)$$

for some $m\geq 2$*, where* $C_n(\alpha_0)$ *and* $K_{in}(\alpha_0)$ *are uniformly bounded in both row and column sums in absolute value for* $i = 1, \cdots, m-1$*, and* $K_{mn}(\alpha)$ *is uniformly bounded in both row and column sums in absolute value, uniformly in a small neighborhood of* α_0*. Then, (a)* $\frac{1}{n}z_{1n}'(C_n(\hat\alpha_n) - C_n(\alpha_0))z_{2n} = o_p(1)$*; (b)* $\frac{1}{\sqrt{n}}z_{1n}'(C_n(\hat\alpha_n) - C_n(\alpha_0))A_n\epsilon_n = o_p(1)$*; (c)* $\frac{1}{n}\epsilon_n' B_{1n}'(C_n(\hat\alpha_n) - C_n(\alpha_0))B_{2n}\epsilon_n = o_p(1)$*, if (21) holds for* $m\geq 3$*; and (d)* $\frac{1}{\sqrt{n}}\epsilon_n'(C_n(\hat\alpha_n) - C_n(\alpha_0))\epsilon_n = o_p(1)$*, if (21) holds for* $m\geq 4$ *with* $tr(K_{in}(\alpha_0)) = 0$ *for* $i = 1, \cdots, m-1$*.*

Furthermore, suppose $\sqrt{n}(\hat{\gamma}_n - \gamma_0) = O_p(1)$ where γ_0 is a s-dimensional vector in the interior of its parameter space, and the matrix $D_n(\hat{\gamma}_n)$ has the expansion that

$$
\begin{aligned}
D_n(\hat{\gamma}_n) - D_n(\gamma_0) &= \sum_{i=1}^{m-1} \sum_{j_1=1}^{s} \cdots \sum_{j_i=1}^{s} (\hat{\gamma}_{nj_1} - \gamma_{0j_1}) \cdots (\hat{\gamma}_{nj_i} - \gamma_{0j_i}) L_{in}(\gamma_0) \\
&\quad + \sum_{j_1=1}^{s} \cdots \sum_{j_m=1}^{s} (\hat{\gamma}_{nj_1} - \gamma_{0j_1}) \cdots (\hat{\gamma}_{nj_m} - \gamma_{0j_m}) L_{mn}(\hat{\gamma}_n), \qquad (22)
\end{aligned}
$$

for some $m \geq 2$, where $D_n(\gamma_0)$ and $L_{in}(\gamma_0)$ are uniformly bounded in both row and column sums in absolute value for $i = 1, \cdots, m-1$, and $L_{mn}(\gamma)$ is uniformly bounded in both row and column sums in absolute value, uniformly in a small neighborhood of γ_0. Then, (a') $\frac{1}{n} z'_{1n}(C_n(\hat{\alpha}_n) - C_n(\alpha_0))(D_n(\hat{\gamma}_n) - D_n(\gamma_0)) z_{2n} = o_p(1)$; (b') $\frac{1}{\sqrt{n}} z'_{1n}(C_n(\hat{\alpha}_n) - C_n(\alpha_0))(D_n(\hat{\gamma}_n) - D_n(\gamma_0)) A_n \epsilon_n = o_p(1)$; (c') $\frac{1}{n} \epsilon'_n B'_{1n}(C_n(\hat{\alpha}_n) - C_n(\alpha_0))(D_n(\hat{\gamma}_n) - D_n(\gamma_0)) B_{2n} \epsilon_n = o_p(1)$, if (21) and (22) hold for $m \geq 3$; and (d') $\frac{1}{\sqrt{n}} \epsilon'_n(C_n(\hat{\alpha}_n) - C_n(\alpha_0))(D_n(\hat{\gamma}_n) - D_n(\gamma_0)) \epsilon_n = o_p(1)$, if (21) and (22) hold for $m \geq 4$ with $tr(K_{in}(\alpha_0) L_{jn}(\gamma_0)) = 0$ for $i, j = 1, \cdots, m-1$.

C.12 *Suppose that z_{1n} and z_{2n} are n-dimensional column vectors of constants of which their elements are uniformly bounded, the $n \times n$ constant matrix A_n is uniformly bounded in column sums in absolute value, and the $n \times n$ constant matrices B_{1n} and B_{2n} are uniformly bounded in both row and column sums in absolute value. Let $\hat{\theta}_n, \hat{\sigma}^2_n, \hat{\mu}_{3n}$ and $\hat{\mu}_{4n}$ be, respectively, $\sqrt{n}$-consistent estimates of $\theta_0, \sigma^2_0, \mu_3$ and μ_4. Let C_n be either $\bar{G}_{jn}$ or H_{kn}, and C^*_n be either $\bar{G}^*_{jn}$, H^*_{kn}, or $D(\bar{X}^*_{nl})$ for $j = 1, \cdots, p$, $k = 1, \cdots, q$ and $l = 1, \cdots, k^*_x$. Let $\hat{C}_n$ and $\hat{C}^*_n$ be the estimated counterparts of C_n and C^*_n. For these C_n (resp. C^*_n) matrices, C^L_n (resp. C^{*L}_n) represents its linear transformed matrix which preserves the uniform boundedness in row and column sums property. Furthermore, let $\hat{D}_n$ be a stochastic matrix that can be expanded to the form of (21). Then, under Assumptions 1-3, (a) $\frac{1}{n} z'_{1n}(\hat{C}'_n - C'_n)^L z_{2n} = o_p(1)$, $\frac{1}{\sqrt{n}} z'_{1n}(\hat{C}'_n - C'_n)^L A_n \epsilon_n = o_p(1)$, $\frac{1}{n} \epsilon'_n B'_{1n}(\hat{C}_n - C_n)^L B_{2n} \epsilon_n = o_p(1)$, $\frac{1}{\sqrt{n}} \epsilon'_n(\hat{C}_n - C_n)^t \epsilon_n = o_p(1)$; (b) $\frac{1}{n} z'_{1n}(\hat{C}^{*\prime}_n - C^{*\prime}_n)^L z_{2n} = o_p(1)$, $\frac{1}{\sqrt{n}} z'_{1n}(\hat{C}^{*\prime}_n - C^{*\prime}_n)^L A_n \epsilon_n = o_p(1)$, $\frac{1}{n} \epsilon'_n B'_{1n}(\hat{C}^*_n - C^*_n)^L B_{2n} \epsilon_n = o_p(1)$, $\frac{1}{\sqrt{n}} \epsilon'_n(\hat{C}^*_n - C^*_n)^t \epsilon_n = o_p(1)$; (c) $\frac{1}{n} vec'_D(\hat{C}_n - C_n)^L z_{2n} = o_p(1)$, $\frac{1}{\sqrt{n}} vec'_D(\hat{C}_n - C_n)^L A_n \epsilon_n = o_p(1)$; and (d) $\frac{1}{n} vec'_D(\hat{C}^*_n - C^*_n)^L z_{2n} = o_p(1)$, $\frac{1}{n} tr[A'_n(\hat{C}^*_n - C^*_n)^L] = o_p(1)$.*

In addition, if $D_n(\gamma)$ is uniformly bounded in both row and column sums in absolute value, uniformly in a small neighborhood of γ_0 that is in the interior of its parameter space, then (e) $\frac{1}{n} tr[D'_n(\hat{\gamma}_n)(\hat{C}^*_n - C^*_n)^L] = o_p(1)$, where $\hat{\gamma}_n - \gamma_0 = o_p(1)$.

C.13 *Suppose that z_n is an n-dimensional column vector of constants which are uniformly bounded, and the $n \times n$ constant matrix A_n is uniformly bounded in column sums in absolute value. Let $\hat{\theta}_n, \hat{\sigma}_n^2, \hat{\mu}_{3n}$ and $\hat{\mu}_{4n}$ be, respectively, $\sqrt{n}$-consistent estimates of $\theta_0, \sigma_0^2, \mu_3$ and μ_4. Let C_n be either $\bar{G}_{jn}$ or H_{kn}, for $j = 1, \cdots, p$ and $k = 1, \cdots, q$, with $\hat{C}_n$ being the estimated counterpart. Let $T_{1n} = \bar{X}_n + \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}(\bar{X}_n - \frac{1}{n}l_n l_n' \bar{X}_n)$, $T_{2n} = C_n \bar{X}_n \beta_0 + \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}(C_n \bar{X}_n \beta_0 - \frac{1}{n} l_n l_n' C_n \bar{X}_n \beta_0)$, and $T_{3n} = \frac{2\sigma_0 \eta_3}{(\eta_4-1)-\eta_3^2} vec_D(C_n^t)$, with $\hat{T}_{1n}$, $\hat{T}_{2n}$, and $\hat{T}_{3n}$ being their estimated counterparts. Then, under Assumptions 1-3, (a) $\frac{1}{n}(\hat{T}_{in} - T_{in})' z_n = o_p(1)$; and (b) $\frac{1}{\sqrt{n}}(\hat{T}_{in} - T_{in})' A_n \epsilon_n = o_p(1)$, for $i = 1, 2, 3$.*

Furthermore, let $D_n(\hat{\gamma}_n)$ be a stochastic matrix that can be expanded to the form of (22) for some $m \geq 3$. Then, (c) $\frac{1}{n}(\hat{T}_{in} - T_{in})' D_n(\hat{\gamma}_n) = o_p(1)$, for $i = 1, 2, 3$.

To show the proposed moment conditions are optimal, we show adding additional moment conditions to the moment conditions does not increase the asymptotic efficiency of the GMM estimator using the conditions for redundancy in Breusch et al. (1999). Their definition of redundancy is given as follows. "Let $\hat{\theta}$ be the optimal GMM estimator based on a set of (unconditional) moment conditions $E[g_1(y,\theta)] = 0$. Now add some extra moment conditions $E[g_2(y,\theta)] = 0$ and let $\tilde{\theta}$ be the optimal GMM estimator based on the whole set of moment conditions $E[g(y,\theta)] \equiv E[g_1'(y,\theta), g_2'(y,\theta)]' = 0$. We say that the moment conditions $E[g_2(y,\theta)] = 0$ are *redundant* given the moment conditions $E[g_1(y,\theta)] = 0$, or simply that g_2 is *redundant* given g_1, if the asymptotic variances of $\hat{\theta}$ and $\tilde{\theta}$ are the same." (Breusch et al., 1999, p. 90) Let $\Omega \equiv E[g(y,\theta) g'(y,\theta)] = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$, with $\Omega_{jl} = E[g_j(y,\theta) g_l'(y,\theta)]$ for $j, l = 1, 2$. And define $D_j = E[\partial g_j(y,\theta)/\partial\theta']$ for $j = 1, 2$. Suppose the dimensions of $g_1(y,\theta)$, $g_2(y,\theta)$ and θ are, respectively, k_1, k_2 and k_θ.

