Journal of Econometrics 159 (2010) 303–319

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

An efficient GMM estimator of spatial autoregressive models☆

Xiaodong Liu[a], Lung-fei Lee[b,∗], Christopher R. Bollinger[c]

[a] *Department of Economics, University of Colorado, Boulder, CO 80309, United States*

[b] *Department of Economics, Ohio State University, Columbus, OH 43210, United States*

[c] *Department of Economics, University of Kentucky, Lexington, KY 40506, United States*

ARTICLE INFO

Article history:
Received 11 October 2006
Received in revised form
28 April 2010
Accepted 19 August 2010
Available online 26 August 2010

JEL classification:
C13
C21
R15

Keywords:
Spatial autoregressive models
Spatial correlated disturbances
GMM
QMLE
Efficiency

ABSTRACT

In this paper, we consider GMM estimation of the regression and MRSAR models with SAR disturbances. We derive the best GMM estimator within the class of GMM estimators based on linear and quadratic moment conditions. The best GMM estimator has the merit of computational simplicity and asymptotic efficiency. It is asymptotically as efficient as the ML estimator under normality and asymptotically more efficient than the Gaussian QML estimator otherwise. Monte Carlo studies show that, with moderate-sized samples, the best GMM estimator has its biggest advantage when the disturbances are asymmetrically distributed. When the diagonal elements of the spatial weights matrix have enough variation, incorporating kurtosis of the disturbances in the moment functions will also be helpful.

1. Introduction

Spatial econometrics models have broad applications in various fields of economics such as regional, urban and public economics. These models address relationships across geographic observations in cross-sectional or panel data. Spatial models have a long history in both statistics and econometrics. Excellent surveys and early developments can be found in Cliff and Ord (1973), Anselin (1988), Cressie (1993) and Anselin and Bera (1998).

Among spatial econometric models, spatial autoregressive (SAR) models by Cliff and Ord (1973) have received the most attention in economics. The first order SAR model can be estimated by the maximum-likelihood (ML) method (see Ord, 1975; Smirnov and Anselin, 2001). Lee (2004) investigates asymptotic properties of the ML estimator (MLE) taking into account various features of the spatial weights matrix. When the sample size is large, the ML method can be computationally demanding for some spatial weights matrices. Alternative estimation methods have subsequently been proposed.

In the presence of exogenous variables in addition to spatial lag variables, the model is known as a mixed regressive, spatial autoregressive model (MRSAR).[1] With the presence of exogenous variables, instrumental variables (IV) can be constructed as functions of the exogenous variables and the spatial weights matrix. The two-stage least squares (2SLS) method has been noted for the estimation of the MRSAR model in Anselin (1988, 1990), Kelejian and Robinson (1993), Kelejian and Prucha (1997, 1998) and Lee (2003), among others. The 2SLS estimator (2SLSE) has been shown to be consistent and asymptotically normal (Kelejian and Prucha, 1998). For the estimation of the linear simultaneous equation model, the 2SLSE is known to be asymptotically as efficient as the limited information MLE (see, e.g., Amemiya, 1985). This is not so for the estimation of the MRSAR model, as it is not a usual linear simultaneous equation model. Lee (2003) discusses the best 2SLSE (B2SLSE) within the class of IV estimators. By comparing the limiting variance matrices, the 2SLSE and B2SLSE are less efficient relative to the MLE when

☆ We thank the editor, the associate editor, and three referees for helpful comments. Lee gratefully acknowledges 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.

∗ Corresponding author. Tel.: +1 614 292 5508; fax: +1 614 292 4192.
E-mail address: lflee@econ.ohio-state.edu (L.-f. Lee).

[1] For simplicity, some authors prefer the terminology, the SAR model, in place of the MRSAR model.

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2010.08.001

the disturbances are normally distributed. For a regression model with SAR disturbances, a method of moments (MOM) approach has been introduced in Kelejian and Prucha (2001). The MOM is computationally simpler than the ML. Their MOM estimator is consistent but can be less efficient relative to the MLE. In order to improve upon the 2SLS, B2SLS and MOM, Lee (2007) has proposed a general GMM estimation framework. For the estimation of the MRSAR model, the proposed GMM method explores both IV (linear) as well as quadratic moment functions. The GMM estimation for those models can be computationally simpler than the MLE. The proposed GMM estimator (GMME) can be asymptotically more efficient than the 2SLSE. With carefully selected linear and quadratic moments, the resulting GMME can be asymptotically as efficient as the MLE when the disturbances are normally distributed. Similarly, for the estimation of a SAR process with normally distributed disturbances, best quadratic moments exist and the resulting GMM estimator can be asymptotically as efficient as the Gaussian MLE.

The best GMM (BGMM) based on the linear and quadratic moments in Lee (2007) assumes that the disturbances of the model are normally distributed. When the disturbances are not normally distributed, such estimators are still consistent and asymptotically normal but may not be efficient. This paper demonstrates that a distribution-free BGMM estimator (BGMME) exists within the class of GMMEs based on the linear and quadratic moments.

Specifically, in this paper, we derive the BGMME for the regression model with SAR disturbances and the MRSAR model with and without SAR disturbances, within the class of GMMEs based on linear and quadratic moment conditions. The BGMME proposed here has the merit of computational simplicity and asymptotic efficiency. It is asymptotically as efficient as the MLE when the disturbances are normally distributed, and asymptotically more efficient than the Gaussian QMLE otherwise.

Recently, Robinson (2010) has proposed an adaptive estimator for the MRSAR models with i.i.d. disturbances ϵ_{ni}'s that follow an unknown distribution. The adaptive estimator is as efficient as ML estimators based on a correct form of distribution. However, in order for the adaptive estimation to be feasible, there are orthogonality conditions which need to be satisfied. In adaptive estimation, one estimates the unknown distribution of the innovations and uses the estimated distribution to construct the score (likelihood) for the estimation of the unknown coefficients of the model. The orthogonality condition requires the estimation error of the distribution to be asymptotically irrelevant for the estimation of the coefficients. For the estimation of the SAR model (even with the normally distributed errors), the ML estimator of the variance of the disturbance is in general asymptotically correlated with that of the spatial lag coefficient. This hints that the adaptive estimation of the model would not be feasible. However, there are special circumstances where the orthogonality condition would hold. One case is the spatial scenario where each spatial unit is influenced by many neighbors whose influences are uniformly small. This case has been studied in Lee (2002) for the OLS approach. In Robinson (2010), he also focuses on such a "many neighbors" case by assuming that the spatial weights matrix W_n has nonnegative elements that are uniformly of order $O(1/h_n)$, where h_n increases with the sample size n such that (1) $h_n/n^{1/2} \to \infty$ as $n \to \infty$, or (2) $h_n \to \infty$ as $n \to \infty$ and either W_n is symmetric or the disturbance ϵ_{ni} is symmetrically distributed. However, the "many neighbors" assumption may not be reasonable in some practical circumstances. The GMM estimation approach proposed in this paper, on the other hand, does not need this assumption. As we have focused on the spatial scenario with a finite number of neighbors, our paper and Robinson (2010) are complementary to each other. Also, the adaptive estimator in Robinson (2010) would not be applicable when all exogenous variables in the model are really irrelevant. The GMM approach in this paper may be used to estimate a pure spatial autoregressive model (without explanatory variables).

This paper is organized as follows. In Section 2, we consider the GMM estimation of the MRSAR model with SAR disturbances. It is interesting and informative to then consider two special cases: the first is estimation of a regression model with SAR disturbances and then an MRSAR model without SAR disturbances. The selection of the best moment functions is discussed and efficiency is considered. All the proofs of the results are collected in the appendices. Section 3 provides some Monte Carlo results for the comparison of finite sample properties of estimators. Section 4 briefly concludes. A list of notations has been collected in Appendix A for convenient reference.

2. GMM estimation and the BGMME

2.1. GMM estimation of the MRSAR model with SAR disturbances

The general MRSAR model with SAR disturbances is given by

$$Y_n = X_n\beta + \lambda W_n Y_n + u_n, \quad u_n = \rho M_n u_n + \epsilon_n, \tag{1}$$

where n is the total number of spatial units, X_n is an $n \times k$-dimensional matrix of nonstochastic exogenous variables, W_n and M_n are zero diagonal spatial weights matrix of known constants that may or may not be equal. The disturbances $\epsilon_{n1}, \ldots, \epsilon_{nn}$ of the n-dimensional vector ϵ_n are i.i.d. $(0, \sigma^2)$. The W_nY_n term is a spacial lag in the dependent variable and its coefficient represents the spatial influence due to neighbors' realized dependent variable. The M_nu_n term is a spacial lag in the disturbances and its coefficient represents the spacial effect of unobservables on neighboring units. In order to distinguish the true parameters from other possible values in the parameter space, we denote β_0, λ_0, ρ_0, and σ_0^2 as the true parameters that generate the observed sample. Let $R_n(\rho) = I_n - \rho M_n$ and $S_n(\lambda) = I_n - \lambda W_n$. At the true parameter values, let $R_n = R_n(\rho_0)$ and $S_n = S_n(\lambda_0)$ for simplicity. The model represents an equilibrium, and so R_n and S_n are assumed to be invertible. The equilibrium vector Y_n is given by $Y_n = S_n^{-1}X_n\beta_0 + S_n^{-1}R_n^{-1}\epsilon_n$. It follows that $W_nY_n = G_nX_n\beta_0 + G_nR_n^{-1}\epsilon_n$, where $G_n = W_nS_n^{-1}$. W_nY_n is correlated with ϵ_n because $E((G_nR_n^{-1}\epsilon_n)'\epsilon_n) = \sigma_0^2 \mathrm{tr}(G_nR_n^{-1}) \neq 0$.

For the estimation of the model (1), we consider the transformed equation, $R_nY_n = R_nX_n\beta_0 + \lambda_0 R_nW_nY_n + \epsilon_n$. Let Q_n be an $n \times q$ matrix of IVs constructed as functions of the regressors and spatial weights matrices. Denote $\epsilon_n(\theta) = R_n(\rho)[S_n(\lambda)Y_n - X_n\beta]$, where $\theta = (\rho, \lambda, \beta')'$. Thus, $\epsilon_n = \epsilon_n(\theta_0)$. The moment functions corresponding to the orthogonality conditions of X_n and ϵ_n are $Q_n'\epsilon_n(\theta)$. In addition to $Q_n'\epsilon_n(\theta)$, Lee (2001b, 2007) suggests the use of the quadratic moment $\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta)$, where P_{jn}'s are $n \times n$ constant matrices such that $\mathrm{tr}(P_{jn}) = 0$ for $j = 1, \ldots, m$. With the selected P_{jn}'s and Q_n, the GMM uses the empirical moments

$$g_n(\theta) = (Q_n, P_{1n}\epsilon_n(\theta), \ldots, P_{mn}\epsilon_n(\theta))'\epsilon_n(\theta). \tag{2}$$

At θ_0, $g_n(\theta_0) = (Q_n, P_{1n}\epsilon_n, \ldots, P_{mn}\epsilon_n)'\epsilon_n$ has a zero mean because $E(Q_n'\epsilon_n) = Q_n'E(\epsilon_n) = 0$ and $E(\epsilon_n'P_{jn}'\epsilon_n) = \sigma_0^2\mathrm{tr}(P_{jn}) = 0$ for $j = 1, \ldots, m$. Lee (2007) has shown the consistency and asymptotic normality of the GMME for the MRSAR model with i.i.d. disturbances. Similar properties for the MRSAR model with SAR disturbances can be found in Lee (2001b). In addition, Lee (2001b) provides identification conditions for (1). In Lee (2001b, 2007), the best moments have been pointed out when ϵ_{ni}'s are normally distributed. In this paper, our interest is on the best selection of P_{jn}'s and Q_n without distributional assumptions on ϵ_n.

We follow the regularity assumptions specified in Lee (2001a, 2007). Henceforth, uniformly bounded in row (column) sums in absolute value of a sequence of square matrices $\{A_n\}$ will be abbreviated as UBR (UBC), and uniformly bounded in both row and column sums in absolute value as UB.[2]

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

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

Assumption 3. The sequences of matrices $\{W_n\}$, $\{M_n\}$, $\{S_n^{-1}\}$ and $\{R_n^{-1}\}$ are UB. $\{S_n^{-1}(\lambda)\}$ and $\{R_n^{-1}(\rho)\}$ are either UBR or UBC, uniformly in λ and ρ in a compact parameter space.

Assumption 4. The sequences of matrices $\{P_{jn}\}$ with $\mathrm{tr}(P_{jn}) = 0$ are UB for $j = 1, \ldots, m$. The elements of Q_n are uniformly bounded.

The assumption that ϵ_{ni} have existing moments higher than the fourth is needed in order to apply a central limit theorem due to Kelejian and Prucha (2001). In general, μ_3 and μ_4 denote, respectively, the third and fourth moments of ϵ_{ni}'s. The uniform boundedness of $\{W_n\}$, $\{M_n\}$, $\{S_n^{-1}\}$ *and* $\{R_n^{-1}\}$ in Assumption 3 limits 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). The additional uniform boundedness of $\{S_n^{-1}(\lambda)\}$ and $\{R_n^{-1}(\rho)\}$ in λ and ρ is required only to justify the QML but not the GMM.[3] Uniform boundedness conditions for X_n, P_{jn}'s and Q_n in Assumptions 2 and 4 are for analytic tractability.

The following assumption summarizes some sufficient identification conditions of θ_0 from the moment equations $E(g_n(\theta_0)) = 0$. Let $H_n = M_n R_n^{-1}$, and $A^{(s)} = A + A'$ for any square matrix A. Let $\alpha_{\rho,j} = \mathrm{tr}(P_{jn}^{(s)}H_n)$, $\alpha_{\lambda,j} = \mathrm{tr}(P_{jn}^{(s)}\bar{G}_n)$, $\alpha_{\rho^2,j} = \mathrm{tr}(H_n'P_{jn}H_n)$, $\alpha_{\lambda^2,j} = \mathrm{tr}(\bar{G}_n'P_{jn}\bar{G}_n)$, $\alpha_{\rho\lambda,j} = \mathrm{tr}(P_{jn}^{(s)}H_n\bar{G}_n + H_n'P_{jn}^{(s)}\bar{G}_n)$, $\alpha_{\rho^2\lambda,j} = \mathrm{tr}(H_n'P_{jn}^{(s)} \times H_n\bar{G}_n)$, $\alpha_{\rho\lambda^2,j} = \mathrm{tr}(\bar{G}_n'P_{jn}^{(s)}H_n\bar{G}_n)$ and $\alpha_{\rho^2\lambda^2,j} = \mathrm{tr}(\bar{G}_n'H_n'P_{jn}H_n\bar{G}_n)$, where $\bar{G}_n = R_nG_nR_n^{-1}$.

Assumption 5. Either (i) $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\rho)(X_n, G_nX_n\beta_0)$ has full rank $(k+1)$ for each possible ρ in its parameter space, and $\lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(P_{jn}H_n) \neq 0$ for some j, $\lim_{n\to\infty}\frac{1}{n}(\mathrm{tr}(P_{1n}^{(s)}H_n), \ldots, \mathrm{tr}(P_{mn}^{(s)}H_n))'$ is linearly independent of $\lim_{n\to\infty}\frac{1}{n}(\mathrm{tr}(H_n'P_{1n}H_n), \ldots, \mathrm{tr}(H_n'P_{mn}H_n))'$; or (ii) $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\rho)X_n$ has full rank k for each possible ρ in its parameter space, $W_n \neq M_n$, and the vectors α's do not have a linear combination with nonlinear coefficients in the form that $\alpha_\rho\delta_1 + \alpha_\lambda\delta_2 + \alpha_{\rho^2}\delta_1^2 + \alpha_{\lambda^2}\delta_2^2 + \alpha_{\rho\lambda}\delta_1\delta_2 + \alpha_{\rho^2\lambda}\delta_1^2\delta_2 + \alpha_{\rho\lambda^2}\delta_1\delta_2^2 + \alpha_{\rho^2\lambda^2}\delta_1^2\delta_2^2 = 0$, for some constants δ_1 and δ_2 with $(\delta_1, \delta_2) \neq 0$.

Assumption 5(i) corresponds to the possible estimation of λ_0 and β_0 by the use of IVs, i.e., linear moments, and ρ_0 from the SAR process of the disturbances. When $G_nX_n\beta_0$ and X_n are linearly dependent, which includes the case that all exogenous variables X_n are irrelevant, (ii) assures the identification of ρ_0 and λ_0 from the quadratic moments as the unique solution of $E[\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta)] = 0$ for $j = 1, \ldots, m$. The identification corresponds to the identification of (ρ_0, λ_0) from the spatial process $v_n = S_n^{-1}R_n^{-1}\epsilon_n$.[4] The details can be found in Lee (2001b).

Assumption 6. Let $\Omega_n = \mathrm{var}(g_n(\theta_0))$. The limit of $\frac{1}{n}\Omega_n$ exists and is a nonsingular matrix.[5]

Assumption 7. The θ_0 is in the interior of the parameter space $\Theta \subset R^{k+2}$.[6]

The GMME $\hat{\theta}_P = \arg\min_{\theta\in\Theta} g_n'(\theta)a_n'a_ng_n(\theta)$ is $\sqrt{n}$-consistent and asymptotically normal. Let $\mathrm{vec}_D(A)$ be the column vector formed by the diagonal elements of a square matrix A. The optimal weighting matrix $a_n'a_n$ is Ω_n^{-1} by the generalized Schwarz inequality, where

$$\Omega_n = \mathrm{var}(g_n(\theta_0)) = \begin{bmatrix} \sigma_0^2 Q_n'Q_n & \mu_3 Q_n'\omega_{mn} \\ \mu_3\omega_{mn}'Q_n & (\mu_4 - 3\sigma_0^4)\omega_{mn}'\omega_{mn} + \sigma_0^4\Delta_m \end{bmatrix},$$

with $\omega_{mn} = [\mathrm{vec}_D(P_{1n}), \ldots, \mathrm{vec}_D(P_{mn})]$ and $\Delta_m = [\mathrm{vec}(P_{1n}^{(s)}), \ldots, \mathrm{vec}(P_{mn}^{(s)})]'[\mathrm{vec}(P_{1n}), \ldots, \mathrm{vec}(P_{mn})]$. Let $\mathcal{M}_n$ be the class of optimal GMMEs derived from $\min_{\theta\in\Theta} g_n'(\theta)\Omega_n^{-1}g_n(\theta)$, where $g_n(\theta)$ is given by (2). To show the existence of the BGMME within $\mathcal{M}_n$, we follow Breusch et al. (1999) in demonstrating that additional moment conditions are redundant to the set of the selected ones.[7] If an intercept appears in $\bar{X}_n = R_nX_n$, 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.[8] Otherwise $\bar{X}_n^* \equiv \bar{X}_n$. Suppose there are k^* 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 $\bar{X}_n^*$. For an $n\times n$ matrix A, let $A^{(t)} = A - \frac{1}{n}\mathrm{tr}(A)I_n$. Let $D(A)$ be a diagonal matrix with diagonal elements being A if A is a vector, or diagonal elements of A if A is a square matrix. Let $\eta_3 = \mu_3/\sigma_0^3$ and $\eta_4 = \mu_4/\sigma_0^4$ be the skewness and kurtosis of the disturbance.

Proposition 1. *Suppose Assumptions 1–7 are satisfied. Let* $P_{1n}^* = \bar{G}_n^{(t)}$, $P_{2n}^* = D(\bar{G}_n^{(t)})$, $P_{3n}^* = D(\bar{G}_n\bar{X}_n\beta_0)^{(t)}$, $P_{4n}^* = H_n^{(t)}$, $P_{5n}^* = D(H_n^{(t)})$ *and* $P_{l+5,n}^* = D(\bar{X}_{nl}^*)^{(t)}$, *for* $l = 1, \ldots, k^*$, *be the weighting matrices of the quadratic moments. Furthermore, let* $Q_{1n}^* = \bar{X}_n^*$, $Q_{2n}^* = \bar{G}_n\bar{X}_n\beta_0$, $Q_{3n}^* = l_n$, $Q_{4n}^* = \mathrm{vec}_D(\bar{G}_n^{(t)})$ *and* $Q_{5n}^* = \mathrm{vec}_D(H_n^{(t)})$ *be the* IV *matrices.*

Denote $g_n^*(\theta) = (Q_n^*, P_{1n}^*\epsilon_n(\theta), \ldots, P_{k^*+5,n}^*\epsilon_n(\theta))'\epsilon_n(\theta)$ *and* $\Omega_n^* = \mathrm{var}(g_n^*(\theta_0))$, *where* $Q_n^* = (Q_{1n}^*, Q_{2n}^*, Q_{3n}^*, Q_{4n}^*, Q_{5n}^*)$. *Then,* $\hat{\theta}_B =$

[2] A sequence of square matrices $\{A_n\}$, where $A_n = [A_{n,ij}]$, is said to be UBR (UBC) if the sequence of row sum matrix norm $\|A_n\|_\infty = \max_{i=1,\ldots,n}\sum_{j=1}^n |A_{n,ij}|$ (column sum matrix norm $\|A_n\|_1 = \max_{j=1,\ldots,n}\sum_{i=1}^n |A_{n,ij}|$) are bounded (Horn and Johnson, 1985).

[3] For the GMM approach, it is sufficient to assume the parameter space to be a bounded set. This is so because the moment functions are linear and quadratic, and they do not involve complicated nonlinearity.

[4] The conditions in (ii) rule out the case $W_n = M_n$. In that case, ρ_0 and λ_0 can be exchanged in the process $v_n = S_n^{-1}R_n^{-1}\epsilon_n$, and they can only be locally identifiable (Anselin, 1988).

[5] Assumptions 5 and 6 exclude the case of large (group) interactions in Lee (2004). These can simplify the presentation of our results. The cases under our assumptions here are relevant to spatial scenario, where interactions are usually among a few neighbors.

[6] In our analysis, the mean value theorem is used occasionally for functions at ρ_0, the interior assumption implicitly implies the existence of a convex neighborhood around ρ_0 for the validity of the mean value theorem.

[7] In Appendix B, we investigate the efficient MOM estimation of a simple SAR process. Due to the simple structure of that model, we have a constructive approach based on the Schwartz inequality to derive the best moments. The feature of the best moment conditions for the simple SAR process sheds light on the search for the best moment conditions for the more general MRSAR model. From the simple model, we realize that some diagonal matrices, with the diagonal elements being (1) the diagonal elements of the best quadratic moment matrices P_n's under normality and (2) the best instruments under normality, can be used to construct additional quadratic moment conditions to improve efficiency when errors follow a non-normal distribution. Also, some vectors with elements being the diagonal elements of the best P_n's under normality can be used as additional instruments to improve efficiency. We thus find candidate moment conditions of these forms for the general model and use the results in Breusch et al. (1999) to verify the best ones and show that any additional linear and quadratic moment conditions are redundant.

[8] When M_n is row-normalized, $M_nl_n = l_n$ and $(I_n - \rho_0M_n)^{-1}l_n = (1-\rho_0)^{-1}l_n$. Hence, $R_nl_n = M_n(I_n - \rho_0M_n)^{-1}l_n = (I_n - \rho_0M_n)^{-1}M_nl_n = (1-\rho_0)^{-1}l_n$. In this case, $c_n(\rho_0) = (1-\rho_0)^{-1}$. If M_n is not row-normalized, $\bar{X}_n$ will, in general, not have a column proportional to l_n.

$$\Sigma_B = \lim_{n\to\infty} \frac{1}{n} \begin{bmatrix} \mathrm{tr}(P^{*(s)}_{\rho n} H_n) & \mathrm{tr}(P^{*(s)}_{\lambda n} H_n) & -\frac{2\eta_3}{\sigma_0(\eta_4 - 1 - \eta_3^2)} \mathrm{vec}'_D(H_n^{(t)})\bar{X}_n \\ * & \sigma_0^{-2}(\bar{G}_n \bar{X}_n \beta_0)' Q^*_{\lambda n} + \mathrm{tr}(P^{*(s)}_{\lambda n} \bar{G}_n) & \sigma_0^{-2} Q^{*\prime}_{\lambda n} \bar{X}_n \\ * & * & \sigma_0^{-2} \bar{X}'_n Q^*_{\beta n} \end{bmatrix}, \quad (3)$$

Box I.

arg $\min_{\theta\in\Theta} g_n^{*\prime}(\theta)\Omega_n^{*-1}g_n^*(\theta)$ *is the BGMME within* $\mathcal{M}_n$*, and it has the asymptotic distribution that* $\sqrt{n}(\hat{\theta}_B - \theta_0) \xrightarrow{D} N(0, \Sigma_B^{-1})$*, where* $\Sigma_B = \lim_{n\to\infty} \frac{1}{n} D_n^{*\prime}\Omega_n^{*-1}D_n^*$ *and*

$$D_n^* = E\left(\frac{\partial}{\partial\theta'} g_n^*(\theta_0)\right) = -\begin{bmatrix} 0 & Q_n^{*\prime}\bar{G}_n\bar{X}_n\beta_0 & Q_n^{*\prime}\bar{X}_n \\ \sigma_0^2 \mathrm{tr}(P^{*(s)}_{1n} H_n) & \sigma_0^2 \mathrm{tr}(P^{*(s)}_{1n}\bar{G}_n) & 0 \\ \vdots & \vdots & \vdots \\ \sigma_0^2 \mathrm{tr}(P^{*(s)}_{k^*+5,n} H_n) & \sigma_0^2 \mathrm{tr}(P^{*(s)}_{k^*+5,n}\bar{G}_n) & 0 \end{bmatrix}.$$

As shown in the proof of Proposition 1, Σ_B has an explicit form as in Eq. (3) given in Box I, where $P^*_{\lambda n} = P^*_{1n} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2}P^*_{2n} - \frac{\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2}P^*_{3n}$, $P^*_{\rho n} = P^*_{4n} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2}P^*_{5n}$, $Q^*_{\beta n} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}\bar{X}_n - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}Q^*_{3n}\left(\frac{1}{n}l'_n\bar{X}_n\right)$, and $Q^*_{\lambda n} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}Q^*_{2n} - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}Q^*_{3n}\left(\frac{1}{n}l'_n\bar{G}_n\bar{X}_n\beta_0\right) - \frac{2\sigma_0\eta_3}{(\eta_4-1)-\eta_3^2}Q^*_{4n}$. From our proof, the best moments in Proposition 1 is equivalent to their linear combinations given by

$$g_n^{\#}(\theta) = (Q_n^{\#}, P^*_{\lambda n}\epsilon_n(\theta), P^*_{\rho n}\epsilon_n(\theta), P^*_{6n}\epsilon_n(\theta), \ldots, P^*_{k^*+5,n}\epsilon_n(\theta))'\epsilon_n(\theta) \quad (4)$$

with $Q_n^{\#} = (Q^*_{\beta n}, Q^*_{\lambda n}, Q^*_{5n})$.[9] When ϵ_n is normally distributed so that $\eta_3 = 0$ and $\eta_4 = 3$, we have $P^*_{\lambda n} = \bar{G}_n^{(t)}$, $P^*_{\rho n} = H_n^{(t)}$, $Q^*_{\beta n} = \bar{X}_n$ and $Q^*_{\lambda n} = \bar{G}_n\bar{X}_n\beta_0$. Following Breusch et al. (1999), $Q^{*\prime}_{5n}\epsilon_n(\theta)$ and $(P^*_{6n}\epsilon_n(\theta), \ldots, P^*_{k^*+5,n}\epsilon_n(\theta))'\epsilon_n(\theta)$ can be shown redundant given the best moment functions $[\bar{X}_n, \bar{G}_n\bar{X}_n\beta_0, \bar{G}_n^{(t)}\epsilon_n(\theta), H_n^{(t)}\epsilon_n(\theta)]'\epsilon_n(\theta)$ under normality in Lee (2001b).[10] When ϵ_n is not normally distributed, the additional moments in Proposition 1 improve efficiency as they capture the skewness and kurtosis of the error distribution.