C.14 (Theorem 1 in Breusch et al., 1999) *The following statements are equivalent. (a) g_2 is redundant given g_1; (b) $D_2 = \Omega_{21}\Omega_{11}^{-1} D_1$; and (c) there exists a $k_1 \times k_\theta$ matrix A such that $D_1 = \Omega_{11} A$ and $D_2 = \Omega_{21} A$.*

C.15 (Theorem 2 in Breusch et al., 1999) *Suppose $E[g(\theta)] \equiv E[g_1'(\theta), g_2'(\theta), g_3'(\theta)]' = 0$, or simply $g = (g_1', g_2', g_3')'$. Then $(g_2', g_3')'$ is redundant given g_1 if and only if g_2 is redundant given g_1 and g_3 is redundant given g_1.*

D Proofs

Proof of C.12. As $S_n - \hat{S}_n = \sum_{j=1}^{p}(\hat{\lambda}_{nj} - \lambda_{0j})W_{jn}$, it follows that $\hat{S}_n^{-1} - S_n^{-1} = \hat{S}_n^{-1}(S_n - \hat{S}_n)S_n^{-1} = \hat{S}_n^{-1}[\sum_{j=1}^{p}(\hat{\lambda}_{nj} - \lambda_{0j})G_{jn}]$. By induction,

$$\begin{aligned}\hat{S}_n^{-1} - S_n^{-1} &= S_n^{-1}\sum_{i=1}^{m-1}[\sum_{j=1}^{p}(\hat{\lambda}_{nj} - \lambda_{0j})G_{jn}]^i + \hat{S}_n^{-1}[\sum_{j=1}^{p}(\hat{\lambda}_{nj} - \lambda_{0j})G_{jn}]^m \\ &= \sum_{i=1}^{m-1}\sum_{j_1=1}^{p}\cdots\sum_{j_i=1}^{p}(\hat{\lambda}_{nj_1} - \lambda_{0j_1})\cdots(\hat{\lambda}_{nj_i} - \lambda_{0j_i})(S_n^{-1}G_{j_1n}\cdots G_{j_in}) \\ &\quad + \sum_{j_1=1}^{p}\cdots\sum_{j_m=1}^{p}(\hat{\lambda}_{nj_1} - \lambda_{0j_1})\cdots(\hat{\lambda}_{nj_m} - \lambda_{0j_m})(\hat{S}_n^{-1}G_{j_1n}\cdots G_{j_mn}), \qquad (23)\end{aligned}$$

for any $m \geq 2$. Hence, it follows that

$$\begin{aligned}(\hat{G}_{ln} - G_{ln})^L &= \sum_{i=1}^{m-1}\sum_{j_1=1}^{p}\cdots\sum_{j_i=1}^{p}(\hat{\lambda}_{nj_1} - \lambda_{0j_1})\cdots(\hat{\lambda}_{nj_i} - \lambda_{0j_i})(G_{ln}G_{j_1n}\cdots G_{j_in})^L \\ &\quad + \sum_{j_1=1}^{p}\cdots\sum_{j_m=1}^{p}(\hat{\lambda}_{nj_1} - \lambda_{0j_1})\cdots(\hat{\lambda}_{nj_m} - \lambda_{0j_m})(\hat{G}_{ln}G_{j_1n}\cdots G_{j_mn})^L, \qquad (24)\end{aligned}$$

which conforms to the expansion (21) with $K_{in}(\lambda_0) = (G_{ln}G_{j_1n}\cdots G_{j_in})^L$ and $K_{mn}(\hat{\lambda}_n) = (\hat{G}_{ln}G_{j_1n}\cdots G_{j_mn})^L$. Analogously, we have,

$$\begin{aligned}\hat{R}_n^{-1} - R_n^{-1} &= \sum_{i=1}^{m-1}\sum_{k_1=1}^{q}\cdots\sum_{k_i=1}^{q}(\hat{\rho}_{nk_1} - \rho_{0k_1})\cdots(\hat{\rho}_{nk_i} - \rho_{0k_i})(R_n^{-1}H_{k_1n}\cdots H_{k_in}) \\ &\quad + \sum_{k_1=1}^{q}\cdots\sum_{k_m=1}^{q}(\hat{\rho}_{nk_1} - \rho_{0k_1})\cdots(\hat{\rho}_{nk_m} - \rho_{0k_m})(\hat{R}_n^{-1}H_{k_1n}\cdots H_{k_mn}), \qquad (25)\end{aligned}$$

for any $m \geq 2$, and

$$\begin{aligned}(\hat{H}_{ln} - H_{ln})^L &= \sum_{i=1}^{m-1}\sum_{k_1=1}^{q}\cdots\sum_{k_i=1}^{q}(\hat{\rho}_{nk_1} - \rho_{0k_1})\cdots(\hat{\rho}_{nk_i} - \rho_{0k_i})(H_{ln}H_{k_1n}\cdots H_{k_in})^L \\ &\quad + \sum_{k_1=1}^{q}\cdots\sum_{k_m=1}^{q}(\hat{\rho}_{nk_1} - \rho_{0k_1})\cdots(\hat{\rho}_{nk_m} - \rho_{0k_m})(\hat{H}_{ln}H_{k_1n}\cdots H_{k_mn})^L. \qquad (26)\end{aligned}$$

(26) conforms to the expansion (21) with $K_{in}(\rho_0) = (H_{ln}H_{k_1n}\cdots H_{k_in})^L$, and $K_{mn}(\hat{\rho}_n) = (\hat{H}_{ln}H_{k_1n}\cdots H_{k_mn})^L$. As $\bar{G}_n = R_nG_nR_n^{-1}$, we have $\hat{\bar{G}}_n - \bar{G}_n = (\hat{R}_n - R_n)\hat{G}_n\hat{R}_n^{-1} + R_n(\hat{G}_n - G_n)(\hat{R}_n^{-1} - R_n^{-1}) + R_n(\hat{G}_n - G_n)R_n^{-1} + R_nG_n(\hat{R}_n^{-1} - R_n^{-1})$. $(\hat{R}_n - R_n)\hat{G}_n\hat{R}_n^{-1} = \sum_{k=1}^{q}(\rho_{0k} - \hat{\rho}_{nk})M_{kn}\hat{G}_n\hat{R}_n^{-1}$. On the other hand,

$\hat{G}_n$ and $\hat{R}_n^{-1}$ can be expanded to the form of (21) by (24) and (25) respectively. Note that when the transformation $\cdot^t$is taken, the deterministic parts of the expansions of $R_n(\hat{G}_n - G_n)(\hat{R}_n^{-1} - R_n^{-1})$, $R_n(\hat{G}_n - G_n)R_n^{-1}$ and $R_n G_n(\hat{R}_n^{-1} - R_n^{-1})$ have a zero trace by construction. Hence (a) follows from C.11, where the uniform boundedness in a neighborhood of the true parameters of the relevant matrices in the remainder terms follow from C.8.

For (b), first consider the case that C_n^* is either $\bar{G}_{jn}^*$ or H_{kn}^* for $j = 1, \cdots, p$ and $k = 1, \cdots, q$. We have $\bar{G}_{jn}^* = \bar{G}_{jn} - \frac{\kappa - 2\sigma_0^6}{\kappa} D(\bar{G}_{jn}) - \frac{\sigma_0^2 \mu_3}{\kappa} D(\bar{G}_{jn}\bar{X}_n\beta_0)$ and $H_{kn}^* = H_{kn} - \frac{\kappa-2\sigma_0^6}{\kappa} D(H_{kn})$, for $j = 1, \cdots, p$ and $k = 1, \cdots, q$, where $\kappa = \sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2$. Let $\hat{\kappa}_n$ be κ's estimated counterpart, and $U_{1n} = [\hat{C}_n - \frac{\hat{\kappa}_n - 2(\hat{\sigma}_n^2)^3}{\hat{\kappa}_n} D(\hat{C}_n)] - [C_n - \frac{\kappa - 2\sigma_0^6}{\kappa} D(C_n)] = (\hat{C}_n - C_n) - (1 - \frac{2(\hat{\sigma}_n^2)^3}{\hat{\kappa}_n}) D(\hat{C}_n - C_n) + (\frac{2(\hat{\sigma}_n^2)^3}{\hat{\kappa}_n} - \frac{2\sigma_0^6}{\kappa}) D(C_n)$. As $(2(\hat{\sigma}_n^2)^3/\hat{\kappa}_n - 2\sigma_0^6/\kappa) = o_p(1)$, it follows from (a) and C.1 that $\frac{1}{n} z_{1n}'(U_{1n}')^L z_{2n} = o_p(1)$. On the other hand, let $U_{2n} = -\frac{\hat{\sigma}_n^2 \hat{\mu}_{3n}}{\hat{\kappa}_n} D(\hat{C}_n \bar{X}_n(\hat{\rho}_n)\hat{\beta}_n - C_n \bar{X}_n \beta_0) - (\frac{\hat{\sigma}_n^2 \hat{\mu}_{3n}}{\hat{\kappa}_n} - \frac{\sigma_0^2 \mu_3}{\kappa}) D(C_n \bar{X}_n \beta_0) = -\frac{\hat{\sigma}_n^2 \hat{\mu}_{3n}}{\hat{\kappa}_n} D[(\hat{C}_n - C_n)\bar{X}_n \beta_0 + \hat{C}_n \bar{X}_n(\hat{\beta}_n - \beta_0) + \hat{C}_n(\bar{X}_n(\hat{\rho}_n) - \bar{X}_n)\hat{\beta}_n] - (\frac{\hat{\sigma}_n^2 \hat{\mu}_{3n}}{\hat{\kappa}_n} - \frac{\sigma_0^2\mu_3}{\kappa}) D(C_n \bar{X}_n \beta_0)$, where $\hat{C}_n - C_n$ takes the general form $\hat{C}_n - C_n = \sum_{i=1}^{m-1} \sum_{j_1=1}^r \cdots \sum_{j_i=1}^r (\hat{\alpha}_{nj_1} - \alpha_{0j_1}) \cdots (\hat{\alpha}_{nj_i} - \alpha_{0j_i}) K_{in}(\alpha_0) + \sum_{j_1=1}^r \cdots \sum_{j_m=1}^r (\hat{\alpha}_{nj_1} - \alpha_{0j_1}) \cdots (\hat{\alpha}_{nj_m} - \alpha_{0j_m}) K_{mn}(\hat{\alpha}_n)$, in the proof of (a). Therefore, $D[(\hat{C}_n - C_n)\bar{X}_n \beta_0] = \sum_{i=1}^{m-1} \sum_{j_1=1}^r \cdots \sum_{j_i=1}^r (\hat{\alpha}_{nj_1} - \alpha_{0j_1}) \cdots (\hat{\alpha}_{nj_i} - \alpha_{0j_i}) D[K_{in}(\alpha_0)\bar{X}_n\beta_0] + \sum_{j_1=1}^r \cdots \sum_{j_m=1}^r (\hat{\alpha}_{nj_1} - \alpha_{0j_1}) \cdots (\hat{\alpha}_{nj_m} - \alpha_{0j_m}) D[K_{mn}(\hat{\alpha}_n)\bar{X}_n\beta_0]$. As conditions in C.11 are satisfied via C.8, it follows that $\frac{1}{n} z_{1n}' D'[(\hat{C}_n - C_n)\bar{X}_n\beta_0]^L z_{2n} = o_p(1)$. Let e_{kj} be the jth unit column vector in R^k, then $\frac{1}{n} z_{1n}' D'[\hat{C}_n \bar{X}_n(\hat{\beta}_n - \beta_0)] z_{2n} = \frac{1}{n}\sum_{i=1}^n z_{1n,i} z_{2n,i} e_{ni}' \hat{C}_n \bar{X}_n(\hat{\beta}_n - \beta_0) = o_p(1)$, as $\frac{1}{n}\sum_{i=1}^n z_{1n,i} z_{2n,i} e_{ni}' \hat{C}_n \bar{X}_n = O_p(1)$ and $\hat{\beta}_n - \beta_0 = o_p(1)$. Similarly, $\frac{1}{n} z_{1n}' D'[\hat{C}_n(\bar{X}_n(\hat{\rho}_n) - \bar{X}_n)\hat{\beta}_n] z_{2n} = \sum_{k=1}^q (\rho_{0k} - \hat{\rho}_{nk}) \frac{1}{n} z_{1n}' D'(\hat{C}_n M_{kn} X_n \hat{\beta}_n) z_{2n} = o_p(1)$. The remaining term in $\frac{1}{n} z_{1n}'(U_{2n}')^L z_{2n}$ is $o_p(1)$ as $(\hat{\sigma}_n^2 \hat{\mu}_{3n}/\hat{\kappa}_n - \sigma_0^2 \mu_3/\kappa) = o_p(1)$. And with similar arguments and corresponding results in C.11, the other results in (b) follow, when C_n^* is either $\bar{G}_{jn}^*$ or H_{kn}^* for $j = 1, \cdots, p$ and $k = 1, \cdots, q$. When $C_n^* = D(\bar{X}_{nl}^*)$, for $l = 1, \cdots, k_x^*$, we have $\hat{C}_n^* - C_n^* = D(R_n(\hat{\rho}_n)X_{nl}^*) - D(R_n(\rho_0)X_{nl}^*) = \sum_{k=1}^q (\rho_{0k} - \hat{\rho}_{nk}) D(M_{kn} X_{nl}^*)$. Because $(\rho_0 - \hat{\rho}_n) = o_p(1)$ and $\sqrt{n}(\rho_0 - \hat{\rho}_n) = O_p(1)$, the 4 claims in (b) hold for $C_n^* = D(\bar{X}_{nj}^*)$ by C.1, C.4, C.3, and C.5 respectively.

For (c), as $vec_D'(\hat{C}_n - C_n)^L = l_n' D(\hat{C}_n - C_n)^L$, the results follow from similar argument as in the proof of (a).

For (d), as $vec_D'(\hat{C}_n^* - C_n^*)^L = l_n' D(\hat{C}_n^* - C_n^*)^L$, it follows from similar arguments as in the proof of (b) that $\frac{1}{n} vec_D'(\hat{C}_n^* - C_n^*)^L z_{2n} = o_p(1)$. To prove $\frac{1}{n} tr[A_n'(\hat{C}_n^* - C_n^*)^L] = o_p(1)$, first we consider the case when C_n^* is either $\bar{G}_{jn}^*$ or H_{kn}^* for $j = 1, \cdots, p$ and $k = 1, \cdots, q$. As in the proof of (a), for $m = 2$, $\hat{C}_n - C_n = \sum_{j=1}^r (\hat{\alpha}_{nj} - \alpha_{0j}) K_{1n}(\alpha_0) + \sum_{j_1=1}^r \sum_{j_2=1}^r (\hat{\alpha}_{nj_1} - \alpha_{0j_1})(\hat{\alpha}_{nj_2} - \alpha_{0j_2}) K_{2n}(\hat{\alpha}_n)$. Hence, it follows $\frac{1}{n}tr[A_n'(\hat{C}_n - C_n)^L] = \sum_{j=1}^r (\hat{\alpha}_{nj} - \alpha_{0j})\frac{1}{n}tr(A_n' K_{1n}^L(\alpha_0)) + \sum_{j_1=1}^r \sum_{j_2=1}^r (\hat{\alpha}_{nj_1} - \alpha_{0j_1})(\hat{\alpha}_{nj_2} - \alpha_{0j_2})\frac{1}{n}tr(A_n' K_{2n}^L(\hat{\alpha}_n)) = o_p(1)$, because $\frac{1}{n}tr(A_n' K_{1n}^L(\alpha_0)) = O(1)$, $\frac{1}{n}tr(A_n' K_{2n}^L(\hat{\alpha}_n)) = O_p(1)$, and $\hat{\alpha}_n - \alpha_0 = o_p(1)$. Similarly, $\frac{1}{n}tr[A_n' D(\hat{C}_n \bar{X}_n(\hat{\rho}_n)\hat{\beta}_n - C_n \bar{X}_n \beta_0)] = \frac{1}{n}vec_D'(A_n)[(\hat{C}_n - C_n)\bar{X}_n\beta_0 + \hat{C}_n(\bar{X}_n(\hat{\rho}_n) - \bar{X}_n)\hat{\beta}_n + \hat{C}_n\bar{X}_n(\hat{\beta}_n - \beta_0)] = o_p(1)$, because $\frac{1}{n}vec_D'(A_n)(\hat{C}_n - C_n)\bar{X}_n\beta_0 = \sum_{j=1}^r(\hat{\alpha}_{nj} - \alpha_{0j})\frac{1}{n}vec_D'(A_n)K_{1n}(\alpha_0)\bar{X}_n\beta_0 + \sum_{j_1=1}^r\sum_{j_2=1}^r(\hat{\alpha}_{nj_1} - \alpha_{0j_1})(\hat{\alpha}_{nj_2} - \alpha_{0j_2})\frac{1}{n}vec_D'(A_n)K_{2n}(\hat{\alpha}_n)\bar{X}_n\beta_0$, $\frac{1}{n}vec_D'(A_n)\hat{C}_n(\bar{X}_n(\hat{\rho}_n) - \bar{X}_n)\hat{\beta}_n = \sum_{k=1}^q(\rho_{0k} - \hat{\rho}_{nk})\frac{1}{n}vec_D'(A_n)\hat{C}_n M_{kn}X_n\hat{\beta}_n$, and $\frac{1}{n}vec_D'(A_n)\hat{C}_n\bar{X}_n(\hat{\beta}_n - \beta_0)$ are $o_p(1)$. As $\frac{1}{n}tr[A_n' D(C_n)] = O(1)$, $\frac{1}{n}tr[A_n' D(C_n X_n \beta_0)] = O(1)$, and $\hat{\sigma}_n^2, \hat{\mu}_{3n}, \hat{\kappa}_n$ are consistent estimates, it follows that $\frac{1}{n}tr[A_n'(\hat{C}_n^* - C_n^*)^L] = o_p(1)$. When $C_n^* = D(\bar{X}_{nl}^*)$, for $l = 1, \cdots, k_x^*$, we have $\frac{1}{n}tr[A_n'(\hat{C}_n^* - C_n^*)^L] = \sum_{k=1}^q(\rho_{0k} - \hat{\rho}_{nk})\frac{1}{n}tr[A_n' D(M_{kn}X_{nl}^*)] = o_p(1)$, because $\rho_0 - \hat{\rho}_n = o_p(1)$ and $\frac{1}{n}tr[A_n' D(M_{kn}X_{nl}^*)] = O(1)$.

(e) As $D_n(\gamma)$ is uniformly bounded in both row and column sums in absolute value, uniformly in a small neighborhood of γ_0, and $\hat{\gamma}_n - \gamma_0 = o_p(1)$, it follows that $D_n(\hat{\gamma}_n)$ is uniformly bounded in both row and column sums in absolute value with probability one. The remaining arguments will be similar to those of the part 2 of (d). ■

Proof of C.13. As $\kappa = \sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2$, with $\mu_3 = \eta_3\sigma_0^3$ and $\mu_4 = \eta_4\sigma_0^4$, we have

$$\begin{aligned}\hat{T}_{1n} - T_{1n} &= (\bar{X}_n(\hat{\rho}_n) - \bar{X}_n) + \frac{\hat{\mu}_{3n}^2}{\hat{\kappa}_n}(I_n - \frac{1}{n}l_n l_n')(\bar{X}_n(\hat{\rho}_n) - \bar{X}_n) \\ &\quad + (\frac{\hat{\mu}_{3n}^2}{\hat{\kappa}_n} - \frac{\mu_3^2}{\kappa})(I_n - \frac{1}{n}l_n l_n')\bar{X}_n, \qquad (27)\end{aligned}$$

$$\begin{aligned}\hat{T}_{2n} - T_{2n} &= (\hat{C}_n\bar{X}_n(\hat{\rho}_n)\hat{\beta}_n - C_n\bar{X}_n\beta_0) + \frac{\hat{\mu}_{3n}^2}{\hat{\kappa}_n}(I_n - \frac{1}{n}l_n l_n')(\hat{C}_n\bar{X}_n(\hat{\rho}_n)\hat{\beta}_n - C_n\bar{X}_n\beta_0) \\ &\quad + (\frac{\hat{\mu}_{3n}^2}{\hat{\kappa}_n} - \frac{\mu_3^2}{\kappa})(I_n - \frac{1}{n}l_n l_n')C_n\bar{X}_n\beta_0, \qquad (28)\end{aligned}$$

and $\hat{T}_{3n} - T_{3n} = \frac{2(\hat{\sigma}_n^2)^2\hat{\mu}_{3n}}{\hat{\kappa}_n}vec_D(\hat{C}_n - C_n)^t + (\frac{2(\hat{\sigma}_n^2)^2\hat{\mu}_{3n}}{\hat{\kappa}_n} - \frac{2\sigma_0^4\mu_3}{\kappa})vec_D(C_n^t)$. Let Γ_n be either I_n or $I_n - \frac{1}{n}l_n l_n'$. $\frac{1}{n}(\hat{T}_{1n} - T_{1n})'z_n = o_p(1)$ since $\frac{1}{n}(\bar{X}_n(\hat{\rho}_n) - \bar{X}_n)'\Gamma_n' z_n = \sum_{k=1}^q(\rho_{0k} - \hat{\rho}_{nk})\frac{1}{n}(M_{kn}X_n)'\Gamma_n' z_n = o_p(1)$ and $\hat{\mu}_{3n}^2/\hat{\kappa}_n - \mu_3^2/\kappa = o_p(1)$. For the first two terms in $\frac{1}{n}(\hat{T}_{2n} - T_{2n})'z_n$, since $\frac{1}{n}[(\hat{C}_n - C_n)R_n(\hat{\rho}_n)\bar{X}_n\hat{\beta}_n]'\Gamma_n' z_n = o_p(1)$ by C.12, $\frac{1}{n}[C_n(\bar{X}_n(\hat{\rho}_n) - \bar{X}_n)\hat{\beta}_n]'\Gamma_n' z_n = \sum_{k=1}^q(\rho_{0k} - \hat{\rho}_{nk})\frac{1}{n}(C_n M_{kn}X_n\hat{\beta}_n)'\Gamma_n' z_n = o_p(1)$, and $(\hat{\beta}_n - \beta_0)'\frac{1}{n}(C_n\bar{X}_n)'\Gamma_n' z_n = o_p(1)$, it follows that $\frac{1}{n}(\hat{C}_n\bar{X}_n(\hat{\rho}_n)\hat{\beta}_n - C_n\bar{X}_n\beta_0)'\Gamma_n' z_n = \frac{1}{n}[(\hat{C}_n - C_n)R_n(\hat{\rho}_n)\bar{X}_n\hat{\beta}_n + C_n(\bar{X}_n(\hat{\rho}_n) - \bar{X}_n)\hat{\beta}_n + C_n\bar{X}_n(\hat{\beta}_n - \beta_0)]'\Gamma_n' z_n = o_p(1)$. The remaining term in $\frac{1}{n}(\hat{T}_{2n} - T_{2n})'z_n$ is $o_p(1)$ because $\hat{\mu}_{3n}^2/\hat{\kappa}_n - \mu_3^2/\kappa = o_p(1)$ and $\frac{1}{n}(\Gamma_n C_n\bar{X}_n\beta_0)'z_n = O(1)$.

For the first term in $\frac{1}{n}(\hat{T}_{3n} - T_{3n})'z_n$, it follows from C.12 that $\frac{1}{n}vec_D'(\hat{C}_n - C_n)^t z_n = o_p(1)$. And the remaining term in $\frac{1}{n}(\hat{T}_{3n} - T_{3n})'z_n$ is $o_p(1)$ because $(\hat{\sigma}_n^2)^2\hat{\mu}_{3n}/\hat{\kappa}_n - \sigma_0^4\mu_3/\kappa = o_p(1)$ and $\frac{1}{n}vec_D'(\hat{C}_n^t)z_n = O(1)$. This proves (a).

For (b), the first two terms in $\frac{1}{\sqrt{n}}(\hat{T}_{1n} - T_{1n})'A_n\epsilon_n$ are $o_p(1)$ because $\frac{1}{\sqrt{n}}(\bar{X}_n(\hat{\rho}_n) - \bar{X}_n)'\Gamma_n' A_n\epsilon_n = \sum_{k=1}^q \sqrt{n}(\rho_{0k} - \hat{\rho}_{nk})\frac{1}{n}(M_{kn}X_n)'\Gamma_n' A_n\epsilon_n = o_p(1)$, where $\sqrt{n}(\rho_{0k} - \hat{\rho}_{nk}) = O_p(1)$ and $\frac{1}{n}(M_{kn}X_n)'\Gamma_n' A_n\epsilon_n = o_p(1)$ by C.4 for $k = 1, \cdots, q$. Similarly, the remaining term in $\frac{1}{\sqrt{n}}(\hat{T}_{1n} - T_{1n})'A_n\epsilon_n$ is also $o_p(1)$. For the first two terms in $\frac{1}{\sqrt{n}}(\hat{T}_{2n} - T_{2n})'A_n\epsilon_n$, we have

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}(\hat{C}_n\bar{X}_n(\hat{\rho}_n)\hat{\beta}_n - C_n\bar{X}_n\beta_0)'\Gamma_n' A_n\epsilon_n \\
= \quad & \frac{1}{\sqrt{n}}[(\hat{C}_n - C_n)(R_n(\hat{\rho}_n) - R_n)X_n\hat{\beta}_n]'\Gamma_n' A_n\epsilon_n + \frac{1}{\sqrt{n}}[(\hat{C}_n - C_n)R_nX_n\hat{\beta}_n]'\Gamma_n' A_n\epsilon_n \\
& + \sum_{k=1}^q \sqrt{n}(\rho_{0k} - \hat{\rho}_{nk})\hat{\beta}_n'\frac{1}{n}(M_{kn}X_n)'C_n'\Gamma_n' A_n\epsilon_n + \sqrt{n}(\hat{\beta}_n - \beta_0)'\frac{1}{n}\bar{X}_n' C_n'\Gamma_n' A_n\epsilon_n. \quad (29)
\end{aligned}
$$

The first two terms of (29) are $o_p(1)$ by C.12. And the remaining terms of (29) are $o_p(1)$ because $\frac{1}{n}(M_{kn}X_n)'C_n'\Gamma_n' A_n\epsilon_n = o_p(1)$ and $\frac{1}{n}\bar{X}_n' C_n'\Gamma_n' A_n\epsilon_n = o_p(1)$ by C.4. Similarly, the remaining term in $\frac{1}{\sqrt{n}}(\hat{T}_{2n} - T_{2n})'A_n\epsilon_n$ is also $o_p(1)$. The first term in $\frac{1}{\sqrt{n}}(\hat{T}_{3n} - T_{3n})'A_n\epsilon_n$ is $o_p(1)$ by C.12, and the remaining term is also $o_p(1)$ because $\frac{1}{n}vec_D'(C_n^t)A_n\epsilon_n = o_p(1)$ by C.4. The desired results follow.