The asymptotic efficiency of the MLE depends on the distribution of the disturbances being correctly specified. The likelihood function based on the normal specification is a quasi-likelihood when the disturbances are not truly normal. The resulted estimator is a QMLE. We claim that the BGMME in Proposition 1 is asymptotically more efficient relative to this QMLE. This can be seen as follows. The log-likelihood function for MRSAR model with SAR disturbances is

$$\ln L_n = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln|S_n(\lambda)| + \ln|R_n(\rho)| - \frac{1}{2\sigma^2}[S_n(\lambda)Y_n - X_n\beta]'R'_n(\rho)R_n(\rho)[S_n(\lambda)Y_n - X_n\beta],$$

and the derivatives are $\frac{\partial}{\partial\rho}\ln L_n = -\mathrm{tr}(H_n(\rho)) + \frac{1}{\sigma^2}\epsilon'_n(\theta)H_n(\rho)\epsilon_n(\theta)$,

$$\frac{\partial}{\partial\lambda}\ln L_n = -\mathrm{tr}(\bar{G}_n(\rho,\lambda)) + \frac{1}{\sigma^2}[\bar{G}_n(\rho,\lambda)\bar{X}_n(\rho)\beta]'\epsilon_n(\theta) + \frac{1}{\sigma^2}\epsilon'_n(\theta)\bar{G}_n(\rho,\lambda)\epsilon'_n(\theta),$$

$\frac{\partial}{\partial\beta}\ln L_n = \frac{1}{\sigma^2}\bar{X}'_n(\rho)\epsilon_n(\theta)$, and $\frac{\partial}{\partial\sigma^2}\ln L_n = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\epsilon'_n(\theta)\epsilon_n(\theta)$, where $\bar{X}_n(\rho) = R_n(\rho)X_n$ and $\bar{G}_n(\rho,\lambda) = R_n(\rho)G_n(\lambda)R_n^{-1}(\rho)$. The QMLE of σ_0^2 is given by $\hat{\sigma}^2_{ml}(\theta) = \frac{1}{n}\epsilon'_n(\theta)\epsilon_n(\theta)$ for a given value θ. Substituting $\hat{\sigma}^2_{ml}(\theta)$ into the remaining first order conditions shows that the QMLE is characterized by the moment equations $\epsilon'_n(\theta)H_n^{(t)}(\rho)\epsilon_n(\theta) = 0$, $[\bar{G}_n(\rho,\lambda)\bar{X}_n(\rho)\beta]'\epsilon_n(\theta) + \epsilon'_n(\theta)\bar{G}_n^{(t)}(\rho,\lambda) \times \epsilon_n(\theta) = 0$, and $\bar{X}'_n(\rho)\epsilon_n(\theta) = 0$. Denote the QMLE of θ by $\hat{\theta}_{ml}$. Obviously $\hat{\theta}_{ml}$ is the solution of $a_n\hat{g}_{ml,n}(\theta)$, where $a_n = \begin{bmatrix} I_k & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and

$$\hat{g}_{ml,n}(\theta) = [\bar{X}_n(\hat{\rho}_{ml}), \bar{G}_n(\hat{\rho}_{ml}, \hat{\lambda}_{ml})\bar{X}_n(\hat{\rho}_{ml})\hat{\beta}_{ml}, \bar{G}_n^{(t)}(\hat{\rho}_{ml}, \hat{\lambda}_{ml})\epsilon_n(\theta), H_n^{(t)}(\hat{\rho}_{ml})\epsilon_n(\theta)]'\epsilon_n(\theta).$$

It follows from analogous arguments in the proof of Proposition 3 in Lee (2007) that $a_n\hat{g}_{ml,n}(\theta) = 0$ is asymptotically equivalent to the moment equations $a_n g_{ml,n}(\theta) = 0$, where

$$g_{ml,n}(\theta) = [\bar{X}_n, \bar{G}_n\bar{X}_n\beta_0, \bar{G}_n^{(t)}\epsilon_n(\theta), H_n^{(t)}\epsilon_n(\theta)]'\epsilon_n(\theta),$$

in the sense that their consistent roots have the same limiting distribution. As $g_{ml,n}(\theta)$ consists of linear and quadratic functions of $\epsilon_n(\theta)$, the corresponding optimal GMME derived from $\min g'_{ml,n}(\theta)\Omega_n^{-1}g_{ml,n}(\theta)$ is in $\mathcal{M}_n$. As the BGMME is the most efficient estimator in $\mathcal{M}_n$, hence, the BGMME is efficient relative to the QMLE.

In practice, with initial consistent estimates $\hat{\theta}_n, \hat{\sigma}_n^2, \hat{\mu}_3$ and $\hat{\mu}_4$, P^*_{jn} and Q^*_n can be estimated as $\hat{P}^*_{jn} = P^*_{jn}(\hat{\theta}_n)$ and $\hat{Q}^*_n = Q^*_n(\hat{\theta}_n)$ for $j = 1, \ldots, k^* + 5$. The variance matrix Ω_n^* of the best moment functions can be estimated as $\hat{\Omega}_n^* = \Omega_n^*(\hat{\theta}_n, \hat{\sigma}_n^2, \hat{\mu}_3, \hat{\mu}_4)$. The following proposition shows that the feasible BGMME has the same limiting distribution as the BGMME in Proposition 1.

Proposition 2. *Under Assumptions 1–7, suppose* $\hat{\theta}_n, \hat{\sigma}_n^2, \hat{\mu}_3$ *and* $\hat{\mu}_4$ *are, respectively,* $\sqrt{n}$*-consistent estimates of* $\theta_0, \sigma_0^2, \mu_3$ *and* μ_4*. Then,* $\hat{\theta}_{FB}$ = arg $\min_{\theta\in\Theta} \hat{g}_n^{*\prime}(\theta)\hat{\Omega}_n^{*-1}\hat{g}_n^*(\theta)$*, with* $\hat{\Omega}_n^* = \Omega_n^*(\hat{\theta}_n, \hat{\sigma}_n^2, \hat{\mu}_3, \hat{\mu}_4)$ *and* $\hat{g}_n^*(\theta) = (\hat{Q}_n^*, \hat{P}^*_{1n}\epsilon_n(\theta), \ldots, P^*_{k^*+5,n}\epsilon_n(\theta))'\epsilon_n(\theta)$*, has the same limiting distribution as* $\hat{\theta}_B$ = arg $\min_{\theta\in\Theta} g_n^{*\prime}(\theta)\Omega_n^{*-1}g_n^*(\theta)$.

2.2. GMM estimation of the regression model with SAR disturbances

An important special case of the general MRSAR-SAR model is the regression model with SAR disturbances, that is the case where $\lambda_0 = 0$. Two approaches are interesting to contrast. The first approach estimates ρ and then estimates β using the feasible generalized least squares (FGLS). The second approach uses the full model GMM estimation above to estimate the parameters

[9] We realize that these are not the unique linear combinations. They help to realize how the best moments of the normal distributed case shall be modified to accommodate the non-normal case. They are also helpful for the implementation of the estimation procedure in practice (see the Monte Carlo section).

[10] In the simulation studies, we compare the finite sample performance of the BGMME based on the enlarged set of moment conditions with the Gaussian MLE when ϵ_{ni}'s are normally distributed. For a moderate-sized sample, the performance of the BGMME is as good as that of the MLE.

simultaneously. In this section, we focus on the second approach, and the FGLS approach is discussed in Appendix B.[11]

Let $\mathcal{M}_{\rho n}$ be the class of optimal GMMEs of (ρ_0, β_0') derived from $\min_{\rho,\beta} g_{\rho n}'(\rho,\beta)\Omega_{\rho n}^{-1}g_{\rho n}(\rho,\beta)$, where $\Omega_{\rho n} = \mathrm{var}(g_{\rho n}(\rho_0,\beta_0))$ and $g_{\rho n}(\rho,\beta) = (Q_n, P_{1n}\epsilon_{\rho n}(\rho,\beta),\ldots,P_{mn}\epsilon_{\rho n}(\rho,\beta))'\epsilon_{\rho n}(\rho,\beta)$ with $\epsilon_{\rho n}(\rho,\beta) = R_n(\rho)(Y_n - X_n\beta)$. As a special case of the GMM estimation in Section 2.1 by imposing the restriction that $\lambda_0 = 0$, we find the following result.

Corollary 3 (*To Proposition 1*). *Consider the GMM estimation of the restricted model (1) with $\lambda_0 = 0$ under Assumptions 1–7. Let $P_{1n}^{\dagger} = H_n^{(t)}$, $P_{2n}^{\dagger} = D(H_n^{(t)})$ and $P_{j+2,n}^{\dagger} = D(\bar{X}_{nj}^{*})^{(t)}$ (for $j = 1,\ldots,k^*$) be the weighting matrices of the quadratic moments, and $Q_{1n}^{\dagger} = \bar{X}_n^{*}$, $Q_{2n}^{\dagger} = l_n$ and $Q_{3n}^{\dagger} = \mathrm{vec}_D(H_n^{(t)})$ be the IV matrices.*

Let

$$g_{\rho n}^{\dagger}(\rho,\beta) = (Q_n^{\dagger}, P_{1n}^{\dagger}\epsilon_{\rho n}(\rho,\beta),\ldots,P_{k^*+2,n}^{\dagger}\epsilon_{\rho n}(\rho,\beta))'\epsilon_{\rho n}(\rho,\beta)$$

and $\Omega_{\rho n}^{\dagger} = \mathrm{var}(g_{\rho n}^{\dagger}(\rho_0,\beta_0))$, where $Q_n^{\dagger} = (Q_{1n}^{\dagger}, Q_{2n}^{\dagger}, Q_{3n}^{\dagger})$. Then, $\hat{\rho}_{B\rho}$ and $\hat{\beta}_{B\rho}$ derived from $\min_{\rho,\beta} g_{\rho n}^{\dagger}(\rho,\beta)'(\Omega_{\rho n}^{\dagger})^{-1}g_{\rho n}^{\dagger}(\rho,\beta)$ is the BGMME within $\mathcal{M}_{\rho n}$ with the asymptotic variance matrix $\frac{1}{n}\Sigma_{B\rho}^{-1}$, where

$$\Sigma_{B\rho} = \lim_{n\to\infty}\frac{1}{n} \times \begin{bmatrix} \mathrm{tr}[(P_{\rho n}^{\dagger})^{(s)}H_n] & -\dfrac{2\eta_3}{\sigma_0(\eta_4-1-\eta_3^2)}\mathrm{vec}_D'(H_n^{(t)})\bar{X}_n \\ * & \sigma_0^{-2}\bar{X}_n'Q_{\beta n}^{\dagger} \end{bmatrix}, \quad (5)$$

with $P_{\rho n}^{\dagger} = P_{1n}^{\dagger} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2}P_{2n}^{\dagger}$ and $Q_{\beta n}^{\dagger} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}\bar{X}_n - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}Q_{2n}^{\dagger}\left(\frac{1}{n}l_n'\bar{X}_n\right)$.

By comparing the result in Corollary 3 with the FGLS in Appendix B, we see that when $\eta_3 = 0$, which implies that the linear and quadratic moments are uncorrelated, the best MOM (BMOM) estimator of ρ_0 and the FGLS estimator of β_0 in Appendix B have the same limiting distribution as the BGMME given in Corollary 3. Indeed, when $\eta_3 = 0$, the best P_n^* of the MOM approach given in Proposition 5 is the same as $P_{\rho n}^{\dagger}$ in Corollary 3, and the best linear moment $Q_{\beta n}^{\dagger} = \bar{X}_n$ corresponds to the GLS type moment for the estimation of β_0. However, when $\eta_3 \neq 0$, the BGMME in Corollary 3 can be efficient relative to the FGLS estimator of β_0 as well as the proposed BMOM estimator of ρ_0 in Appendix B. The GMME of β_0 is no longer a linear function of Y_n when $\eta_3 \neq 0$, but the FGLS estimator is. While the FGLS estimator of β_0 remains the best linear unbiased one, it can be inefficient relative to some nonlinear estimators like the one given in Corollary 3. The MLE estimator, under departures from normality, will not always fall in the class of linear unbiased estimators. Hence it is not surprising that improvements on the linear unbiased estimator can be found in general.

2.3. GMM estimation of the MRSAR model with IID disturbances

Another special case of the model (1) is the MRSAR model with i.i.d. disturbances, i.e., $\rho_0 = 0$. The following corollary gives the BGMME of the MRSAR model with i.i.d. disturbances. Let $\mathcal{M}_{\lambda n}$ be the class of optimal GMMEs of (λ_0, β_0') derived from $\min_{\lambda,\beta} g_{\lambda n}'(\lambda,\beta)\Omega_{\lambda n}^{-1}g_{\lambda n}(\lambda,\beta)$, where $\Omega_{\lambda n} = \mathrm{var}(g_{\lambda n}(\lambda_0,\beta_0))$ and $g_{\lambda n}(\lambda,\beta) = (Q_n, P_{1n}\epsilon_{\lambda n}(\lambda,\beta),\ldots,P_{mn}\epsilon_{\lambda n}(\lambda,\beta))'\epsilon_{\lambda n}(\lambda,\beta)$ with $\epsilon_{\lambda n}(\lambda,\beta) = S_n(\lambda)Y_n - X_n\beta$.

[11] Throughout the paper we maintain Assumptions 1–7 (suitably modified for different models).