For (c), as the arguments are similar to those in the proof of (a), we only give the proof for $\frac{1}{n}(\hat{T}_{2n} - T_{2n})'\hat{D}_n = o_p(1)$. For its first two terms, we have

$$
\begin{aligned}
& \frac{1}{n}(\hat{C}_n\bar{X}_n(\hat{\rho}_n)\hat{\beta}_n - C_n\bar{X}_n\beta_0)'\Gamma_n'\hat{D}_n \\
= \quad & \frac{1}{n}[(\hat{C}_n - C_n)\bar{X}_n\hat{\beta}_n]'\Gamma_n'\hat{D}_n + [\sum_{k=1}^q(\rho_{0k} - \hat{\rho}_{nk})\hat{C}_nM_{kn}X_n\hat{\beta}_n + C_n\bar{X}_n(\hat{\beta}_n - \beta_0)]'\Gamma_n'\hat{D}_n, \quad (30)
\end{aligned}
$$

where the first term can be rewritten as $\frac{1}{n}[(\hat{C}_n - C_n)\bar{X}_n\hat{\beta}_n]'\Gamma_n'(\hat{D}_n - D_n) + \frac{1}{n}[(\hat{C}_n - C_n)\bar{X}_n\hat{\beta}_n]'\Gamma_n' D_n$, and it is $o_p(1)$ by C.12. The remaining term of (30) is also $o_p(1)$ because $\hat{\rho}_n - \rho_0 = o_p(1)$, $\hat{\beta}_n - \beta_0 = o_p(1)$, and $(\Gamma_n\hat{C}_nM_{kn}X_n)'\hat{D}_n = O_p(1)$, $(\Gamma_nC_n\bar{X}_n)'\hat{D}_n = O_p(1)$. Similarly, we have $(\hat{\mu}_{3n}^2/\hat{\kappa}_n - \mu_3^2/\kappa)(\Gamma_nC_n\bar{X}_n\beta_0)'\hat{D}_n = o_p(1)$. Hence, $\frac{1}{n}(\hat{T}_{2n} - T_{2n})'\hat{D}_n = o_p(1)$. ■

Proofs of Propositions 2 and 3. With the basic properties in C.1 - C.5 and our assumptions, the proofs of these two propositions will be similar to the arguments in Lee (2007a) and, hence, are omitted. ■

Proof of Proposition 4. Consider the moment conditions $E\left(g_n^*\left(\theta_0\right), g_n\left(\theta_0\right)\right)' = 0$, where $g_n\left(\theta\right)$

is a vector of arbitrary moment functions taking the form of (4). To show the desired results, it is sufficient to show that g_n is redundant given g_n^*, or equivalently that there exists an A_n invariant with P_{in} $(i=1,\cdots,m)$ and Q_n st. $D_2=\Omega_{21}A_n$ according to C.14 (c), where

$$\begin{aligned} D_2 &= \frac{\partial E(g_n(\theta_0))}{\partial\theta'} \\ &= -\begin{pmatrix} 0_{k_{IV}\times q} & Q_n'(\bar{G}_{1n}\bar{X}_n\beta_0,\cdots,\bar{G}_{pn}\bar{X}_n\beta_0) & Q_n'\bar{X}_n \\ \sigma_0^2(tr(P_{1n}^sH_{1n}),\cdots,tr(P_{1n}^sH_{qn})) & \sigma_0^2(tr(P_{1n}^s\bar{G}_{1n}),\cdots,tr(P_{1n}^s\bar{G}_{pn})) & 0_{1\times k_x} \\ \vdots & \vdots & \vdots \\ \sigma_0^2(tr(P_{mn}^sH_{1n}),\cdots,tr(P_{mn}^sH_{qn})) & \sigma_0^2(tr(P_{mn}^s\bar{G}_{1n}),\cdots,tr(P_{mn}^s\bar{G}_{pn})) & 0_{1\times k_x} \end{pmatrix}, \end{aligned}$$

and

$$\begin{aligned} \Omega_{21} &= E\left(g_n\left(\theta_0\right)g_n^{*\prime}\left(\theta_0\right)\right) \\ &= \begin{pmatrix} \sigma_0^2Q_n'Q_n^* & \mu_3Q_n'vec_D\left(P_{1n}^*\right) & \cdots & \mu_3Q_n'vec_D(P_{p+q+k_x^*,n}^*) \\ \mu_3vec_D'\left(P_{1n}\right)Q_n^* & \sigma_0^4tr(P_{1n}^sP_{1n}^*) & \cdots & \sigma_0^4tr(P_{1n}^sP_{p+q+k_x^*,n}^*) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_3vec_D'\left(P_{mn}\right)Q_n^* & \sigma_0^4tr(P_{mn}^sP_{1n}^*) & \cdots & \sigma_0^4tr(P_{mn}^sP_{p+q+k_x^*,n}^*) \end{pmatrix} \\ &\quad +(\mu_4-3\sigma_0^4)\begin{pmatrix} 0_{k_{IV}\times k_{IV}} & 0_{k_{IV}\times(p+q+k_x^*)} \\ 0_{1\times k_{IV}} & vec_D'\left(P_{1n}\right)\left(vec_D\left(P_{1n}^*\right),\cdots,vec_D(P_{p+q+k_x^*,n}^*)\right) \\ \vdots & \vdots \\ 0_{1\times k_{IV}} & vec_D'\left(P_{mn}\right)\left(vec_D\left(P_{1n}^*\right),\cdots,vec_D(P_{p+q+k_x^*,n}^*)\right) \end{pmatrix}. \end{aligned}$$

With $\kappa=\sigma_0^6\left[(\eta_4-1)-\eta_3^2\right]=\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2$, let

$$A_n=-\begin{pmatrix} 0_{q\times k_x} & 0_{q\times p} & -\frac{2\sigma_0^2\mu_3}{\kappa}I_q & 0_{q\times p} & \sigma_0^{-2}I_q & 0_{q\times k_x^*} \\ 0_{p\times k_x} & \sigma_0^{-2}I_p & 0_{p\times q} & \sigma_0^{-2}I_p & 0_{p\times q} & 0_{p\times k_x^*} \\ \sigma_0^{-2}I_{k_x} & 0_{k_x\times p} & 0_{k_x\times q} & 0_{k_x\times p} & 0_{k_x\times q} & b \end{pmatrix}',$$

where $b=(b_1',\cdots,b_{k_x^*}')'$ with $b_l=-\left(\mu_3/\kappa\right)e_{k_xl}'$ for $l=1,\cdots,k_x^*$. To check $D_2=\Omega_{21}A_n$, the following identities are helpful. For $j=1,\cdots,p$, $k=1,\cdots,q$, and $l=1,\cdots,k_x^*$, (a) $vec_D(P_{jn}^*)=\frac{2\sigma_0^6}{\kappa}vec_D(\bar{G}_{jn}-\frac{tr(\bar{G}_{jn})}{n}I_n)-\frac{\sigma_0^2\mu_3}{\kappa}(\bar{G}_{jn}\bar{X}_n\beta_0-\frac{1}{n}l_nl_n'\bar{G}_{jn}\bar{X}_n\beta_0)$; (b) $vec_D(P_{p+k,n}^*)=\frac{2\sigma_0^6}{\kappa}vec_D(H_{kn}-\frac{tr(H_{kn})}{n}I_n)$; (c) $vec_D(P_{p+q+l,n}^*)=\bar{X}_{nl}^*-\frac{1}{n}l_nl_n'\bar{X}_{nl}^*$; and (d) $\sum_{l=1}^{k_x^*}vec_D(P_{p+q+l,n}^*)e_{k_xl}'=\bar{X}_n-\frac{1}{n}l_nl_n'\bar{X}_n$.

It follows from (a), (b) and (d), respectively, to have that (e) $\sigma_0^2 Q^*_{2nj} + \mu_3 vec_D(P^*_{jn}) = \sigma_0^2 \bar{G}_{jn}\bar{X}_n\beta_0$; (f) $\frac{2\sigma_0^6}{\kappa} Q^*_{3nk} = vec_D(P^*_{p+k,n})$, and (g) $Q^*_{1n} - \frac{\mu_3^2}{\kappa}\sum_{l=1}^{k^*_x} vec_D(P^*_{p+q+l,n})e'_{k_x l} = \bar{X}_n$. For an arbitrary $n \times n$ matrix P_n with $tr\,(P_n) = 0$, we have: (h) $vec'_D(P_n)Q^*_{1n} = (\sigma_0^2(\mu_4 - \sigma_0^4)/\kappa)vec'_D(P_n)\bar{X}_n$; (i) $\mu_3 vec'_D(P_n)Q^*_{2nj} + \sigma_0^4 tr(P^s_n P^*_{jn}) + (\mu_4 - 3\sigma_0^4)vec'_D(P_n)vec_D(P^*_{jn}) = \sigma_0^4 tr(P^s_n \bar{G}_{jn})$; (j) $-\frac{2\sigma_0^4\mu_3^2}{\kappa} vec'_D(P_n)Q^*_{3nk} + \sigma_0^4 tr(P^s_n P^*_{p+k,n}) + (\mu_4 - 3\sigma_0^4)vec'_D(P_n)vec_D(P^*_{p+k,n}) = \sigma_0^4 tr(P^s_n H_{kn})$; and (k) $\sigma_0^4 tr(P^s_n P^*_{p+q+l,n}) + (\mu_4 - 3\sigma_0^4)vec'_D(P_n)vec_D(P^*_{p+q+l,n}) = (\mu_4 - \sigma_0^4)vec'_D(P_n)vec_D(P^*_{p+q+l,n})$.

It follows from identity (f) the $(1,1)$ block of $\Omega_{21}A_n$ is $0_{k_{IV}\times q}$, it follows from identity (e) that the $(1,2)$ block of $\Omega_{21}A_n$ is $-Q'_n(\bar{G}_{1n}\bar{X}_n\beta_0, \cdots, \bar{G}_{pn}\bar{X}_n\beta_0)$, and it follows from identity (g) that the $(1,3)$ block of $\Omega_{21}A_n$ is $-Q'_n\bar{X}_n$. Identity (j) implies that the $(i+1,1)$ blocks of $\Omega_{21}A_n$ are $-\sigma_0^2(tr(P^s_{in}H_{1n}), \cdots, tr(P^s_{in}H_{qn}))$ for $i = 1, \cdots, m$, identity (i) implies that the $(i+1,2)$ blocks of $\Omega_{21}A_n$ are $-\sigma_0^2(tr(P^s_{in}\bar{G}_{1n}), \cdots, tr(P^s_{in}\bar{G}_{pn}))$ for $i = 1, \cdots, m$, and identities (d), (h) and (k) imply that the remaining blocks of $\Omega_{21}A_n$ are zeros. Therefore, $\Omega_{21}A_n = D_2$.

Furthermore, as $g^*_n(\theta)$ is a special case of $g_n(\theta)$, and A_n is invariant with P_n's and Q_n, it follows that $D_1 = \Omega_{11}A_n$, and hence $\Omega_{11}^{-1}D_1 = A_n$, where $\Omega_{11} = \Omega^*_n = var\,(g^*_n(\theta_0))$ and $D_1 = E\left(\frac{\partial g^*_n(\theta_0)}{\partial\theta}\right)$. Hence $\Sigma_b = \lim_{n\to\infty}\frac{1}{n}D'_1\Omega_{11}^{-1}D_1 = \lim_{n\to\infty}\frac{1}{n}D'_1A_n$. After some tedious but straightforward algebra, the desired result follows. ∎

Proof of Proposition 5. We shall show that $\digamma^*_n(\theta) = \hat{g}^{*\prime}_n(\theta)\hat{\Omega}^{*-1}_n\hat{g}^*_n(\theta)$ and $\digamma_n(\theta) = g^{*\prime}_n(\theta)\Omega^{*-1}_n g^*_n(\theta)$ will satisfy the conditions in C.6. If so, the GMM estimator from the minimization of $\digamma^*_n(\theta)$ will have the same limiting distribution as that of the minimization of $\digamma_n(\theta)$. The difference of $\digamma^*_n(\theta)$ and $\digamma_n(\theta)$ and its derivatives involve the difference of $\hat{g}^*_n(\theta)$ and $g^*_n(\theta)$ and their derivatives. Furthermore, one has to consider the difference of $\hat{\Omega}^*_n$ and Ω^*_n.

First, consider $\frac{1}{n}(\hat{g}^*_n(\theta) - g^*_n(\theta))$. Let $m^* = k^*_x + p + q$. Explicitly,

$$\frac{1}{n}(\hat{g}^*_n(\theta) - g^*_n(\theta))' = [\frac{1}{n}(\hat{Q}^*_n - Q^*_n)', \frac{1}{n}\epsilon'_n(\theta)(\hat{P}^*_{1n} - P^*_{1n}), \cdots, \frac{1}{n}\epsilon'_n(\theta)(\hat{P}^*_{m^*n} - P^*_{m^*n})]\epsilon_n(\theta).$$

The $\epsilon_n(\theta)$ is related to ϵ_n as $\epsilon_n(\theta) = e_n(\theta) + (I_n + \sum_{k=1}^q(\rho_{0k} - \rho_k)H_{kn})(I_n + \sum_{j=1}^p(\lambda_{0j} - \lambda_j)\bar{G}_{jn})\epsilon_n$ where $e_n(\theta) = (I_n + \sum_{k=1}^q(\rho_{0k} - \rho_k)H_{kn})[\sum_{j=1}^p(\lambda_{0j} - \lambda_j)\bar{G}_{jn}\bar{X}_n\beta_0 + \bar{X}_n(\beta_0 - \beta)]$. It follows that $\frac{1}{n}(\hat{Q}^*_n - Q^*_n)'\epsilon_n(\theta) = \frac{1}{n}(\hat{Q}^*_n - Q^*_n)'(I_n + \sum_{k=1}^q(\rho_{0k} - \rho_k)H_{kn})(I_n + \sum_{j=1}^p(\lambda_{0j} - \lambda_j)\bar{G}_{jn})\epsilon_n + \frac{1}{n}(\hat{Q}^*_n - Q^*_n)'e_n(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$ by C.13. From C.12, it follows that $\frac{1}{n}\epsilon'_n(\theta)(\hat{P}^*_{in} - P^*_{in})\epsilon_n(\theta) = o_p(1)$, for $i = 1, \cdots, m^*$, uniformly in $\theta \in \Theta$. Hence, we conclude that $\frac{1}{n}(\hat{g}^*_n(\theta) - g^*_n(\theta)) = o_p(1)$ uniformly in $\theta \in \Theta$.

Consider the derivatives of $\hat{g}_n^*(\theta)$ and $g_n^*(\theta)$:

$$\frac{\partial g_n^*(\theta)}{\partial\theta'} = \begin{pmatrix} Q_n^{*\prime}\frac{\partial\epsilon_n(\theta)}{\partial\theta'} \\ \epsilon_n'(\theta)P_{1n}^{*s}\frac{\partial\epsilon_n(\theta)}{\partial\theta'} \\ \vdots \\ \epsilon_n'(\theta)P_{m^*n}^{*s}\frac{\partial\epsilon_n(\theta)}{\partial\theta'} \end{pmatrix}, \text{ and } \frac{\partial^2 g_n^*(\theta)}{\partial\theta\partial\theta'} = \begin{pmatrix} Q_n^{*\prime}\frac{\partial^2\epsilon_n(\theta)}{\partial\theta\partial\theta'} \\ \frac{\partial\epsilon_n'(\theta)}{\partial\theta}P_{1n}^{*s}\frac{\partial\epsilon_n(\theta)}{\partial\theta'} + \epsilon_n'(\theta)P_{1n}^{*s}\frac{\partial^2\epsilon_n(\theta)}{\partial\theta\partial\theta'} \\ \vdots \\ \frac{\partial\epsilon_n'(\theta)}{\partial\theta}P_{m^*n}^{*s}\frac{\partial\epsilon_n(\theta)}{\partial\theta'} + \epsilon_n'(\theta)P_{m^*n}^{*s}\frac{\partial^2\epsilon_n(\theta)}{\partial\theta\partial\theta'} \end{pmatrix}.$$

The first order derivatives of $\epsilon_n(\theta)$ are $\frac{\partial\epsilon_n(\theta)}{\partial\theta'} = -[M_{1n}u_n(\delta), \cdots, M_{qn}u_n(\delta), R_n(\rho)W_{1n}Y_n, \cdots, R_n(\rho)W_{pn}Y_n, R_n(\rho)X_n]$, where $u_n(\delta) = (I_n - \sum_{j=1}^p \lambda_j W_{jn})Y_n - X_n\beta$, $R_n(\rho) = I_n - \sum_{k=1}^q \rho_k M_{kn}$ and $Y_n = S_n^{-1}X_n\beta_0 + S_n^{-1}R_n^{-1}\epsilon_n$. The second derivatives of $\epsilon_n(\theta)$ are $\frac{\partial^2\epsilon_n(\theta)}{\partial\lambda_j\partial\rho_k} = M_{kn}W_{jn}Y_n$, $\frac{\partial^2\epsilon_n(\theta)}{\partial\rho_k\partial\beta'} = M_{kn}X_n$ and $\frac{\partial^2\epsilon_n(\theta)}{\partial\lambda_j\partial\beta'} = 0$. It follows from C.12 and C.13 that $\frac{1}{n}(\frac{\partial\hat{g}_n^*(\theta)}{\partial\theta} - \frac{\partial g_n^*(\theta)}{\partial\theta}) = o_p(1)$ and $\frac{1}{n}(\frac{\partial^2\hat{g}_n^*(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 g_n^*(\theta)}{\partial\theta\partial\theta'}) = o_p(1)$ uniformly in $\theta \in \Theta$.

Consider $\frac{1}{n}(\hat{\Omega}_n^* - \Omega_n^*)$, where $\Omega_n^* = E\left[g_n^*(\theta_0)\, g_n^{*\prime}(\theta_0)\right] = \begin{pmatrix} \sigma_0^2 Q_n^{*\prime}Q_n^* & \mu_3 Q_n^{*\prime}\omega_{m^*}^* \\ \mu_3\omega_{m^*}^{*\prime}Q_n^* & \sigma_0^4\Delta_{m^*}^* + (\mu_4 - 3\sigma_0^4)\omega_{m^*}^{*\prime}\omega_{m^*}^* \end{pmatrix}$, with $\omega_{m^*}^* = [vec_D(P_{1n}^*), \cdots, vec_D(P_{m^*n}^*)]$ and

$$\Delta_{m^*}^* = \begin{pmatrix} tr\,(P_{1n}^{*s}P_{1n}^*) & \cdots & tr\,(P_{1n}^{*s}P_{m^*n}^*) \\ \vdots & \ddots & \vdots \\ tr\,(P_{m^*n}^{*s}P_{1n}^*) & \cdots & tr\,(P_{m^*n}^{*s}P_{m^*n}^*) \end{pmatrix}.$$

First, consider the block matrix $\sigma_0^4\Delta_{m^*}^* + (\mu_4 - 3\sigma_0^4)\omega_{m^*}^{*\prime}\omega_{m^*}^*$. It follows from C.12 that $\frac{1}{n}tr(\hat{P}_{in}^{*s}\hat{P}_{jn}^*) - \frac{1}{n}tr(P_{in}^{*s}P_{jn}^*) = \frac{1}{n}tr[(\hat{P}_{in}^{*s} - P_{in}^{*s})\hat{P}_{jn}^* + P_{in}^{*s}(\hat{P}_{jn}^* - P_{jn}^*)]$ and $\frac{1}{n}vec_D'(\hat{P}_{in}^*)vec_D(\hat{P}_{jn}^*) - \frac{1}{n}vec_D'(P_{in}^*)\,vec_D(P_{jn}^*) = \frac{1}{n}vec_D'(\hat{P}_{in}^*)vec_D(\hat{P}_{jn}^* - P_{jn}^*) + \frac{1}{n}vec_D'(\hat{P}_{in}^* - P_{in}^*)vec_D(P_{jn}^*)$ are $o_p(1)$ for $i, j = 1, \cdots, m^*$. Hence, $\frac{1}{n}(\hat{\sigma}_n^2)^2 tr(\hat{P}_{in}^{*s}\hat{P}_{jn}^*) - \frac{1}{n}\sigma_0^4 tr(P_{in}^{*s}P_{jn}^*)$ and $\frac{1}{n}(\hat{\mu}_{4n} - 3(\hat{\sigma}_n^2)^2)vec_D'(\hat{P}_{in}^*)vec_D(\hat{P}_{jn}^*) - \frac{1}{n}(\mu_4 - 3\sigma_0^4)vec_D'(P_{in}^*)vec_D(P_{jn}^*)$ are $o_p(1)$ for $i, j = 1, \cdots, m^*$.

Next consider the block matrix $\mu_3 Q_n^{*\prime}\omega_{m^*}^*$. It follows from C.12 and C.13 that $\frac{1}{n}\hat{Q}_n^{*\prime}vec_D(\hat{P}_{in}^*) - \frac{1}{n}Q_n^{*\prime}vec_D(P_{in}^*) = \frac{1}{n}(\hat{Q}_n^* - Q_n^*)'vec_D(\hat{P}_{in}^*) + \frac{1}{n}Q_n^{*\prime}vec_D(\hat{P}_{in}^* - P_{in}^*) = o_p(1)$, for $i = 1, \cdots, m^*$. Hence, $\frac{1}{n}(\hat{\mu}_{3n}\hat{Q}_n^{*\prime}vec_D(\hat{P}_{in}^*) - \mu_3 Q_n^{*\prime}vec_D(P_{in}^*)) = \hat{\mu}_{3n}\frac{1}{n}(\hat{Q}_n^{*\prime}vec_D(\hat{P}_{in}^*) - Q_n^{*\prime}vec_D(P_{in}^*)) + (\hat{\mu}_{3n} - \mu_3)\frac{1}{n}Q_n^{*\prime}vec_D(P_{in}^*) = o_p(1)$, for $i = 1, \cdots, m^*$.

Lastly, consider the remaining block matrix $\sigma_0^2 Q_n^{*\prime}Q_n^*$. C.13 implies that $\frac{1}{n}(\hat{Q}_{in}^{*\prime}\hat{Q}_{jn}^* - Q_{in}^{*\prime}Q_{jn}^*) = \frac{1}{n}[\hat{Q}_{in}^{*\prime}(\hat{Q}_{jn}^* - Q_{jn}^*) + (\hat{Q}_{in}^* - Q_{in}^*)'Q_{jn}^*] = o_p(1)$, for $i, j = 1, 2, 3$. Therefore, it follows that $\frac{1}{n}(\hat{\sigma}_n^2\hat{Q}_n^{*\prime}\hat{Q}_n^* - \sigma_0^2 Q_n^{*\prime}Q_n^*) = \hat{\sigma}_n^2\frac{1}{n}(\hat{Q}_n^{*\prime}\hat{Q}_n^* - Q_n^{*\prime}Q_n^*) + (\hat{\sigma}_n^2 - \sigma_0^2)\frac{1}{n}Q_n^{*\prime}Q_n^* = o_p(1)$. In conclusion, $\frac{1}{n}\hat{\Omega}_n^* - \frac{1}{n}\Omega_n^* = o_p(1)$.

As the limit of $\frac{1}{n}\Omega_n^*$ exists and is a nonsingular matrix (as the moments are not linearly dependent with probability 1), it follows that $(\frac{1}{n}\hat{\Omega}_n^*)^{-1} - (\frac{1}{n}\Omega_n^*)^{-1} = o_p(1)$ by the continuous mapping theorem (White, 1984, Proposition 2.30).

Furthermore, because $\frac{1}{n}(\hat{g}_n^*(\theta) - g_n^*(\theta)) = o_p(1)$, and $\frac{1}{n}[g_n^*(\theta) - E(g_n^*(\theta))] = o_p(1)$ uniformly in $\theta \in \Theta$, and $\sup_{\theta\in\Theta} \frac{1}{n}|E(g_n^*(\theta))| = O(1)$ (see proof of Proposition 3), hence $\frac{1}{n}g_n^*(\theta)$ and $\frac{1}{n}\hat{g}_n^*(\theta)$ are $O_p(1)$, uniformly in $\theta \in \Theta$. Similarly, $\frac{1}{n}\frac{\partial g_n^*(\theta)}{\partial\theta}$, $\frac{1}{n}\frac{\partial \hat{g}_n^*(\theta)}{\partial\theta}$, $\frac{1}{n}\frac{\partial^2 g_n^*(\theta)}{\partial\theta\partial\theta}$ and $\frac{1}{n}\frac{\partial^2 \hat{g}_n^*(\theta)}{\partial\theta\partial\theta}$ are $O_p(1)$, uniformly in $\theta \in \Theta$.