Corollary 4 (*To Proposition 1*). *Consider the GMM estimation of the restricted model (1) with $\rho_0 = 0$ under Assumptions 1–7. Let $P_{1n}^{\star} = G_n^{(t)}$, $P_{2n}^{\star} = D(G_n^{(t)})$, $P_{3n}^{\star} = D(G_nX_n\beta_0)^{(t)}$, and $P_{j+3,n}^{\star} = D(X_{nj}^{*})^{(t)}$ (for $j = 1,\ldots,k^*$) be the weighting matrices of the quadratic moments, and $Q_{1n}^{\star} = X_n^{*}$, $Q_{2n}^{\star} = G_nX_n\beta_0$, $Q_{3n}^{\star} = l_n$ and $Q_{4n}^{\star} = \mathrm{vec}_D(G_n^{(t)})$ be the IV matrices.*

Let $g_{\lambda n}^{\star}(\lambda,\beta) = (Q_n^{\star}, P_{1n}^{\star}\epsilon_{\lambda n}(\lambda,\beta),\ldots,P_{k^*+3,n}^{\star}\epsilon_{\lambda n}(\lambda,\beta))'\epsilon_{\lambda n}(\lambda,\beta)$ and $\Omega_{\lambda n}^{\star} = \mathrm{var}(g_{\lambda n}^{\star}(\lambda_0,\beta_0))$, where $Q_n^{\star} = (Q_{1n}^{\star}, Q_{2n}^{\star}, Q_{3n}^{\star}, Q_{4n}^{\star})$. Then, $\hat{\lambda}_{B\lambda}$ and $\hat{\beta}_{B\lambda}$ derived from $\min_{\lambda,\beta} g_{\lambda n}^{\star}(\lambda,\beta)'(\Omega_{\lambda n}^{\star})^{-1}g_{\lambda n}^{\star}(\lambda,\beta)$ is the BGMME within $\mathcal{M}_{\lambda n}$ with the asymptotic variance matrix $\frac{1}{n}\Sigma_{B\lambda}^{-1}$, where

$$\Sigma_{B\lambda} = \lim_{n\to\infty}\frac{1}{n} \times \begin{bmatrix} \sigma_0^{-2}(G_nX_n\beta_0)'Q_{\lambda n}^{\star} + \mathrm{tr}[(P_{\lambda n}^{\star})^{(s)}G_n] & \sigma_0^{-2}(Q_{\lambda n}^{\star})'X_n \\ * & \sigma_0^{-2}X_n'Q_{\beta n}^{\star} \end{bmatrix}, \quad (6)$$

with $P_{\lambda n}^{\star} = P_{1n}^{\star} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2}P_{2n}^{\star} - \frac{\eta_3}{\sigma_0((\eta_4-1)-\eta_3^2)}P_{3n}^{\star}$, $Q_{\beta n}^{\star} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}X_n - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}Q_{3n}^{\star}\left(\frac{1}{n}l_n'X_n\right)$ and $Q_{\lambda n}^{\star} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}Q_{2n}^{\star} - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}Q_{3n}^{\star}\left(\frac{1}{n}l_n'G_nX_n\beta_0\right) - \frac{2\sigma_0\eta_3}{(\eta_4-1)-\eta_3^2}Q_{4n}^{\star}$.

When ϵ_n is normally distributed, $\eta_3 = 0$ and $\eta_4 = 3$, and hence, $Q_{\beta n}^{\star} = X_n$, $Q_{\lambda n}^{\star} = G_nX_n\beta_0$ and $P_{\lambda n}^{\star} = G_n^{(t)}$. Based on the characterization of best moments in Breusch et al. (1999), it can be shown that any moment function in the form of (2) is redundant given $(X_n, G_nX_n\beta_0, G_n^{(t)}\epsilon_{\lambda n}(\lambda,\beta))'\epsilon_{\lambda n}(\lambda,\beta)$ under normality, with similar arguments used in the proof of Proposition 1.

On the other hand, the likelihood function of the MRSAR model with i.i.d. disturbances is

$$\ln L_n = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}[S_n(\lambda)Y_n - X_n\beta]'[S_n(\lambda)Y_n - X_n\beta]$$

with its derivatives being $\frac{\partial \ln L_n}{\partial\beta} = \frac{1}{\sigma^2}X_n'\epsilon_{\lambda n}(\lambda,\beta)$, $\frac{\partial \ln L_n}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\epsilon_{\lambda n}'(\lambda,\beta)\epsilon_{\lambda n}(\lambda,\beta)$, and

$$\frac{\partial \ln L_n}{\partial\lambda} = -\mathrm{tr}(G_n(\lambda)) + \frac{1}{\sigma^2}(G_n(\lambda)X_n\beta)'\epsilon_{\lambda n}(\lambda,\beta) + \frac{1}{\sigma^2}\epsilon_{\lambda n}'(\lambda,\beta)G_n(\lambda)\epsilon_{\lambda n}(\lambda,\beta).$$

The score vector of QMLE consists of linear and quadratic moments of $\epsilon_{\lambda n}(\lambda,\beta)$. Hence, the optimal GMME based on that score vector is in $\mathcal{M}_{\lambda n}$, and hence is less efficient relative to the BGMME in Corollary 4.

3. Monte Carlo study

In the Monte Carlo experiments, the model is specified as

$$Y_n = X_{n1}\beta_{10} + X_{n2}\beta_{20} + \lambda_0W_nY_n + u_n, \quad u_n = \rho_0W_nu_n + \epsilon_n. \quad (7)$$

The regressors X_{n1} and X_{n2} are mutually independent vectors of independent standard normal random variables. The error terms, ϵ_{ni}'s, are independently generated from the following two distributions: (a) normal, $\epsilon_{ni} \sim N(0,2)$ and (b) gamma, $\epsilon_{ni} = \gamma_i - 2$, where $\gamma_i \sim \mathrm{gamma}(2,1)$. The ϵ_{ni}'s have mean zero and variance 2. The skewness (η_3) and kurtosis (η_4) of these distributions are correspondingly: (a) $\eta_3 = 0$, $\eta_4 = 3$ and (b) $\eta_3 = \sqrt{2}$, $\eta_4 = 6$. When the disturbances are normally distributed, both the MLE and the BGMME are asymptotically efficient. The gamma distribution is introduced to study the effects of skewness and excess kurtosis on the small sample performance of various estimators. The BGMME is asymptotically more efficient than the

Table 1
The regression model with SAR disturbances ($\lambda_0 = 0$).

	$\rho_0 = 0.3$	$\beta_{10} = 1.0$	$\beta_{20} = -1.0$	Time (s)
$n = 98$	Normal			
GLS1	0.279(0.134)[0.136]	0.999(0.144)[0.144]	−0.998(0.146)[0.146]	0.0071
GLS2	0.278(0.131)[0.132]	0.999(0.144)[0.144]	−0.998(0.146)[0.146]	0.0042
BGMM	0.329(0.143)[0.146]	0.997(0.151)[0.151]	−0.999(0.153)[0.153]	0.0188
Gaussian ML	0.287(0.134)[0.135]	0.999(0.144)[0.144]	−0.998(0.146)[0.146]	0.2726
$n = 490$				
GLS1	0.294(0.055)[0.056]	1.000(0.062)[0.062]	−0.998(0.063)[0.063]	0.0075
GLS2	0.294(0.055)[0.056]	1.000(0.062)[0.062]	−0.998(0.063)[0.063]	0.0418
BGMM	0.305(0.056)[0.056]	1.000(0.064)[0.064]	−0.997(0.064)[0.064]	0.1613
Gaussian ML	0.294(0.055)[0.055]	1.000(0.062)[0.062]	−0.998(0.063)[0.063]	0.4870
$n = 98$	Gamma			
GLS1	0.281(0.130)[0.131]	1.004(0.143)[0.143]	−1.009(0.144)[0.144]	0.0069
GLS2	0.282(0.125)[0.127]	1.004(0.143)[0.143]	−1.009(0.144)[0.144]	0.0042
BGMM	0.331(0.138)[0.141]	1.003(0.113)[0.113]	−1.005(0.115)[0.115]	0.0195
Gaussian QML	0.290(0.129)[0.129]	1.004(0.143)[0.143]	−1.009(0.144)[0.144]	0.2632
Gamma ML	0.299(0.101)[0.101]	1.005(0.093)[0.093]	−1.004(0.093)[0.093]	0.0324
$n = 490$				
GLS1	0.297(0.056)[0.056]	0.996(0.063)[0.063]	−1.003(0.061)[0.061]	0.0075
GLS2	0.297(0.055)[0.055]	0.996(0.063)[0.063]	−1.003(0.061)[0.061]	0.0419
BGMM	0.307(0.055)[0.056]	0.998(0.049)[0.049]	−1.001(0.049)[0.049]	0.1630
Gaussian QML	0.297(0.055)[0.055]	0.996(0.063)[0.063]	−1.003(0.061)[0.061]	0.4939
Gamma ML	0.300(0.034)[0.034]	1.000(0.029)[0.029]	−0.999(0.030)[0.030]	0.0895

Mean (SD) [RMSE].

Gaussian QMLE when ϵ_{ni}'s follow the gamma distribution, as its moment functions incorporate skewness and excess kurtosis of the error distribution.

The number of repetitions is 1000 for each case in the Monte Carlo experiments. The regressors are randomly redrawn for each repetition.[12] In each case, we report the 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. We set $\beta_{10} = 1.0$ and $\beta_{20} = -1.0$ in the data generating process. The variance ratio of $X_{n1}\beta_{10} + X_{n2}\beta_{20}$ with the sum of variances of $X_{n1}\beta_{10} + X_{n2}\beta_{20}$ and ϵ_n is 0.5. If one ignores the interaction term, this ratio would represent $R^2 = 0.5$ in a regression equation. λ_0 and ρ_0 are varied in the experiments. The sample sizes considered are $n = 98$ and $n = 490$.

We take the weights matrix W_A from the study of crimes across 49 districts in Columbus, Ohio in Anselin (1988). For $n = 98$ and 490, the corresponding spatial weights matrices in the Monte Carlo study are given by $I_2 \otimes W_A$ and $I_{10} \otimes W_A$ respectively, where $\otimes$ denotes the Kronecker product operator.

The first case we consider is $\lambda_0 = 0$ and $\rho_0 = 0.3$, so that (7) reduces to the regression model with SAR disturbances. The estimators considered are (i) the GLS1 estimator where ρ_0 is estimated by the MOM in Kelejian and Prucha (1998) and the GLS2 estimator where ρ_0 is estimated by the BMOM in Proposition 5, (ii) the BGMME in Corollary 3, (iii) the Gaussian QMLE,[13] and (iv) the gamma MLE when the innovations follow the gamma distribution.[14] We use preliminary estimates from the GLS1 to implement the GLS2 and the feasible BGMM.

The estimation results for the regression model with SAR errors are reported in Table 1. The GLS estimators and the Gaussian QMLE of ρ_0 are biased downward and the BGMME of ρ_0 is biased upward for a small sample size $n = 98$. The bias reduces as the sample size increases. When the disturbances are normally distributed, the Gaussian MLE is efficient. When $n = 98$, the BGMME of ρ_0 has a slightly bigger SD than the MLE. For a moderate sample size $n = 490$, the performance of the BGMME is as good as that of the MLE. When the innovations follow the gamma distribution, the gamma MLE performs better than the other estimators for both $n = 98$ and $n = 490$. The GLS2 estimator of ρ_0 has a slightly smaller SD than the GLS1 for both sample sizes considered. The BGMME of β_0 has a smaller SD and RMSE than the GLS estimators and the Gaussian QMLE for both $n = 98$ and $n = 490$. For both sample sizes, the percentage reduction in the SD of the BGMME of β_0 relative to the Gaussian QMLE is about 20%. The average CPU time for one repetition is also reported for each estimation method.[15] The GMME significantly reduces the CPU time cost relative to the QMLE.

The second case we consider is $\lambda_0 = 0.3$ and $\rho_0 = 0$, so that the true data generating process in (7) corresponds to the MRSAR model with i.i.d. disturbances. The estimators considered are (i) the 2SLS estimator with IV set $Q_n = (X_n, W_nX_n, W_n^2X_n)$ and the B2SLS estimator with IV set $Q_n = (X_n, \hat{G}_nX_n\hat{\beta}_n)$, (ii) the BGMME in Proposition 2, (iii) the Gaussian QMLE, and (iv) the gamma MLE. We use initial estimates from the 2SLS to implement the B2SLS and feasible GMM estimations.

Table 2 reports the estimation results for the MRSAR model with i.i.d. disturbances. The 2SLS and B2SLS estimators of λ_0 have

12 We have also experimented with the specification where the regressors are fixed across the replications. The simulation results are similar to those reported here.

13 The QMLEs for the regression model with SAR disturbances, the MRSAR model with i.i.d. disturbances, and the MRSAR model with SAR disturbances are calculated, respectively, using sem.m, sar.m, 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.

14 We assume the scale parameter of the gamma density is known and estimate the shape parameter with other unknown parameters in the model using the likelihood method.

15 All the computation is performed using Dell Optiplex 755 with Intel (R) Core (TM) 2 Duo CPU E6850 @ 3.00 GHz and 3.25 GB of RAM.

Table 2
The MRSAR model with i.i.d. disturbances ($\rho_0 = 0$).

	$\lambda_0 = 0.3$	$\beta_{10} = 1.0$	$\beta_{20} = -1.0$	Time (s)
$n = 98$	Normal			
2SLS	0.313(0.176)[0.177]	0.989(0.146)[0.147]	−0.990(0.149)[0.149]	0.0002
B2SLS	0.270(0.214)[0.216]	0.990(0.147)[0.148]	−0.990(0.149)[0.149]	0.0025
BGMM	0.320(0.117)[0.119]	0.987(0.150)[0.151]	−0.991(0.154)[0.155]	0.0188
Gaussian ML	0.287(0.107)[0.108]	0.996(0.145)[0.145]	−0.996(0.147)[0.147]	0.0426
$n = 490$				
2SLS	0.296(0.080)[0.080]	0.998(0.064)[0.064]	−0.996(0.064)[0.064]	0.0022
B2SLS	0.290(0.080)[0.081]	0.998(0.064)[0.064]	−0.996(0.064)[0.064]	0.0399
BGMM	0.301(0.047)[0.047]	0.997(0.065)[0.065]	−0.994(0.064)[0.065]	0.1804
Gaussian ML	0.294(0.046)[0.046]	0.999(0.064)[0.064]	−0.997(0.063)[0.063]	0.2009
$n = 98$	Gamma			
2SLS	0.321(0.172)[0.173]	0.995(0.145)[0.145]	−1.002(0.146)[0.146]	0.0002
B2SLS	0.277(0.198)[0.199]	0.996(0.144)[0.144]	−1.002(0.147)[0.147]	0.0025
BGMM	0.319(0.102)[0.104]	0.996(0.114)[0.114]	−0.999(0.115)[0.115]	0.0194
Gaussian QML	0.290(0.103)[0.104]	1.002(0.143)[0.143]	−1.008(0.145)[0.145]	0.0421
Gamma ML	0.301(0.081)[0.081]	1.001(0.093)[0.093]	−1.002(0.093)[0.093]	0.0340
$n = 490$				
2SLS	0.303(0.076)[0.076]	0.995(0.064)[0.064]	−1.001(0.062)[0.063]	0.0022
B2SLS	0.297(0.077)[0.077]	0.995(0.064)[0.064]	−1.002(0.063)[0.063]	0.0400
BGMM	0.305(0.041)[0.041]	0.997(0.050)[0.050]	−1.000(0.050)[0.050]	0.1806
Gaussian QML	0.298(0.045)[0.045]	0.996(0.064)[0.064]	−1.003(0.062)[0.062]	0.2053
Gamma ML	0.299(0.027)[0.027]	1.000(0.030)[0.030]	−1.000(0.030)[0.030]	0.0894

Mean (SD) [RMSE].

much larger SDs than the other estimators for both sample sizes considered. When the disturbances are normally distributed, the BGMME of λ_0 has a bigger SD than the Gaussian MLE for a small sample size $n = 98$. The performance of the BGMME is as good as the MLE for $n = 490$. When the innovations follow the gamma distribution, the gamma MLE performs the best. The BGMME improves upon the Gaussian QMLE in terms of SD and RMSE for both sample sizes considered. When $n = 98$, the SD of the BGMME of β_0 is about 20% smaller than that of the Gaussian QMLE. When $n = 490$, the percentage reductions in SDs of the BGMMEs of λ_0, β_{10} and β_{20} relative to the Gaussian QMLEs are, respectively, 8.9%, 21.9% and 19.4%.[16]

Lastly, we consider the case that $\lambda_0 = 0.3$ and $\rho_0 = 0.3$. The estimators considered are: (i) the G2SLS estimator in Kelejian and Prucha (1998) and the best G2SLS (B2SLS) estimator in Lee (2003),[17] (ii) the BGMME[18] in Proposition 2, (iii) the Gaussian QMLE, and (iv) the gamma MLE. We use preliminary estimates from the G2SLS to implement the B2SLS and the feasible BGMM.