With the uniform convergence in probability and uniformly stochastic boundedness properties, the difference of $F_n^*(\theta)$ and $F_n(\theta)$ can be investigated. By expansion, $\frac{1}{n}(F_n^*(\theta) - F_n(\theta)) = \frac{1}{n}\hat{g}_n^{*\prime}(\theta)\hat{\Omega}_n^{*-1}(\hat{g}_n^*(\theta) - g_n^*(\theta)) + \frac{1}{n}g_n^{*\prime}(\theta)(\hat{\Omega}_n^{*-1} - \Omega_n^{*-1})\hat{g}_n^*(\theta) + \frac{1}{n}g_n^{*\prime}(\theta)\Omega_n^{*-1}(\hat{g}_n^*(\theta) - g_n^*(\theta)) = o_p(1)$, uniformly in $\theta \in \Theta$. Similarly, for each component θ_l of θ, $\frac{1}{n}\frac{\partial^2 F_n^*(\theta)}{\partial\theta_l\partial\theta'} - \frac{1}{n}\frac{\partial^2 F_n(\theta)}{\partial\theta_l\partial\theta'} = \frac{2}{n}[\frac{\partial \hat{g}_n^{*\prime}(\theta)}{\partial\theta_l}\hat{\Omega}_n^{*-1}\frac{\partial \hat{g}_n^*(\theta)}{\partial\theta'} + \hat{g}_n^{*\prime}(\theta)\hat{\Omega}_n^{*-1}\frac{\partial^2 \hat{g}_n^*(\theta)}{\partial\theta_l\partial\theta'} - (\frac{\partial g_n^{*\prime}(\theta)}{\partial\theta_l}\Omega_n^{*-1}\frac{\partial g_n^*(\theta)}{\partial\theta'} + g_n^{*\prime}(\theta)\Omega_n^{*-1}\frac{\partial^2 g_n^*(\theta)}{\partial\theta_l\partial\theta'})] = o_p(1)$.

Finally, because $(\frac{\partial \hat{g}_n^{*\prime}(\theta_0)}{\partial\theta}\hat{\Omega}_n^{*-1} - \frac{\partial g_n^{*\prime}(\theta_0)}{\partial\theta}\Omega_n^{*-1}) = o_p(1)$ as above, and $\frac{1}{\sqrt{n}}g_n^*(\theta_0) = O_p(1)$ by the central limit theorems in C.4 and C.5, $\frac{1}{\sqrt{n}}(\frac{\partial F_n^*(\theta_0)}{\partial\theta} - \frac{\partial F_n(\theta_0)}{\partial\theta}) = 2\{\frac{\partial \hat{g}_n^{*\prime}(\theta_0)}{\partial\theta}\hat{\Omega}_n^{*-1}\frac{1}{\sqrt{n}}(\hat{g}_n^*(\theta_0) - g_n^*(\theta_0)) + (\frac{\partial \hat{g}_n^{*\prime}(\theta_0)}{\partial\theta}\hat{\Omega}_n^{*-1} - \frac{\partial g_n^{*\prime}(\theta_0)}{\partial\theta}\Omega_n^{*-1})\frac{1}{\sqrt{n}}g_n^*(\theta_0)\} = 2\frac{1}{n}\frac{\partial \hat{g}_n^{*\prime}(\theta_0)}{\partial\theta}(\frac{1}{n}\hat{\Omega}_n^*)^{-1}\frac{1}{\sqrt{n}}(\hat{g}_n^*(\theta_0) - g_n^*(\theta_0)) + o_p(1)$. As $\frac{1}{\sqrt{n}}(\hat{g}_n^*(\theta_0) - g_n^*(\theta_0)) = o_p(1)$ by C.12 and C.13, $\frac{1}{\sqrt{n}}(\frac{\partial F_n^*(\theta_0)}{\partial\theta} - \frac{\partial F_n(\theta_0)}{\partial\theta}) = o_p(1)$. The desired result follows from C.6. ■

Proof of Proposition 6. The log-likelihood function for the SARAR(p, q) model is given by (2) and its derivatives are $\frac{\partial \ln L_n}{\partial\beta} = \frac{1}{\sigma^2}(R_n(\rho)X_n)'\epsilon_n(\theta) = \frac{1}{\sigma^2}\bar{X}_n'(\rho)\epsilon_n(\theta)$, $\frac{\partial \ln L_n}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\epsilon_n'(\theta)\epsilon_n(\theta)$, $\frac{\partial \ln L_n}{\partial\lambda_j} = -tr(\bar{G}_{jn}(\rho,\lambda)) + \frac{1}{\sigma^2}\{\bar{G}_{jn}(\rho,\lambda)\bar{X}_n(\rho)\beta\}'\epsilon_n(\theta) + \frac{1}{\sigma^2}\epsilon_n'(\theta)\bar{G}_{jn}(\rho,\lambda)\epsilon_n(\theta)$, $\frac{\partial \ln L_n}{\partial\rho_k} = -tr(H_{kn}(\rho)) + \frac{1}{\sigma^2}\epsilon_n'(\theta)H_{kn}\epsilon_n(\theta)$, where $\bar{Y}_n(\rho) = R_n(\rho)Y_n$, $\bar{X}_n(\rho) = R_n(\rho)X_n$, $\bar{W}_{jn}(\rho) = R_n(\rho)W_{jn}R_n^{-1}(\rho)$, $\bar{S}_n(\rho,\lambda) = R_n(\rho)S_n(\lambda)R_n^{-1}(\rho)$, and $\bar{G}_{jn}(\rho,\lambda) = \bar{W}_{jn}(\rho)\bar{S}_{jn}^{-1}(\lambda)$, for $j = 1, \cdots, p$ and $k = 1, \cdots, q$. The QML estimator of σ^2 is $\hat{\sigma}^2_{ml,n}(\theta) = \frac{1}{n}\epsilon_n'(\theta)\epsilon_n(\theta)$ for a given value θ. Substitution of $\hat{\sigma}^2_{ml,n}(\theta)$ in the remaining likelihood equations shows that the QML estimator is characterized by the equations $\bar{X}_n'(\rho)\epsilon_n(\theta) = 0$, $[\bar{G}_{jn}(\rho,\lambda)\bar{X}_n(\rho)\beta]'\epsilon_n(\theta) + \epsilon_n'(\theta)[\bar{G}_{jn}(\rho,\lambda) - \frac{1}{n}tr(\bar{G}_{jn}(\rho,\lambda))]\epsilon_n(\theta) = 0$, and $\epsilon_n'(\theta)[H_{kn}(\rho) - \frac{1}{n}tr(H_{kn}(\rho))]\epsilon_n(\theta) = 0$, for $j = 1, \cdots, p$ and $k = 1, \cdots, q$. Denote the QML estimator of θ by $\hat{\theta}_{ml,n}$. Obviously $\hat{\theta}_{ml,n}$ is the solution of $a_n\hat{g}_{ml,n}(\theta) = 0$, with $a_n = \begin{pmatrix} I_{k_x} & 0 & 0 & 0 \\ 0 & I_p & I_p & 0 \\ 0 & 0 & 0 & I_q \end{pmatrix}$ and $\hat{g}_{ml,n}(\theta) = [\bar{X}_n(\hat{\rho}_{ml,n}), \bar{G}_{1n}(\hat{\rho}_{ml,n}, \hat{\lambda}_{ml,n})\bar{X}_n(\hat{\rho}_{ml,n})\hat{\beta}_{ml,n}, \cdots, \bar{G}_{pn}(\hat{\rho}_{ml,n}, \hat{\lambda}_{ml,n})\bar{X}_n(\hat{\rho}_{ml,n})\hat{\beta}_{ml,n}, \bar{G}^t_{1n}(\hat{\rho}_{ml,n}, \hat{\lambda}_{ml,n})\epsilon_n(\theta), \cdots, \bar{G}^t_{pn}(\hat{\rho}_{ml,n}, \hat{\lambda}_{ml,n})\epsilon_n(\theta), H^t_{1n}(\hat{\rho}_{ml,n})\epsilon_n(\theta), \cdots, H^t_{qn}(\hat{\rho}_{ml,n})\epsilon_n(\theta)]'\epsilon_n(\theta)$, where

$A^t = A - \frac{1}{n} tr(A) I_n$ for a square matrix A. And it follows by similar arguments as in the proof Proposition 5 that $a_n \hat{g}_{ml,n}(\theta) = 0$ is asymptotically equivalent to $a_n g_{ml,n}(\theta) = 0$, where $g_{ml,n}(\theta) = [\bar{X}_n, \bar{G}_{1n}\bar{X}_n\hat{\beta}_{ml,n}, \cdots, \bar{G}_{pn}\bar{X}_n\hat{\beta}_{ml,n}, \bar{G}^t_{1n}\epsilon_n(\theta), \cdots, \bar{G}^t_{pn}\epsilon_n(\theta), H^t_{1n}\epsilon_n(\theta), \cdots, H^t_{qn}\epsilon_n(\theta)]'\epsilon_n(\theta)$, in the sense that their consistent roots have the same limiting distribution. The vector of empirical moments $g_{ml,n}(\theta)$ consists of linear and quadratic functions of $\epsilon_n(\theta)$, hence the corresponding optimal GMM estimator derived from $\min g'_{ml,n}(\theta)\Omega_n^{-1} g_{ml,n}(\theta)$ is in the class $\mathcal{M}_n$. As the BGMM estimator is the most efficient estimator in $\mathcal{M}_n$, the BGMM estimator is efficient relative to the QML estimator.

The best G2SLS estimator of Lee (2003) is consistent and asymptotic normal with $\sqrt{n}(\hat{\delta}_{bg2sls,n} - \delta_0) \xrightarrow{D} N(0, \sigma_0^2(\lim_{n\to\infty} \frac{1}{n}\Lambda'_n R'_n R_n \Lambda_n)^{-1})$, where $\Lambda_n = (G_{1n}X_n\beta_0, \cdots, G_{pn}X_n\beta_0, X_n)$. The asymptotic variance of the best G2SLS estimator can be easily compared with the asymptotic variance of the BGMM estimator in $\mathcal{P}_{2n}$. With the best P_n's in $\mathcal{P}_{2n}$, the asymptotic variance of $\hat{\theta}_{b,n}$ is the inverse of the asymptotic precision matrix in (14). By the inverse formula of a partitioned matrix, the corresponding asymptotic variance of the component $\hat{\delta}_{b,n}$ of $\hat{\theta}_{b,n}$ is

$$\left(\frac{C_n}{\sigma_0^2} + \begin{pmatrix} (A_n B_n^{-1} A'_n)_{22} - (A_n B_n^{-1} A'_n)_{21} (A_n B_n^{-1} A'_n)_{11}^{-1} (A_n B_n^{-1} A'_n)_{12} & 0_{p\times k_x} \\ 0_{k_x \times p} & 0_{k_x\times k_x} \end{pmatrix} \right)^{-1},$$

where $(A_n B_n^{-1} A'_n)_{11}$ is the first $q \times q$ diagonal block in $A_n B_n^{-1} A'_n$, $(A_n B_n^{-1} A'_n)_{22}$ is the other $p \times p$ diagonal block in $A_n B_n^{-1} A'_n$, and $(A_n B_n^{-1} A'_n)_{21}$ and $(A_n B_n^{-1} A'_n)_{12}$ are, respectively, the $p \times q$ lower block and the $q \times p$ upper block in $A_n B_n^{-1} A'_n$. In $\mathcal{P}_{2n}$, the best selection of IVs is given by $Q_n^* = R_n \Lambda_n$ and, hence, $C_n = \Lambda'_n R'_n R_n \Lambda_n$. As $A_n B_n^{-1} A_n$ is nonnegative definite, the asymptotic variance of the BGMM estimator in $\mathcal{P}_{2n}$ is relatively smaller than the asymptotic variance of the best G2SLS estimator. As $\mathcal{P}_{1n}$ is a broader class containing $\mathcal{P}_{2n}$, the BGMM estimator in $\mathcal{P}_{1n}$ given in Proposition 4 is therefore efficient relative to the best G2SLS estimator. ■

Table 1: Estimation of the SARAR(1,1) model with strong x's (normal)

	$\lambda_0 = 0.4$	$\rho_0 = 0.4$	$\beta_{10} = 1.0$	$\beta_{20} = -1.0$
	$n = 245$			
G2SLS	.412(.137)[.138]	.351(.154)[.162]	.995(.087)[.087]	−.998(.092)[.092]
B2SLS	.387(.159)[.160]	.351(.154)[.162]	.992(.091)[.091]	−.996(.092)[.092]
QML	.389(.135)[.136]	.383(.153)[.154]	.993(.087)[.087]	−.996(.092)[.092]
GMM1	.387(.136)[.137]	.393(.152)[.153]	.993(.087)[.087]	−.996(.093)[.093]
GMM2	.387(.136)[.137]	.392(.152)[.152]	.993(.087)[.088]	−.996(.092)[.093]
BGMM	.384(.149)[.150]	.400(.162)[.162]	.992(.089)[.089]	−.995(.095)[.095]
	$n = 490$			
G2SLS	.400(.096)[.096]	.381(.110)[.112]	.998(.062)[.063]	−.996(.064)[.064]
B2SLS	.389(.095)[.096]	.381(.110)[.112]	.997(.063)[.063]	−.995(.064)[.064]
QML	.388(.096)[.096]	.398(.106)[.106]	.997(.062)[.063]	−.995(.064)[.064]
GMM1	.386(.096)[.097]	.403(.105)[.105]	.997(.063)[.063]	−.995(.064)[.064]
GMM2	.386(.096)[.097]	.403(.105)[.105]	.997(.063)[.063]	−.995(.064)[.064]
BGMM	.385(.099)[.100]	.406(.107)[.107]	.996(.064)[.064]	−.994(.064)[.064]

Mean(SD)[RMSE]

Table 2: Estimation of the SARAR(1,1) model with strong x's (symmetric mixture normal)

	$\lambda_0 = 0.4$	$\rho_0 = 0.4$	$\beta_{10} = 1.0$	$\beta_{20} = -1.0$
	$n = 245$			
G2SLS	.413(.135)[.136]	.350(.155)[.163]	.993(.089)[.089]	−1.001(.090)[.090]
B2SLS	.390(.137)[.137]	.350(.155)[.163]	.991(.089)[.090]	−.999(.091)[.091]
QML	.391(.134)[.135]	.383(.153)[.154]	.991(.089)[.089]	−1.000(.090)[.090]
GMM1	.389(.134)[.135]	.393(.149)[.149]	.991(.089)[.090]	−.999(.090)[.090]
GMM2	.389(.133)[.133]	.392(.148)[.148]	.991(.089)[.089]	−.999(.090)[.090]
BGMM	.384(.137)[.138]	.401(.147)[.148]	.991(.085)[.085]	−.998(.084)[.085]
	$n = 490$			
G2SLS	.404(.094)[.094]	.378(.107)[.109]	.998(.064)[.064]	−1.000(.063)[.063]
B2SLS	.394(.093)[.093]	.378(.107)[.109]	.998(.064)[.064]	−1.000(.063)[.063]
QML	.394(.094)[.094]	.392(.105)[.105]	.998(.063)[.063]	−1.000(.063)[.063]
GMM1	.392(.095)[.095]	.398(.104)[.104]	.997(.064)[.064]	−.999(.063)[.063]
GMM2	.392(.094)[.094]	.398(.104)[.104]	.997(.064)[.064]	−.999(.063)[.063]
BGMM	.392(.093)[.093]	.400(.103)[.103]	.997(.061)[.061]	−.998(.061)[.061]

Mean(SD)[RMSE]

Table 3: Estimation of the SARAR(1,1) model with strong x's (gamma)

	$\lambda_0 = 0.4$	$\rho_0 = 0.4$	$\beta_{10} = 1.0$	$\beta_{20} = -1.0$
	$n = 245$			
G2SLS	.411(.133)[.133]	.354(.154)[.161]	.998(.087)[.087]	−.996(.094)[.094]
B2SLS	.383(.150)[.151]	.354(.154)[.161]	.995(.090)[.091]	−.995(.096)[.096]
QML	.383(.137)[.138]	.389(.155)[.156]	.995(.087)[.087]	−.993(.094)[.094]
GMM1	.380(.139)[.141]	.400(.151)[.151]	.995(.088)[.088]	−.993(.095)[.095]
GMM2	.380(.141)[.143]	.400(.154)[.154]	.994(.088)[.089]	−.993(.095)[.095]
BGMM	.385(.121)[.122]	.402(.139)[.139]	.997(.069)[.069]	−.994(.073)[.073]
	$n = 490$			
G2SLS	.411(.092)[.092]	.373(.109)[.112]	.995(.064)[.064]	−.996(.063)[.063]
B2SLS	.400(.093)[.093]	.373(.109)[.112]	.995(.064)[.064]	−.995(.063)[.063]
QML	.399(.095)[.095]	.388(.108)[.109]	.994(.064)[.064]	−.995(.063)[.063]
GMM1	.398(.094)[.094]	.394(.105)[.106]	.994(.064)[.064]	−.995(.063)[.063]
GMM2	.398(.095)[.095]	.393(.107)[.107]	.994(.064)[.065]	−.995(.063)[.063]
BGMM	.397(.073)[.073]	.399(.091)[.091]	.996(.049)[.049]	−.996(.048)[.048]

Mean(SD)[RMSE]

Table 4: Estimation of the SARAR(1,1) model with weak x's (n=490)

	$\lambda_0 = 0.4$	$\rho_0 = 0.4$	$\beta_{01} = 0.4$	$\beta_{02} = -0.4$
	normal			
G2SLS	.449(.243)[.247]	.282(.245)[.272]	.395(.064)[.065]	−.393(.065)[.065]
B2SLS	.346(.296)[.301]	.432(.245)[.247]	.394(.066)[.066]	−.392(.066)[.066]
QML	.407(.203)[.203]	.347(.216)[.223]	.396(.063)[.063]	−.394(.063)[.064]
GMM1	.352(.205)[.211]	.411(.199)[.199]	.395(.063)[.063]	−.393(.064)[.064]
GMM2	.352(.205)[.211]	.411(.199)[.200]	.395(.063)[.063]	−.393(.063)[.064]
BGMM	.344(.224)[.231]	.416(.214)[.215]	.393(.064)[.065]	−.392(.064)[.065]
	symmetric mixture normal			
G2SLS	.453(.248)[.254]	.276(.239)[.269]	.394(.065)[.065]	−.397(.064)[.064]
B2SLS	.361(.269)[.272]	.427(.237)[.238]	.394(.066)[.067]	−.397(.066)[.067]
QML	.417(.200)[.201]	.337(.214)[.223]	.397(.063)[.063]	−.399(.063)[.063]
GMM1	.368(.205)[.207]	.397(.200)[.200]	.395(.064)[.064]	−.397(.063)[.063]
GMM2	.371(.198)[.200]	.395(.194)[.194]	.395(.064)[.064]	−.398(.063)[.063]
BGMM	.369(.201)[.203]	.399(.196)[.196]	.395(.062)[.062]	−.397(.061)[.061]
	gamma			
G2SLS	.485(.219)[.235]	.253(.233)[.275]	.392(.065)[.066]	−.394(.064)[.065]
B2SLS	.377(.318)[.319]	.424(.251)[.252]	.390(.069)[.069]	−.394(.066)[.066]
QML	.428(.195)[.197]	.329(.209)[.221]	.394(.064)[.064]	−.395(.063)[.063]
GMM1	.376(.197)[.198]	.391(.193)[.193]	.392(.064)[.065]	−.394(.063)[.063]
GMM2	.374(.203)[.205]	.392(.198)[.198]	.392(.064)[.065]	−.394(.063)[.064]
BGMM	.365(.223)[.226]	.400(.213)[.213]	.392(.050)[.051]	−.393(.050)[.050]

Mean(SD)[RMSE]

Table 5: Estimation of the SARAR(2,1) model (normal)

	$\lambda_{01} = 0.4$	$\lambda_{02} = 0.2$	$\rho_0 = 0.4$	$\beta_{01} = 1.0$	$\beta_{02} = -1.0$
G2SLS	.419(.090)[.092]	.241(.134)[.140]	.339(.104)[.120]	.995(.059)[.059]	−.996(.055)[.055]
B2SLS	.389(.201)[.201]	.157(.746)[.747]	.339(.104)[.120]	.993(.064)[.065]	−.995(.061)[.061]
GMM1	.393(.092)[.092]	.204(.089)[.089]	.399(.105)[.105]	.994(.059)[.059]	−.995(.055)[.055]
GMM2	.393(.097)[.097]	.202(.088)[.088]	.400(.108)[.108]	.993(.059)[.060]	−.995(.056)[.056]
BGMM	.392(.098)[.098]	.201(.087)[.087]	.403(.111)[.111]	.993(.060)[.060]	−.995(.056)[.057]
	Under the exclusion restriction $\lambda_{02} = 0$				
G2SLS	.428(.090)[.095]	–	.378(.111)[.113]	.999(.059)[.059]	−.999(.055)[.055]
B2SLS	.418(.089)[.091]	–	.378(.111)[.113]	.998(.059)[.059]	−.999(.055)[.055]
QML	.417(.095)[.097]	–	.398(.111)[.111]	.998(.059)[.059]	−.998(.056)[.056]
GMM1	.414(.099)[.100]	–	.404(.112)[.113]	.997(.059)[.059]	−.998(.056)[.056]
GMM2	.414(.098)[.099]	–	.404(.112)[.112]	.997(.059)[.059]	−.998(.056)[.056]
BGMM	.414(.100)[.101]	–	.405(.113)[.113]	.997(.059)[.060]	−.998(.056)[.056]

Mean(SD)[RMSE]

Table 6: Estimation of the SARAR(2,1) model (symmetric mixture normal)

	$\lambda_{01} = 0.4$	$\lambda_{02} = 0.2$	$\rho_0 = 0.4$	$\beta_{01} = 1.0$	$\beta_{02} = -1.0$
G2SLS	.420(.086)[.088]	.245(.131)[.138]	.339(.100)[.117]	.999(.056)[.056]	−.999(.056)[.056]
B2SLS	.395(.104)[.104]	.186(.360)[.361]	.339(.100)[.117]	.998(.057)[.057]	−.998(.057)[.057]
GMM1	.392(.105)[.105]	.206(.104)[.105]	.402(.106)[.106]	.997(.058)[.058]	−.998(.057)[.057]
GMM2	.391(.108)[.108]	.205(.106)[.106]	.403(.105)[.105]	.997(.058)[.058]	−.998(.057)[.057]
BGMM	.392(.098)[.098]	.198(.094)[.094]	.406(.106)[.106]	.997(.055)[.055]	−.998(.055)[.055]
	Under the exclusion restriction $\lambda_{02} = 0$				
G2SLS	.428(.088)[.092]	–	.378(.106)[.108]	1.003(.057)[.057]	−1.003(.056)[.056]
B2SLS	.418(.087)[.088]	–	.378(.106)[.108]	1.002(.057)[.057]	−1.002(.056)[.056]
QML	.417(.092)[.093]	–	.402(.108)[.108]	1.002(.057)[.057]	−1.002(.056)[.056]
GMM1	.414(.094)[.095]	–	.407(.109)[.109]	1.002(.057)[.057]	−1.002(.056)[.056]
GMM2	.415(.093)[.094]	–	.406(.108)[.109]	1.002(.057)[.057]	−1.002(.056)[.056]
BGMM	.415(.092)[.093]	–	.407(.106)[.106]	1.001(.055)[.055]	−1.002(.055)[.055]

Mean(SD)[RMSE]

Table 7: Estimation of the SARAR(2,1) model (gamma)

	$\lambda_{01} = 0.4$	$\lambda_{02} = 0.2$	$\rho_0 = 0.4$	$\beta_{01} = 1.0$	$\beta_{02} = -1.0$
G2SLS	.416(.093)[.094]	.244(.132)[.139]	.341(.105)[.120]	.998(.056)[.056]	−.997(.059)[.059]
B2SLS	.392(.163)[.164]	.185(.800)[.800]	.341(.105)[.120]	.996(.059)[.059]	−.998(.079)[.080]
GMM1	.388(.101)[.102]	.200(.081)[.081]	.405(.107)[.107]	.997(.057)[.057]	−.996(.059)[.060]
GMM2	.388(.105)[.106]	.199(.085)[.085]	.405(.107)[.108]	.997(.058)[.058]	−.996(.060)[.060]
BGMM	.393(.080)[.080]	.197(.081)[.082]	.403(.093)[.093]	.998(.043)[.043]	−1.000(.044)[.044]
	Under the exclusion restriction $\lambda_{02} = 0$				
G2SLS	.423(.094)[.097]	–	.382(.111)[.112]	1.002(.056)[.056]	−1.001(.059)[.059]
B2SLS	.414(.093)[.094]	–	.382(.111)[.112]	1.002(.056)[.056]	−1.000(.059)[.059]
QML	.411(.100)[.101]	–	.403(.113)[.113]	1.001(.057)[.057]	−1.000(.059)[.059]
GMM1	.410(.099)[.100]	–	.409(.111)[.112]	1.001(.056)[.056]	−1.000(.059)[.059]
GMM2	.409(.100)[.100]	–	.409(.112)[.112]	1.001(.056)[.056]	−1.000(.059)[.059]
BGMM	.414(.079)[.080]	–	.408(.094)[.095]	1.001(.042)[.042]	−1.004(.044)[.044]