The estimation results of the MRSAR model with SAR disturbances is given in Table 3. When the disturbances are normally distributed, the Gaussian MLE performs better than the BGMME if the sample size is small, and the BGMME is as good as the MLE if the sample size is moderate. When the innovations follow the gamma distribution, the BGMMEs of λ_0 and ρ_0 have bigger SDs than the Gaussian QMLEs but the BGMME of β_0 has a smaller SD than the QMLE if $n = 98$, and the BGMMEs of λ_0, ρ_0, and β_0 have smaller SDs than the Gaussian QMLEs if $n = 490$. Table 3 also reports some results with misspecifications in that the effect captured by either λ_0 or ρ_0 were ignored, and the restricted models are estimated. When the model is estimated under the restriction that $\lambda_0 = 0$, the various estimators of ρ_0 are biased upwards by about 80%. The estimates of β_0 are only trivially affected. On the other hand, when the model is estimated under the restriction that $\rho_0 = 0$, the QMLE and BGMME of λ_0 are upwards biased, while the G2SLS and B2SLS estimators are quite robust to this misspecification. For both misspecified models, the finite sample performance of the BGMME is as good as the MLE when ϵ_{ni}'s are normally distributed, and the BGMME of β_0 has a smaller SD than the Gaussian QMLE when the innovations follow the gamma distribution.

In summary, in the absence of specific and correct knowledge of the underlying distribution, the BGMME improves on the Gaussian QMLE as the former incorporates correlation between linear and quadratic moment conditions when the disturbances are skewed. The BGMMEs of both the spatial effects λ_0 and ρ_0 and regression coefficient β_0 have smaller SDs and RMSEs than the Gaussian QMLE for a moderate-sized sample. The BGMME is also computationally more efficient than the Gaussian QMLE.

4. Conclusion

In this paper, we consider improved GMM estimation of regression and MRSAR models with SAR disturbances. When the disturbances are normally distributed, the MLE approach for such models is efficient. Lee (2007) has shown the existence of the GMME based on linear and quadratic moment conditions that can attain the same limiting distribution as the MLE. When the disturbances are not normally distributed, the MLE based on the normal likelihood specification is a QMLE. This paper improves upon the QMLE approach by incorporating potential skewness and kurtosis of the disturbances into the linear and quadratic moment conditions used in the GMM framework. The proposed BGMME is asymptotically as efficient as MLE under normality, and more

[16] Given the data generating process of X_n, we evaluate the asymptotic variance of the BGMMEs and QMLEs in addition to the empirical SDs. With the gamma distribution and $n = 490$, the percentage of reductions in asymptotic SDs of the BGMMEs of λ_0, β_{10} and β_{20} relative to the QMLEs are, on average over all the repetitions, respectively, 9.8%, 22.5% and 22.5%.

[17] We use $Q_n = (X_n, W_nX_n, W_n^2X_n)$ as the IV matrix for the G2SLS.

[18] In the Monte Carlo experiments, as $W_n = M_n$, $\bar{G}_n = R_nG_nR_n^{-1} = (I_n - \rho_0M_n)M_n(I_n - \lambda_0M_n)^{-1}(I_n - \rho_0M_n)^{-1} = M_n(I_n - \rho_0M_n)(I_n - \lambda_0M_n)^{-1}(I_n - \rho_0M_n)^{-1} = H_n$ if $\lambda_0 = \rho_0$. Though the estimated $\bar{G}_n$ and H_n would not be exactly the same, they can be very close to each other and the finite sample performance of the BGMME might be affected. So we use linear combinations of the moment functions in Proposition 1 in this Monte Carlo study. The linear combinations are given in (4), and can be shown asymptotically equivalent to those in Proposition 1.

Table 3
The MRSAR model with SAR disturbances.

	$\lambda_0 = 0.3$	$\rho_0 = 0.3$	$\beta_{10} = 1.0$	$\beta_{20} = -1.0$	Time (s)
$n = 98$	Normal				
G2SLS	0.345(0.207)[0.212]	0.197(0.240)[0.261]	0.992(0.147)[0.147]	−0.990(0.149)[0.150]	0.0076
B2SLS	0.332(2.28)[2.28]	0.197(0.240)[0.261]	0.997(0.285)[0.285]	−0.997(0.422)[0.422]	0.0026
BGMM	0.243(0.309)[0.315]	0.318(0.324)[0.324]	0.976(0.161)[0.163]	−0.974(0.162)[0.164]	0.0394
Gaussian ML	0.284(0.206)[0.206]	0.261(0.241)[0.244]	0.990(0.146)[0.147]	−0.988(0.146)[0.147]	0.1110
$n = 490$					
G2SLS	0.301(0.094)[0.094]	0.285(0.109)[0.110]	0.998(0.063)[0.063]	−0.996(0.064)[0.064]	0.0076
B2SLS	0.289(0.094)[0.095]	0.285(0.109)[0.110]	0.997(0.063)[0.063]	−0.995(0.064)[0.064]	0.0382
BGMM	0.287(0.098)[0.099]	0.306(0.109)[0.110]	0.997(0.064)[0.064]	−0.994(0.064)[0.065]	0.3311
Gaussian ML	0.291(0.094)[0.094]	0.296(0.107)[0.107]	0.997(0.063)[0.063]	−0.995(0.064)[0.064]	0.5505
$n = 490$	Estimated model: the regression model with SAR disturbances				
GLS1	–	0.538(0.042)[0.241]	0.950(0.060)[0.078]	−0.948(0.060)[0.080]	0.0074
GLS2	–	0.535(0.041)[0.239]	0.950(0.060)[0.078]	−0.948(0.060)[0.080]	0.0422
BGMM	–	0.556(0.043)[0.260]	0.948(0.061)[0.080]	−0.946(0.061)[0.082]	0.1668
Gaussian ML	–	0.545(0.042)[0.249]	0.949(0.060)[0.079]	−0.947(0.060)[0.080]	0.4792
$n = 490$	Estimated model: the MRSAR model with i.i.d. disturbances				
2SLS	0.300(0.096)[0.096]	–	0.995(0.065)[0.066]	−0.994(0.065)[0.065]	0.0033
B2SLS	0.286(0.099)[0.100]	–	0.996(0.066)[0.066]	−0.995(0.065)[0.065]	0.0406
BGMM	0.481(0.042)[0.186]	–	0.975(0.067)[0.071]	−0.972(0.065)[0.071]	0.1737
Gaussian ML	0.471(0.041)[0.176]	–	0.985(0.064)[0.066]	−0.983(0.064)[0.066]	0.2013
$n = 98$	Gamma				
G2SLS	0.350(0.208)[0.214]	0.194(0.229)[0.252]	0.996(0.144)[0.144]	−1.003(0.146)[0.146]	0.0074
B2SLS	0.259(0.521)[0.522]	0.194(0.229)[0.252]	0.993(0.151)[0.151]	−0.999(0.165)[0.165]	0.0024
BGMM	0.251(0.295)[0.299]	0.315(0.301)[0.301]	0.984(0.130)[0.131]	−0.986(0.130)[0.131]	0.0413
Gaussian QML	0.291(0.205)[0.205]	0.258(0.241)[0.244]	0.993(0.144)[0.144]	−1.001(0.146)[0.146]	0.1107
Gamma ML	0.302(0.169)[0.169]	0.271(0.207)[0.209]	0.996(0.099)[0.099]	−0.995(0.105)[0.105]	0.1030
$n = 490$					
G2SLS	0.309(0.090)[0.090]	0.280(0.107)[0.109]	0.995(0.064)[0.064]	−1.002(0.062)[0.062]	0.0075
B2SLS	0.298(0.090)[0.090]	0.280(0.107)[0.109]	0.995(0.064)[0.064]	−1.002(0.062)[0.062]	0.0381
BGMM	0.299(0.069)[0.069]	0.299(0.087)[0.087]	0.997(0.050)[0.050]	−1.000(0.049)[0.049]	0.3353
Gaussian QML	0.299(0.090)[0.090]	0.291(0.106)[0.106]	0.995(0.064)[0.064]	−1.001(0.062)[0.062]	0.5512
Gamma ML	0.299(0.046)[0.047]	0.299(0.061)[0.061]	0.999(0.031)[0.031]	−0.999(0.030)[0.030]	0.2859
$n = 490$	Estimated model: the regression model with SAR disturbances				
GLS1	–	0.539(0.043)[0.243]	0.946(0.060)[0.081]	−0.952(0.058)[0.075]	0.0096
GLS2	–	0.538(0.042)[0.241]	0.946(0.060)[0.081]	−0.953(0.058)[0.075]	0.0427
BGMM	–	0.559(0.043)[0.262]	0.946(0.048)[0.072]	−0.949(0.048)[0.070]	0.1656
Gaussian QML	–	0.548(0.042)[0.252]	0.945(0.060)[0.082]	−0.951(0.058)[0.076]	0.5083
Gamma ML	–	0.559(0.039)[0.262]	0.946(0.046)[0.071]	−0.946(0.046)[0.071]	0.0910
$n = 490$	Estimated model: the MRSAR model with i.i.d. disturbances				
2SLS	0.308(0.093)[0.093]	–	0.994(0.066)[0.066]	−1.000(0.064)[0.064]	0.0034
B2SLS	0.294(0.096)[0.097]	–	0.994(0.066)[0.066]	−1.001(0.064)[0.064]	0.0407
BGMM	0.460(0.040)[0.165]	–	0.979(0.052)[0.056]	−0.982(0.051)[0.054]	0.1752
Gaussian QML	0.474(0.041)[0.179]	–	0.983(0.064)[0.066]	−0.990(0.063)[0.064]	0.2000
Gamma ML	0.438(0.040)[0.144]	–	0.988(0.045)[0.047]	−0.988(0.047)[0.048]	0.1271

Mean (SD) [RMSE].

efficient than the QMLE when the innovations are not normal. Monte Carlo studies show that the potential inefficiency of the QMLE in finite sample for the MRSAR model mainly comes from the possible correlation between linear and quadratic moment conditions in the likelihood function. Hence, the proposed BGMME has its biggest advantage when the skewness of the disturbances is nonzero. In the event that the diagonal elements of H_n have enough variation,[19] then, taking into account kurtosis may also be valuable.

In the Monte Carlo studies, the (infeasible) exact MLE performs better than the Gaussian QMLE and the BGMME for the case of non-normal errors, which suggests the possibility to further improve the efficiency of the Gaussian QMLE by considering higher order moment conditions in the GMM framework. However, some complications would occur as more high order moment conditions are used for the GMM estimation. First, additional high order moments of the unknown innovation distribution might involve more unknown parameters for estimation. Second, the finite sample properties of the GMM estimator can be sensitive to the number of moment conditions. And as the number of moment conditions increases with the sample size, the GMM estimator could even be asymptotically biased (Han and Phillips, 2006). A more difficult problem in the literature of GMM estimators with many moments occurs when the (optimal) weighting matrix involves preliminary estimates of parameters nonlinearly (see Han and Phillips, 2006, for a discussion). It would be quite difficult if not impossible to derive the asymptotic properties of such an estimator. As the optimal weighting matrix of the moment

19 H_n can be expanded as $H_n = M_n(I_n - \rho_0 M_n)^{-1} = M_n + \lambda_0 M_n^2 + \cdots$. As $D(M_n) = 0$, the empirical variance of the diagonal elements of H_n is largely determined by that of M_n^2.

conditions of the BGMME in this paper involve initial estimates, we expect this technical difficulty would occur if many higher moments are considered.

The models considered so far in this paper have concentrated on the regression and MRSAR models with SAR disturbances, where the spatial lags are all of the first order, i.e., there is a single spatial weights matrix in the main equation or the disturbance process. It is of interest to consider models with high order spatial lags. Those models would be more complicated in structure, which will result in more complex identification and estimation issues. The details will be reported in a separate paper.

Appendix A. Summary of notation

- $D(A) = \text{Diag}(A)$ is a diagonal matrix with diagonal elements being A if A is a vector, or diagonal elements of A if A is a square matrix.
- $\text{vec}_D(A)$ is a column vector formed by the diagonal elements of a square matrix A.
- $A^{(s)} = A + A'$, where A is a square matrix.
- $A^{(t)} = A - \frac{1}{n}\text{tr}(A)I_n$, where A is an $n \times n$ matrix.
- $A^{(l)}$ is a linearly transformed matrix of A that preserves the uniform boundedness property.
- $\theta' = (\rho, \lambda, \beta')$; $\theta_0' = (\rho_0, \lambda_0, \beta_0')$. $\delta' = (\theta', \sigma^2)$; $\delta_0' = (\theta_0', \sigma_0^2)$.
- $R_n(\rho) = I_n - \rho M_n$; $R_n = R_n(\rho_0)$. $S_n(\lambda) = I_n - \lambda W_n$; $S_n = S_n(\lambda_0)$.
- $H_n(\rho) = M_n R_n^{-1}(\rho)$; $H_n = H_n(\rho_0)$. $G_n(\lambda) = W_n S_n^{-1}(\lambda)$; $G_n = G_n(\lambda_0)$.
- $\bar{X}_n(\rho) = R_n(\rho)X_n$; $\bar{X}_n = R_n X_n$. $\bar{G}_n(\rho, \lambda) = R_n(\rho)G_n(\lambda)R_n^{-1}(\rho)$; $\bar{G}_n = R_n G_n R_n^{-1}$.
- If an intercept appears in $\bar{X}_n$, we have $\bar{X}_n = [\bar{X}_n^*, c(\rho_0)l_n]$. Otherwise $\bar{X}_n^* \equiv \bar{X}_n$.
- l_n is an $n \times 1$ vector of ones. $J_n = I_n - \frac{1}{n}l_n l_n'$. e_{kj} is the jth unit vector in R^k.
- For the MRSAR model with SAR disturbances, $\epsilon_n(\theta) = R_n(\rho)[S_n(\lambda)Y_n - X_n\beta]$, $g_n(\theta) = (Q_n, P_{1n}\epsilon_n(\theta), \ldots, P_{mn}\epsilon_n(\theta))'\epsilon_n(\theta)$, and $\Omega_n = \text{var}(g_n(\theta_0))$. The class of GMMEs of θ_0 that minimize $g_n'(\theta)\Omega_n^{-1}g_n(\theta)$ is denoted by $\mathcal{M}_n$.
- For the regression model with SAR disturbances, $\epsilon_{\rho n}(\rho, \beta) = R_n(\rho)(Y_n - X_n\beta)$, $g_{\rho n}(\rho, \beta) = (Q_n, P_{1n}\epsilon_{\rho n}(\rho, \beta), \ldots, P_{mn}\epsilon_{\rho n}(\rho, \beta))'\epsilon_{\rho n}(\rho, \beta)$, and $\Omega_{\rho n} = \text{var}(g_{\rho n}(\rho_0, \beta_0))$. The class of GMMEs of (ρ_0, β_0') that minimize $g_{\rho n}'(\rho, \beta)\Omega_{\rho n}^{-1}g_{\rho n}(\rho, \beta)$ is denoted by $\mathcal{M}_{\rho n}$.
- For the MRSAR model with i.i.d. disturbances, $\epsilon_{\lambda n}(\lambda, \beta) = S_n(\lambda)Y_n - X_n\beta$, $g_{\lambda n}(\lambda, \beta) = (Q_n, P_{1n}\epsilon_{\lambda n}(\lambda, \beta), \ldots, P_{mn}\epsilon_{\lambda n}(\lambda, \beta))'\epsilon_{\lambda n}(\lambda, \beta)$, and $\Omega_{\lambda n} = \text{var}(g_{\lambda n}(\lambda_0, \beta_0))$. The class of GMMEs of (λ_0, β_0') that minimize $g_{\lambda n}'(\lambda, \beta)\Omega_{\lambda n}^{-1}g_{\lambda n}(\lambda, \beta)$ is denoted by $\mathcal{M}_{\lambda n}$.

Appendix B. FGLS and MOM estimation of the regression model with SAR disturbances

The regression model with SAR disturbances is a generalized linear model with variance $\sigma_0^2 R_n^{-1}R_n'^{-1}$ for u_n and the parameter of interest in this discussion is ρ_0. A consistent estimator of ρ_0 can be used as an initial estimator for the FGLS estimation of the regression coefficient β_0. Kelejian and Prucha (1999) have considered the MOM estimation of ρ_0 and the FGLS estimation of β_0. If the purpose is solely for the estimation of β_0 via the GLS, efficient estimation of ρ_0 is not an issue as the asymptotic distribution of the FGLS estimator does not depend on the asymptotic distribution of the initial consistent estimator of ρ_0. However, efficiency in estimation of ρ_0 improves the power of tests for the presence of SAR disturbances (the test for $\rho_0 = 0$) as well as other inference on ρ_0.

B.1. FGLS estimation

Let $\hat{\beta}_L = (X_n'X_n)^{-1}X_n'Y_n$ be the OLS estimator. u_n can be estimated by the estimated residual $\hat{u}_n = Y_n - X_n\hat{\beta}_L$. Following Lee (2001a), ρ_0 can then be estimated by the GMM:

$$\hat{\rho}_P = \arg\min_{\rho} \hat{g}_n'(\rho)a_n'a_n\hat{g}_n(\rho), \tag{8}$$

based on the quadratic moment conditions of ϵ_n

$$\hat{g}_n(\rho) = [P_{1n}R_n(\rho)\hat{u}_n, \ldots, P_{mn}R_n(\rho)\hat{u}_n]'R_n(\rho)\hat{u}_n, \tag{9}$$

where P_{jn}'s are $n \times n$ constant matrices such that $\text{tr}(P_{jn}) = 0$ for $j = 1, \ldots, m$.

Under Assumptions 1–7, Lee (2001a) has shown that the GMME $\hat{\rho}_P$ is $\sqrt{n}$-consistent and has a limiting distribution equivalent to the GMME when u_n is observed. Furthermore, with a consistent estimator of ρ_0, the FGLS estimator $\hat{\beta}_{FG} = (X_n'\hat{R}_n'\hat{R}_nX_n)^{-1}X_n'\hat{R}_n'\hat{R}_nY_n$ is asymptotically equivalent to the exact GLS estimator $\hat{\beta}_G = (X_n'R_n'R_nX_n)^{-1}X_n'R_n'R_nY_n$.

B.2. BMOM estimation

Within the class of GMMEs given by (8), efficiency hinges on the selection of P_{jn}'s. Lee (2001a) gives the best one when ϵ_n is normally distributed. Here, we derive the BGMM (or BMOM) estimator within this class without the normality assumption. The optimal choice of the weighting matrix $a_n'a_n$ in (8) is, as usual, a matrix proportional to Ω_n^{-1}. The approach used in the general model above hinges on the characterization of best moments in terms of any additional moments being redundant in Breusch et al. (1999). In this section, we derive the analytically best P_n^* directly. Let $\mathcal{M}_n$ be the class of optimal GMMEs from $\min_{\rho\in\Lambda} g_n'(\rho)\Omega_n^{-1}g_n(\rho)$, where $g_n(\rho) = [P_{1n}R_n(\rho)u_n, \ldots, P_{mn}R_n(\rho)u_n]'R_n(\rho)u_n$ is a vector of moment functions with P_{jn}'s satisfying Assumption 4. We are interested in the BGMME within $\mathcal{M}_n$ without any distributional assumption. Following Lee (2001a), the asymptotic variance of the consistent GMME $\sqrt{n}\hat{\rho}_P$ based on the quadratic moment $u_n'R_n'(\rho)P_nR_n(\rho)u_n$ with $\text{tr}(P_n) = 0$ is $(\lim_{n\to\infty}\frac{1}{n}\Sigma_{P,n})^{-1}$, where $\Sigma_{P,n} = \frac{\text{tr}^2(P_n^{(s)}H_n)}{(\eta_4-3)\sum_{i=1}^n P_{n,ii}^2 + \text{tr}(P_nP_n^{(s)})}$. The best P_n with $\text{tr}(P_n) = 0$ will minimize the asymptotic variance or, equivalently, maximize the corresponding precision measure $\Sigma_{P,n}$. As $\text{tr}(P_n^{(s)}P_n) = \text{tr}[(P_n - D(P_n))^{(s)}P_n] + 2\sum_{i=1}^n P_{n,ii}^2$, the denominator of $\Sigma_{P,n}$ is $(\eta_4 - 3)\sum_{i=1}^n P_{n,ii}^2 + \text{tr}(P_n^{(s)}P_n) = (\eta_4 - 1)\sum_{i=1}^n P_{n,ii}^2 + \text{tr}[(P_n - D(P_n))^{(s)}P_n]$, where $\eta_4 > 1$ by Jensen's inequality for a concave function. Let $P_n^+ = P_n + \left(\sqrt{\frac{1}{2}(\eta_4 - 1)} - 1\right)D(P_n)$. As $\text{tr}(P_n) = 0$, $\text{tr}(P_n^+) = 0$. The square of the Euclidean norm of $(P_n^+)^{(s)}$ is $\text{tr}[(P_n^+)^{(s)}(P_n^+)^{(s)}] = 2\left\{(\eta_4 - 1)\sum_{i=1}^n P_{n,ii}^2 + \text{tr}[(P_n - D(P_n))^{(s)}P_n]\right\}$. P_n and P_n^+ have a one-to-one relation. Given P_n^+, P_n can be recovered as $P_n = P_n^+ + \left(\sqrt{\frac{2}{\eta_4-1}} - 1\right)D(P_n^+)$. Because $\text{tr}(P_n^{(s)}H_n) = \text{tr}(P_n^{(s)}H_n^{(t)}) = \frac{1}{2}\text{tr}(P_n^{(s)}(H_n^{(t)})^{(s)})$, the maximization problem is thus equivalent to $\max_{P_n^+} \frac{\text{tr}^2\{[P_n^+ + (\sqrt{2/(\eta_4-1)}-1)D(P_n^+)]^{(s)}(H_n^{(t)})^{(s)}\}}{\text{tr}[(P_n^+)^{(s)}(P_n^+)^{(s)}]}$. To solve this, we shall look for a matrix A_n such that $\text{tr}\left\{\left[P_n^+ + \left(\sqrt{\frac{2}{\eta_4-1}} - 1\right)D(P_n^+)\right]^{(s)}(H_n^{(t)})^{(s)}\right\} = \text{tr}[(P_n^+)^{(s)}(H_n^{(t)} + A_n)^{(s)}]$. This identity is equivalent to $\left(\sqrt{\frac{2}{\eta_4-1}} - 1\right)\text{tr}[D(P_n^+)^{(s)}(H_n^{(t)})^{(s)}] = \text{tr}[(P_n^+)^{(s)}A_n^{(s)}]$. If A_n is taken to be a diagonal matrix, then $\text{tr}[(P_n^+)^{(s)}A_n^{(s)}] = \text{tr}[D(P_n^+)^{(s)}A_n^{(s)}]$. One possible A_n is $A_n = \left(\sqrt{\frac{2}{\eta_4-1}} - 1\right)D(H_n^{(t)})$, which is determined by H_n alone. Thus the optimization becomes $\max_{P_n^+} \frac{\text{tr}^2[(P_n^+)^{(s)}(H_n^{(t)} + A_n)^{(s)}]}{\text{tr}[(P_n^+)^{(s)}(P_n^+)^{(s)}]}$. For any square conformable matrices B and C, $\text{tr}^2(BC) \le \text{tr}(B^2)\text{tr}(C^2)$ is a version of the Cauchy inequality. Hence the optimum P_n^+ is $P_n^{+*} = H_n^{(t)} + A_n = H_n^{(t)} + \left(\sqrt{\frac{2}{\eta_4-1}} - 1\right)D(H_n^{(t)})$. In terms of the original P_n^*, one has

$P_n^* = P_n^{+*} + \left(\sqrt{\frac{2}{\eta_4-1}} - 1\right) D(P_n^{+*}) = H_n^{(t)} - \frac{\eta_4-3}{\eta_4-1} D(H_n^{(t)})$, because $D(P_n^{+*}) = \sqrt{\frac{2}{\eta_4-1}} D(H_n^{(t)})$.

The form of the best P_n^* here motivates the selection of best moments for the regression model. The following proposition gives the BMOM estimator of ρ_0 for the SAR process.

Proposition 5. *Under Assumptions 1–7, $\hat{\rho}_B = \arg\min_{\rho\in\Lambda}[u_n' R_n'(\rho) P_n^* R_n(\rho) u_n]^2$ is the BMOM estimator within $\mathcal{M}_n$, with $\sqrt{n}(\hat{\rho}_B - \rho_0) \xrightarrow{D} N(0, \Sigma_B^{-1})$ and $\Sigma_B = \lim_{n\to\infty} \frac{1}{n}\operatorname{tr}(P_n^{*(s)} H_n)$.*

B.3. Variance reduction

Let $\mathcal{P}_{1n}$ be the class of constant $n \times n$ matrices P_{jn}'s satisfying Assumption 4. When ϵ_n is normally distributed, Lee (2001a) has shown that $H_n^{(t)}$ is the best selection in $\mathcal{P}_{1n}$. This is the special case of P_n^* in Proposition 5 with $\eta_4 = 3$. Furthermore, Lee (2001a) has shown that the GMME $\hat{\rho}_{H1}$ based on the quadratic moment $u_n' R_n'(\rho) H_n^{(t)} R_n(\rho) u_n$ has the same limiting distribution as the QML derived from $\max \ln L_n(\rho, \sigma^2)$, where $L_n(\rho, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} |R_n(\rho)| \exp\left(-\frac{1}{2\sigma^2} u_n' R_n'(\rho) R_n(\rho) u_n\right)$, regardless of ϵ_n's distribution. Thus it is of interest to compare the efficiency gain of the BGMME $\hat{\rho}_B$ with $\hat{\rho}_{H1}$. The limiting variance of $\sqrt{n}\hat{\rho}_{H1}$ is $\Sigma_{H1}^{-1} = \left(\lim_{n\to\infty} \frac{1}{n}\Sigma_{H1,n}\right)^{-1}$, where $\Sigma_{H1,n} = \frac{\operatorname{tr}^2[(H_n^{(t)})^{(s)} H_n]}{(\eta_4-3)\sum_{i=1}^n (H_{n,ii}^{(t)})^2 + \operatorname{tr}[(H_n^{(t)})^{(s)} H_n]}$. The limiting variance of $\sqrt{n}\hat{\rho}_B$ is $(\lim_{n\to\infty} \frac{1}{n}\Sigma_{B,n})^{-1}$, where $\Sigma_{B,n} = \operatorname{tr}(P_n^{*(s)} H_n) = \operatorname{tr}[(H_n^{(t)})^{(s)} H_n] - 2\left(\frac{\eta_4-3}{\eta_4-1}\right)\operatorname{tr}[D(H_n^{(t)}) H_n]$. To simplify notation, denote

$$v_H^2 = v^2(H) = \frac{1}{n}\sum_{i=1}^n (H_{n,ii}^{(t)})^2 = \frac{1}{n}\sum_{i=1}^n \left(H_{n,ii} - \frac{1}{n}\sum_{j=1}^n H_{n,jj}\right)^2 \tag{10}$$

the empirical variance formed by the diagonal elements of H_n. Furthermore, denote $l_{H,1}^2 = \frac{1}{n}\operatorname{tr}[(H_n^{(t)})^{(s)} H_n] = \frac{1}{2n}\operatorname{tr}[(H_n^{(t)})^{(s)} (H_n^{(t)})^{(s)}]$ and $l_{H,2}^2 = \frac{1}{n}\operatorname{tr}[(H_n - D(H_n))^{(s)} H_n] = \frac{1}{2n}\operatorname{tr}[(H_n - D(H_n))^{(s)} (H_n - D(H_n))^{(s)}]$, which are, respectively, $\frac{1}{2n}$ of the square of the Euclidean norm of $(H_n^{(t)})^{(s)}$ and $(H_n - D(H_n))^{(s)}$.