Mean(SD)[RMSE]

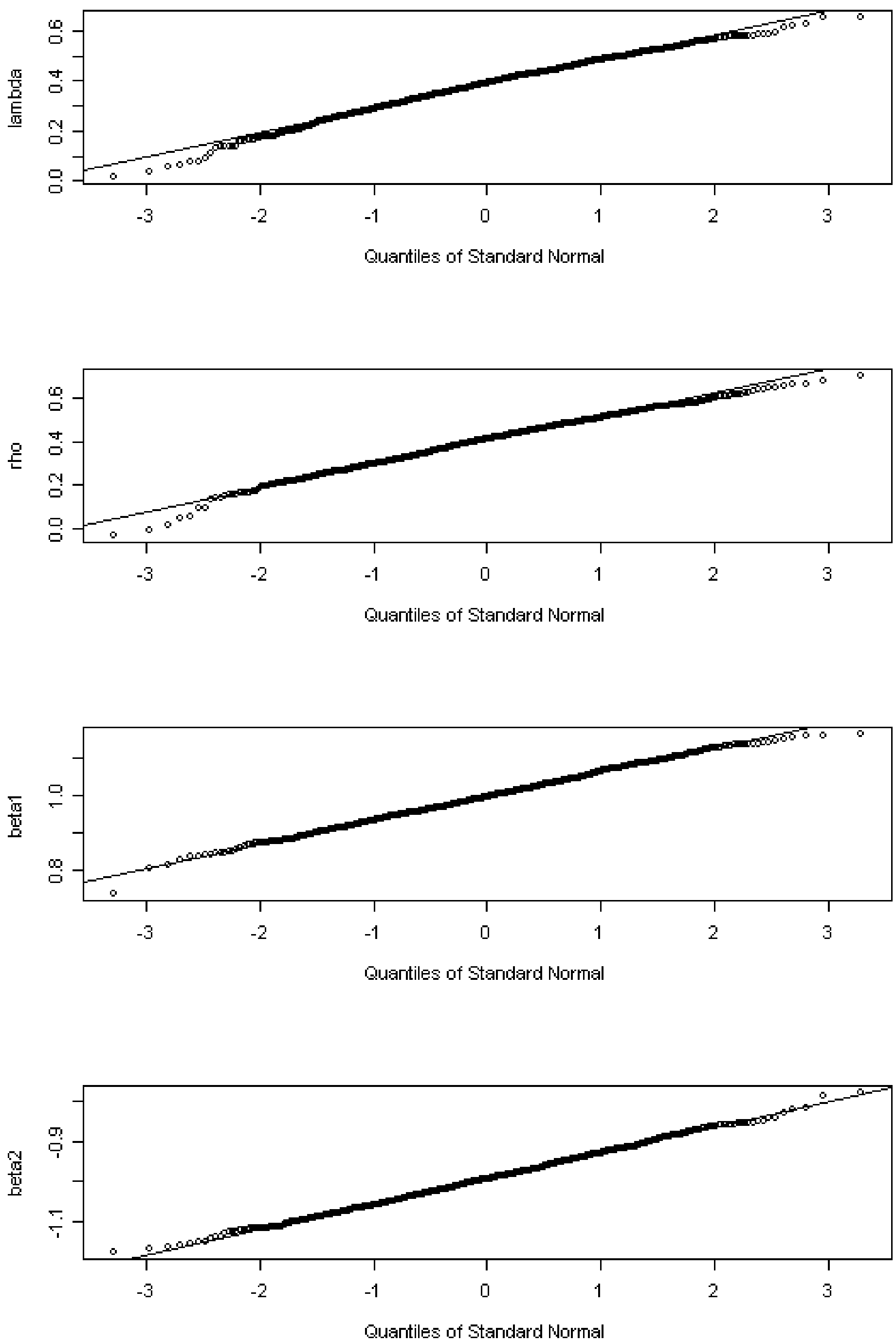

Figure 1: Quantile-quantile Plots for the BGMMEs (normal)

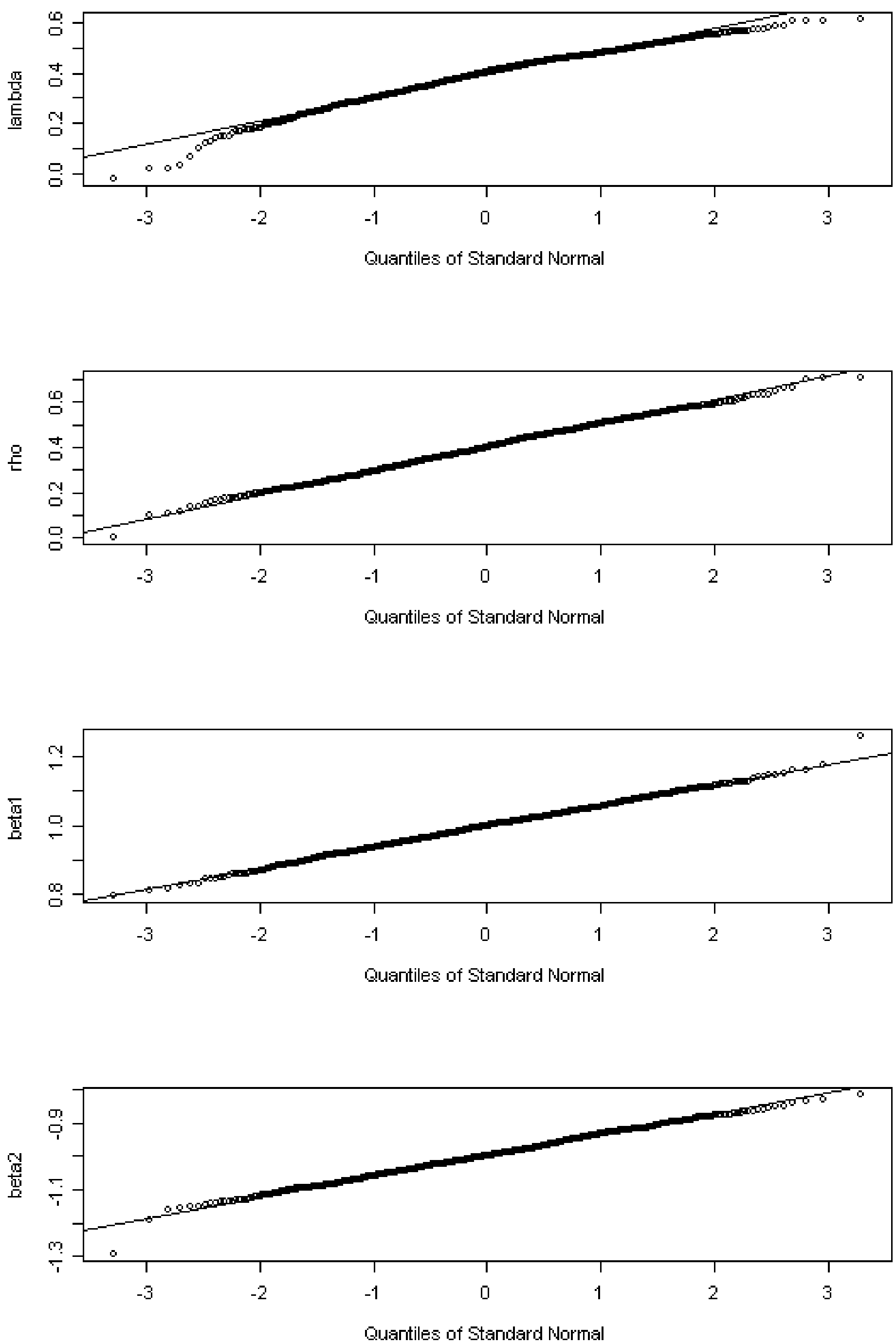

Figure 2: Quantile-quantile Plots for the BGMMEs (symmetric mixture normal)

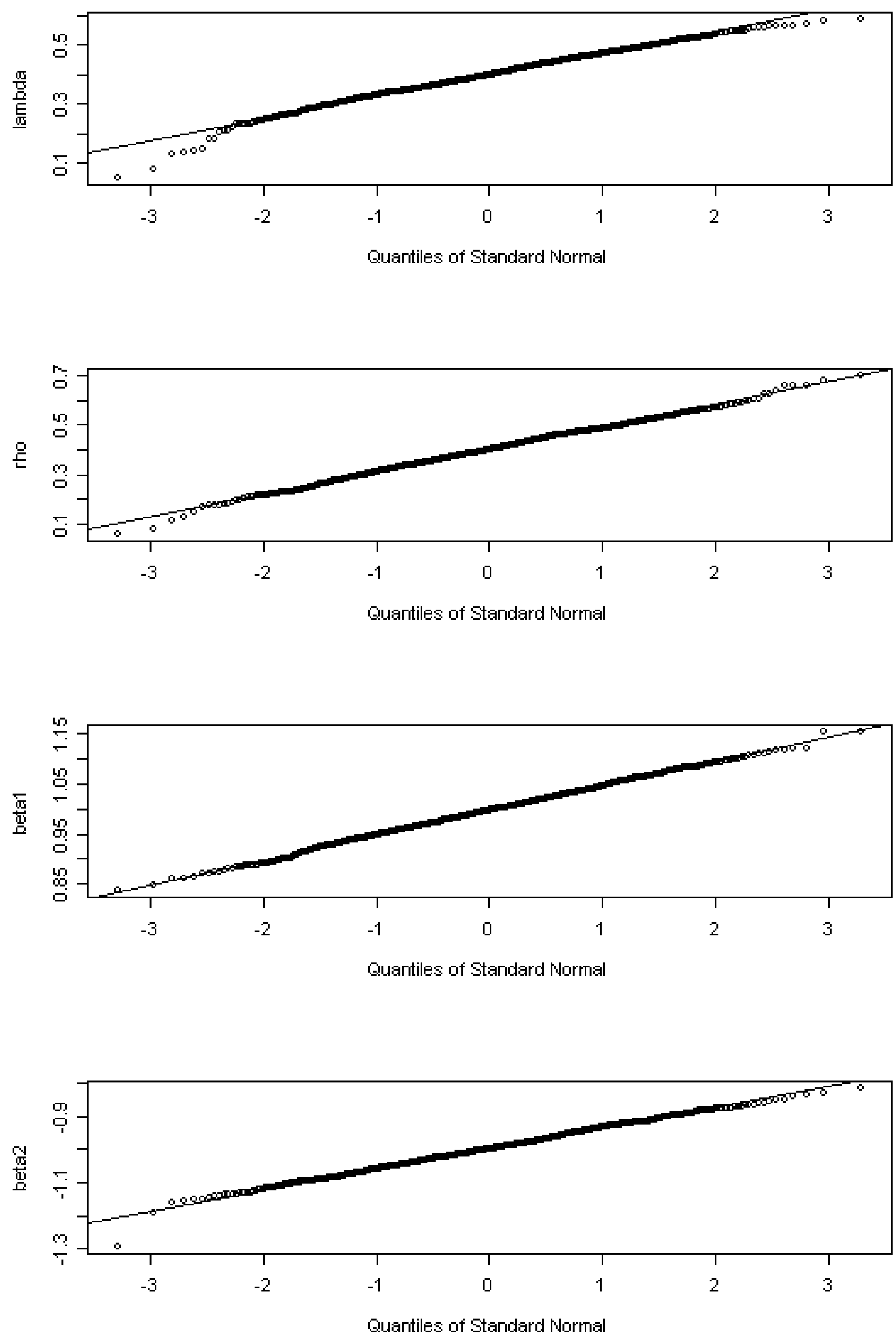

Figure 3: Quantile-quantile Plots for the BGMMEs (gamma)