Instead of comparing the variances of these two estimators we compare the precision measures $\frac{1}{n}\Sigma_{H1,n}$ and $\frac{1}{n}\Sigma_{B,n}$. As $\frac{1}{n}\Sigma_{H1,n} = l_{H,1}^4/[(\eta_4-3)v_H^2 + l_{H,1}^2]$ and $\frac{1}{n}\Sigma_{B,n} = l_{H,1}^2 - 2\left(\frac{\eta_4-3}{\eta_4-1}\right)v_H^2$, it follows that

$$\frac{1}{n}\Sigma_{B,n} - \frac{1}{n}\Sigma_{H1,n} = \frac{(\eta_4-3)^2 v_H^2 (l_{H,1}^2 - 2v_H^2)}{(\eta_4-1)[(\eta_4-3)v_H^2 + l_{H,1}^2]} = \frac{(\eta_4-3)^2 v_H^2 l_{H,2}^2}{(\eta_4-1)[(\eta_4-1)v_H^2 + l_{H,2}^2]},$$

because $l_{H,1}^2 - l_{H,2}^2 = \frac{1}{n}\operatorname{tr}[(H_n^{(t)})^{(s)} H_n] - \frac{1}{n}\operatorname{tr}[(H_n - D(H_n))^{(s)} H_n] = \frac{1}{n}\operatorname{tr}\left[\left(D(H_n) - \frac{\operatorname{tr}(H_n)}{n} I_n\right)^{(s)} H_n\right] = 2v_H^2$. As $\eta_4 > 1$ and $l_{H,2}^2 > 0$, it follows that $\frac{1}{n}\Sigma_{B,n} \geq \frac{1}{n}\Sigma_{H1,n}$. Hence $\hat{\rho}_B$ is efficient relative to $\hat{\rho}_{H1}$. When $\eta_4 \neq 3$, the percentage loss of asymptotic efficiency of $\hat{\rho}_{H1}$ can be evaluated as

$$1 - \frac{\Sigma_{H1,n}}{\Sigma_{B,n}} = \frac{(\eta_4-3)^2 v_H^2 l_{H,2}^2}{[(\eta_4-1)v_H^2 + l_{H,2}^2] \cdot [4v_H^2 + (\eta_4-1) l_{H,2}^2]}. \tag{11}$$

Note that the variance is the inverse of the precision measure. So, $1 - \frac{\Sigma_{H1,n}}{\Sigma_{B,n}} = \frac{\Sigma_{H1,n}^{-1} - \Sigma_{B,n}^{-1}}{\Sigma_{H1,n}^{-1}}$ is also the percentage of reduction in asymptotic variance of $\hat{\rho}_B$ relative to $\hat{\rho}_{H1}$.

A subclass $\mathcal{P}_{2n}$ of $\mathcal{P}_{1n}$ consisting of P_{jn}'s with a zero diagonal is also interesting, as the corresponding GMME is robust against unknown heteroskedasticity (Lin and Lee, 2010) and distributional assumptions. Lee (2001a) has shown the best selection of P_n from $\mathcal{P}_{2n}$ is $H_n - D(H_n)$. Similarly, we can compare the efficiency gain of $\hat{\rho}_B$ relative to the GMME $\hat{\rho}_{H2}$ derived based on the quadratic moment $u_n' R_n'(\rho)(H_n - D(H_n)) R_n(\rho) u_n$. Following Lee (2001a), the limiting variance of $\hat{\rho}_{H2}$ is $\Sigma_{H2}^{-1} = \left(\lim_{n\to\infty} \frac{1}{n}\Sigma_{H2,n}\right)^{-1}$, where $\frac{1}{n}\Sigma_{H2,n} = \frac{1}{n}\operatorname{tr}[(H_n - D(H_n))^{(s)} H_n] = l_{H,2}^2$. It follows that $\frac{1}{n}\Sigma_{B,n} - \frac{1}{n}\Sigma_{H2,n} = l_{H,1}^2 - 2\left(\frac{\eta_4-3}{\eta_4-1}\right)v_H^2 - l_{H,2}^2 = \frac{4}{\eta_4-1}v_H^2$, because $l_{H,1}^2 - l_{H,2}^2 = 2v_H^2$. As $\eta_4 > 1$, we have $\frac{1}{n}\Sigma_{B,n} \geq \frac{1}{n}\Sigma_{H2,n}$. The percentage loss of asymptotic efficiency of $\hat{\rho}_{H2}$ can be evaluated as

$$1 - \frac{\Sigma_{H2,n}}{\Sigma_{B,n}} = \frac{4v_H^2}{4v_H^2 + (\eta_4-1) l_{H,2}^2}, \tag{12}$$

which is also the percentage of reduction in asymptotic variance of $\hat{\rho}_B$ relative to $\hat{\rho}_{H2}$. From this, $\hat{\rho}_B$ is more precise as it takes into account the variation of the diagonal elements of H_n.

Appendix C. Joint GMM estimation approach

Here we consider the joint estimation of σ_0^2 and θ_0 in the GMM framework. Let $\delta = (\theta', \sigma^2)'$. The optimal GMMEs are derived from $\min_\delta \dot{g}_n'(\delta)\dot{\Omega}_n^{-1}\dot{g}_n(\delta)$, where $\dot{\Omega}_n = \operatorname{var}(\dot{g}_n(\delta_0))$ and

$$\dot{g}_n(\delta) = (\epsilon_n'(\theta)\dot{Q}_n, \epsilon_n'(\theta)\dot{P}_{1n}\epsilon_n(\theta) - \sigma^2\operatorname{tr}(\dot{P}_{1n}), \ldots, \epsilon_n'(\theta)\dot{P}_{mn}\epsilon_n(\theta) - \sigma^2\operatorname{tr}(\dot{P}_{mn}))',$$

with $\dot{Q}_n$ being an arbitrary $n \times q$ matrix of IVs, and $\dot{P}_{jn}$'s being arbitrary $n \times n$ matrices, not necessarily with zero traces. At δ_0, $\dot{g}_n(\delta_0) = [\epsilon_n'\dot{Q}_n, \epsilon_n'\dot{P}_{jn}\epsilon_n - \sigma_0^2\operatorname{tr}(\dot{P}_{jn})]'$, which has a zero mean because $E(\dot{Q}_n'\epsilon_n) = \dot{Q}_n'E(\epsilon_n) = 0$ and $E(\epsilon_n'\dot{P}_{jn}'\epsilon_n) = \sigma_0^2\operatorname{tr}(\dot{P}_{jn})$ for $j = 1, \ldots, m$. By comparing the asymptotic variance matrix of the BGMME derived from the joint GMM estimation approach with that of the BGMME in Proposition 1, we conclude that there is no efficiency loss in the estimation of θ_0 by concentrating σ_0^2 out.

For simplicity, we focus on the case that $\bar{X}_n$ does not have a column proportional to l_n so that $\bar{X}_n^* = \bar{X}_n$. When $\bar{X}_n$ has a column proportional to l_n, the result follows by similar arguments. Let $\dot{P}_{1n}^* = \bar{G}_n - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2} D(\bar{G}_n) - \frac{\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2} D(\bar{G}_n\bar{X}_n\beta_0)$, $\dot{P}_{2n}^* = H_n - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2} D(H_n)$, $\dot{P}_{3n}^* = I_n$, $\dot{P}_{j+3,n}^* = D(\bar{X}_{nj}^*)$ for $j = 1, \ldots, k^*$, and $\dot{Q}_n^* = (\dot{Q}_{1n}^*, \dot{Q}_{2n}^*, \dot{Q}_{3n}^*, \dot{Q}_{4n}^*)$, with $\dot{Q}_{1n}^* = \bar{X}_n^*$, $\dot{Q}_{2n}^* = l_n$, $\dot{Q}_{3n}^* = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}\bar{G}_n\bar{X}_n\beta_0 - \frac{2\sigma_0\eta_3}{(\eta_4-1)-\eta_3^2}\operatorname{vec}_D(\bar{G}_n)$ and $\dot{Q}_{4n}^* = \operatorname{vec}_D(H_n)$. Let $\dot{g}_n^*(\delta) = [\epsilon_n'(\theta)\dot{Q}_n^*, \epsilon_n'(\theta)\dot{P}_{1n}^*\epsilon_n(\theta) - \sigma^2\operatorname{tr}(\dot{P}_{1n}^*), \ldots, \epsilon_n'(\theta)\dot{P}_{k^*+3,n}^* \times \epsilon_n(\theta) - \sigma^2\operatorname{tr}(\dot{P}_{k^*+3,n}^*)]'$ and $\dot{\Omega}_n^* = \operatorname{var}(\dot{g}_n^*(\delta_0))$. $\hat{\delta}_{BJ} = \arg\min_\delta \dot{g}_n^{*\prime}(\delta)\dot{\Omega}_n^{*-1}\dot{g}_n^*(\delta)$ is the BGMME within the class of optimal joint GMMEs as shown below.

Analogous to the proof of Proposition 1, the BGMME can be confirmed by showing that there exists a matrix $\dot{A}_n$ invariant with $\dot{P}_{jn}$'s and $\dot{Q}_n$ such that $\dot{D}_2 = \dot{\Omega}_{21}\dot{A}_n$, where

$$\dot{D}_2 = E\left(\frac{\partial}{\partial\delta'}\dot{g}_n(\delta_0)\right) = -\begin{bmatrix} 0 & Q_n'\bar{G}_n\bar{X}_n\beta_0 & Q_n'\bar{X}_n & 0 \\ \sigma_0^2\operatorname{tr}(P_{1n}^{(s)}H_n) & \sigma_0^2\operatorname{tr}(P_{1n}^{(s)}\bar{G}_n) & 0 & \operatorname{tr}(\dot{P}_{1n}) \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_0^2\operatorname{tr}(P_{mn}^{(s)}H_n) & \sigma_0^2\operatorname{tr}(P_{mn}^{(s)}\bar{G}_n) & 0 & \operatorname{tr}(\dot{P}_{mn}) \end{bmatrix},$$

$$\dot{\Omega}_{21} = E\left[\dot{g}_n(\delta_0)\dot{g}_n^{*\prime}(\delta_0)\right]$$
$$= \begin{bmatrix} \sigma_0^2\dot{Q}_n'\dot{Q}_n^* & \mu_3\dot{Q}_n'\mathrm{vec}_D(\dot{P}_{1n}^*) & \cdots & \mu_3\dot{Q}_n'\mathrm{vec}_D(\dot{P}_{k^*+3,n}^*) \\ \mu_3\mathrm{vec}_D'(\dot{P}_{1n})\dot{Q}_n^* & \sigma_0^4\mathrm{tr}(\dot{P}_{1n}^{(s)}\dot{P}_{1n}^*) & \cdots & \sigma_0^4\mathrm{tr}(\dot{P}_{1n}^{(s)}\dot{P}_{k^*+3,n}^*) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_3\mathrm{vec}_D'(\dot{P}_{mn})\dot{Q}_n^* & \sigma_0^4\mathrm{tr}(\dot{P}_{mn}^{(s)}\dot{P}_{1n}^*) & \cdots & \sigma_0^4\mathrm{tr}(\dot{P}_{mn}^{(s)}\dot{P}_{k^*+3,n}^*) \end{bmatrix}$$
$$+ (\mu_4 - 3\sigma_0^4)\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \mathrm{vec}_D'(\dot{P}_{1n})\mathrm{vec}_D(\dot{P}_{1n}^*) & \cdots & \mathrm{vec}_D'(\dot{P}_{1n})\mathrm{vec}_D(\dot{P}_{k^*+3,n}^*) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \mathrm{vec}_D'(\dot{P}_{mn})\mathrm{vec}_D(\dot{P}_{1n}^*) & \cdots & \mathrm{vec}_D'(\dot{P}_{mn})\mathrm{vec}_D(\dot{P}_{k^*+3,n}^*) \end{bmatrix}$$

Box II.

$$\dot{A}_n = -\begin{bmatrix} 0 & 0 & 0 & -\frac{2\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2} & 0 & \sigma_0^{-2} & 0 & 0 \\ 0 & 0 & \sigma_0^{-2} & 0 & \sigma_0^{-2} & 0 & 0 & 0 \\ \frac{\sigma_0^{-2}(\eta_4-1)}{(\eta_4-1)-\eta_3^2}I_{k^*} & 0 & 0 & 0 & 0 & 0 & 0 & b' \\ 0 & -\frac{\sigma_0^{-3}\eta_3}{(\eta_4-1)-\eta_3^2} & 0 & 0 & 0 & 0 & \frac{\sigma_0^{-4}}{(\eta_4-1)-\eta_3^2} & 0 \end{bmatrix}',$$

where $b = (b_1', \ldots, b_{k^*}')'$ with $b_l = -\frac{\sigma_0^{-3}\eta_3}{(\eta_4-1)-\eta_3^2}e_{kl}'$ for $l = 1, \ldots, k^*$.

Box III.

$$\dot{D}_1'\dot{A}_n = \begin{bmatrix} \mathrm{tr}(\dot{P}_{2n}^{*(s)}H_n) & \mathrm{tr}(\dot{P}_{1n}^{*(s)}H_n) & -\frac{2\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2}\mathrm{vec}_D'(H_n)\bar{X}_n & \frac{2\sigma_0^{-2}}{(\eta_4-1)-\eta_3^2}\mathrm{tr}(H_n) \\ * & \sigma_0^{-2}(\bar{G}_n\bar{X}_n\beta_0)'\dot{Q}_{3n}^* + \mathrm{tr}(P_{1n}^{*(s)}\bar{G}_n) & \sigma_0^{-2}\dot{Q}_{3n}^{*\prime}\bar{X}_n & \sigma_0^{-2}\mathrm{tr}(\dot{P}_{1n}^*) \\ * & * & \frac{\sigma_0^{-2}(\eta_4-1)}{(\eta_4-1)-\eta_3^2}\bar{X}_n'\bar{X}_n & -\frac{\sigma_0^{-3}\eta_3}{(\eta_4-1)-\eta_3^2}\bar{X}_n'l_n \\ * & * & * & \frac{n\sigma_0^{-4}}{(\eta_4-1)-\eta_3^2} \end{bmatrix}$$

Box IV.

and $\dot{\Omega}_{21}$ is as given in Box II. Let $\dot{A}_n$ is as given in Box III. Straightforward but tedious algebra leads to $\dot{D}_2 = \dot{\Omega}_{21}\dot{A}_n$. Furthermore, as $\dot{A}_n$ is invariant with $\dot{P}_{jn}$'s and $\dot{Q}_n$, $\dot{\Omega}_{11}^{-1}\dot{D}_1 = \dot{\Omega}_{21}^{-1}\dot{D}_2 = \dot{A}_n$, where $\dot{\Omega}_{11} = \mathrm{var}(\dot{g}_n^*(\delta_0))$ and $\dot{D}_1 = E\left(\frac{\partial}{\partial\delta'}\dot{g}_n^*(\delta_0)\right)$. The asymptotic precision matrix of $\hat{\delta}_{BJ}$ is $\Sigma_{BJ} = \lim_{n\to\infty}\frac{1}{n}\dot{D}_1'\dot{A}_n$, where $\dot{D}_1'\dot{A}_n$ is as given in Box IV. From the inverse of a partitioned matrix, we have $Avar(\hat{\theta}_{BJ}) = (n\Sigma_B)^{-1}$, with Σ_B given in (3). Hence the efficiency property of the BGMME of θ_0 is not affected by concentrating σ^2 out in the GMM estimation.

Appendix D. Some useful lemmas

In this Appendix D, we list some useful lemmas for the proofs of the results in the text. The central limit theorem D.5 is in Kelejian and Prucha (2001). The other properties in Lemmas D.1–D.9 are either trivial or can be found in Lee (2001a, 2004, 2007).

Lemma D.1. *Suppose that z_{1n} and z_{2n} are n-dimensional column vectors of constants which are uniformly bounded. If $\{A_n\}$ is either UBR or UBC, then $|z_{1n}'A_nz_{2n}| = O(n)$.*

Lemma D.2. *Suppose that $\epsilon_{n1}, \ldots, \epsilon_{nn}$ 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,*

$$E(\epsilon_n'A_n\epsilon_n\cdot\epsilon_n'B_n\epsilon_n) = (\mu_4 - 3\sigma^4)\mathrm{vec}_D'(A_n)\mathrm{vec}_D(B_n) + \sigma^4\left[\mathrm{tr}(A_n)\mathrm{tr}(B_n) + \mathrm{tr}(A_nB_n^{(s)})\right],$$

where $B_n^{(s)} = B_n + B_n'$.

Lemma D.3. *Suppose that $\{A_n\}$ is a sequence of $n\times n$ UB matrices, and $\epsilon_{n1}, \ldots, \epsilon_{nn}$ are i.i.d. with zero mean and finite fourth moment. Then, $E(\epsilon_n'A_n\epsilon_n) = O(n)$, $\mathrm{var}(\epsilon_n'A_n\epsilon_n) = O(n)$, $\epsilon_n'A_n\epsilon_n = O_p(n)$, and $\frac{1}{n}\epsilon_n'A_n\epsilon_n - \frac{1}{n}E(\epsilon_n'A_n\epsilon_n) = o_p(1)$.*

Lemma D.4. *Suppose that $\{A_n\}$ is a sequence of $n\times n$ UBC matrices, elements of the $n\times k$ matrix C_n are uniformly bounded, and $\epsilon_{n1}, \ldots, \epsilon_{nn}$ 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 $\frac{1}{n}C_n'A_nA_n'C_n$ exists and is positive definite, then $\frac{1}{\sqrt{n}}C_n'A_n\epsilon_n \xrightarrow{D} N\left(0, \sigma^2\lim_{n\to\infty}\frac{1}{n}C_n'A_nA_n'C_n\right)$.*

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

Lemma D.6. *Suppose that* $\frac{1}{n}(g_n(\lambda) - \bar{g}_n(\lambda)) \to 0$ *in probability uniformly in* $\lambda \in \Lambda$, *which is a compact set, and* $\lim_{n\to\infty} \frac{1}{n}\bar{g}_n(\lambda) = 0$ *has a unique root at* λ_0 *in* Λ. *The* $\hat{\lambda}_n$ *and* $\hat{\lambda}_n^*$ *are, respectively, the roots of* $g_n(\lambda) = 0$ *and* $g_n^*(\lambda) = 0$. *If* $\frac{1}{n}(g_n^*(\lambda) - g_n(\lambda)) = o_p(1)$ *uniformly in* $\lambda \in \Lambda$, *then both* $\hat{\lambda}_n$ *and* $\hat{\lambda}_n^*$ *converge in probability to* λ_0.

In addition, suppose that $\frac{1}{n}\frac{\partial g_n(\lambda)}{\partial \lambda}$ *converges in probability to a well-defined nonzero limit function uniformly in* $\lambda \in \Lambda$, *and* $\frac{1}{\sqrt{n}} g_n(\lambda_0) = O_p(1)$. *If* $\frac{1}{n}\left(\frac{\partial g_n^*(\lambda)}{\partial \lambda} - \frac{\partial g_n(\lambda)}{\partial \lambda}\right) = o_p(1)$ *uniformly in* $\lambda \in \Lambda$, *and* $\frac{1}{\sqrt{n}}(g_n^*(\lambda_0) - g_n(\lambda_0)) = o_p(1)$, *then both* $\sqrt{n}(\hat{\lambda}_n - \lambda_0)$ *and* $\sqrt{n}(\hat{\lambda}_n^* - \lambda_0)$ *have the same limiting distribution.*

Lemma D.7. *Let* $\hat{\theta}_n$ *and* $\hat{\theta}_n^*$ *be, respectively, the minimizers of* $\digamma_n(\theta)$ *and* $\digamma_n^*(\theta)$ *in the compact set* Θ. *Suppose that* $\frac{1}{n}(\digamma_n(\theta) - \bar{\digamma}_n(\theta)) \to 0$ *in probability uniformly in* $\theta \in \Theta$, *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* $\frac{1}{n}\left(\frac{\partial^2 \digamma_n^*(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \digamma_n(\theta)}{\partial\theta\partial\theta'}\right) = o_p(1)$ *uniformly in* $\theta \in \Theta$ *and* $\frac{1}{\sqrt{n}}\left(\frac{\partial \digamma_n^*(\theta_0)}{\partial\theta} - \frac{\partial \digamma_n(\theta_0)}{\partial\theta}\right) = o_p(1)$, *then* $\sqrt{n}(\hat{\theta}_n^* - \theta_0)$ *and* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *have the same limiting distribution.*

Lemma D.8. *Under Assumption 2, the sequences of projectors* $\{Z_n\}$ *and* $\{I_n - Z_n\}$ *with* $Z_n = X_n(X_n'X_n)^{-1}X_n'$ *are UB.*

Lemma D.9. *Suppose that* $\{\|W_n\|\}$, $\{\|M_n\|\}$, $\{\|S_n^{-1}\|\}$, *and* $\{\|R_n^{-1}\|\}$, *where* $\|\cdot\|$ *is a matrix norm, are bounded. 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 properties are specific to the model in this paper.

Lemma D.10. *Suppose that* z_{1n} *and* z_{2n} *are n-dimensional column vectors of constants which are uniformly bounded, the sequence of* $n \times n$ *constant matrices* $\{A_n\}$ *is UBC, and* $\{B_{1n}\}$ *and* $\{B_{2n}\}$ *are UB, and* $\epsilon_{n1}, \ldots, \epsilon_{nn}$ *are i.i.d. with zero mean and finite second moment.* $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_p(1)$, *where* α_0 *is a p-dimensional vector in the interior of its convex parameter space. For notational simplicity, denote* $(\hat{\alpha}_n - \alpha_0)^{\langle i\rangle} = \sum_{j_1=1}^p \cdots \sum_{j_i=1}^p (\hat{\alpha}_{nj_1} - \alpha_{j_10}) \cdots (\hat{\alpha}_{nj_i} - \alpha_{j_i0})$. *The matrix* $C_n(\hat{\alpha}_n)$ *has the expansion that*

$$C_n(\hat{\alpha}_n) - C_n(\alpha_0) = \sum_{i=1}^{m-1}(\hat{\alpha}_n - \alpha_0)^{\langle i\rangle} K_{in}(\alpha_0) + (\hat{\alpha}_n - \alpha_0)^{\langle m\rangle} K_{mn}(\hat{\alpha}_n), \quad (13)$$

for some $m \geq 2$, *where* $\{C_n(\alpha_0)\}$ *and* $\{K_{in}(\alpha_0)\}$ *are UB for* $i = 1, \ldots, m-1$, *and* $\{K_{mn}(\alpha)\}$ *is UB uniformly in a small neighborhood of* α_0. *Then, for* $\Delta_{1n} = C_n(\hat{\alpha}_n) - C_n(\alpha_0)$, (a) $\frac{1}{n} z_{1n}'\Delta_{1n} z_{2n} = o_p(1)$; (b) $\frac{1}{\sqrt{n}} z_{1n}'\Delta_{1n}A_n\epsilon_n = o_p(1)$; (c) $\frac{1}{n}\epsilon_n' B_{1n}'\Delta_{1n}B_{2n}\epsilon_n = o_p(1)$, *if (13) holds for* $m > 2$; *and* (d) $\frac{1}{\sqrt{n}}\epsilon_n'\Delta_{1n}\epsilon_n = o_p(1)$, *if (13) holds for* $m > 3$ *with* $\mathrm{tr}(K_{in}(\alpha_0)) = 0$ *for* $i = 1, \ldots, m-1$.

Furthermore, suppose another matrix $D_n(\hat{\gamma}_n)$ *has the expansion that*

$$D_n(\hat{\gamma}_n) - D_n(\gamma_0) = \sum_{i=1}^{m-1}(\hat{\gamma}_n - \gamma_0)^{\langle i\rangle} L_{in}(\gamma_0) + (\hat{\gamma}_n - \gamma_0)^{\langle m\rangle} L_{mn}(\hat{\gamma}_n), \quad (14)$$

for some $m \geq 2$, *where all the components on the right hand side have the same properties of corresponding ones in (13). Then, for* $\Delta_{2n} = (C_n(\hat{\alpha}_n) - C_n(\alpha_0))(D_n(\hat{\gamma}_n) - D_n(\gamma_0))$, (a′) $\frac{1}{n} z_{1n}'\Delta_{2n} z_{2n} = o_p(1)$; (b′) $\frac{1}{\sqrt{n}} z_{1n}'\Delta_{2n}A_n\epsilon_n = o_p(1)$; (c′) $\frac{1}{n}\epsilon_n' B_{1n}'\Delta_{2n}B_{2n}\epsilon_n = o_p(1)$, *if (13) and (14) hold for* $m > 2$; *and* (d′) $\frac{1}{\sqrt{n}}\epsilon_n'\Delta_{2n}\epsilon_n = o_p(1)$, *if (13) and (14) hold for* $m > 3$ *with* $\mathrm{tr}(K_{in}(\alpha_0)L_{jn}(\gamma_0)) = 0$ *for* $i, j = 1, \ldots, m-1$.

Proof. Let $T_n = \frac{1}{n} z_{1n}'(C_n(\hat{\alpha}_n) - C_n(\alpha_0))z_{2n}$. With (13), $T_n = T_{n1} + T_{n2}$, where $T_{n1} = \sum_{i=1}^{m-1}(\hat{\alpha}_n - \alpha_0)^{\langle i\rangle}\frac{1}{n}z_{1n}'K_{in}(\alpha_0)z_{2n}$ and $T_{n2} = (\hat{\alpha}_n - \alpha_0)^{\langle m\rangle}\frac{1}{n}z_{1n}'K_{mn}(\hat{\alpha}_n)z_{2n}$. $T_{n1} = o_p(1)$ because $\frac{1}{n}z_{1n}'K_{in}(\alpha_0) \times z_{2n} = O(1)$ by Lemma D.1, and $\hat{\alpha}_n - \alpha_0 = o_p(1)$. Similarly, as $\{K_{mn}(\alpha)\}$ is UB uniformly in a small neighborhood of α_0, and $\hat{\alpha}_n - \alpha_0 = o_p(1)$, it follows that $\{K_{mn}(\hat{\alpha}_n)\}$ is UB in probability. Hence $\frac{1}{n}z_{1n}'K_{mn}(\hat{\alpha}_n)z_{2n} = O_p(1)$ by Lemma D.1, which implies $T_{n2} = o_p(1)$. This proves (a).

Similarly, let $U_n = \frac{1}{\sqrt{n}}z_{1n}'(C_n(\hat{\alpha}_n) - C_n(\alpha_0))A_n\epsilon_n$. Then, with (13), $U_n = U_{n1} + U_{n2}$, where $U_{n1} = \sum_{i=1}^{m-1}(\hat{\alpha}_n - \alpha_0)^{\langle i\rangle}\frac{1}{\sqrt{n}}z_{1n}'K_{in}(\alpha_0)A_n\epsilon_n = o_p(1)$, because $\frac{1}{\sqrt{n}}z_{1n}'K_{in}(\alpha_0)A_n\epsilon_n = O_p(1)$ by Lemma D.4, and $U_{n2} = (\hat{\alpha}_n - \alpha_0)^{\langle m\rangle}\frac{1}{\sqrt{n}}z_{1n}'K_{mn}(\hat{\alpha}_n)A_n\epsilon_n$. Let $\|\cdot\|_1$ be the maximum column sum norm. Because the product of UBC matrices is UBC, $\|K_{mn}(\hat{\alpha}_n)A_n\|_1 \leq c_1$ for some constant c_1 for all n. As elements of z_{1n} are uniformly bounded, $\|z_{1n}'\|_1 \leq c_2$ for some constant c_2. It follows that

$$\begin{aligned}\|U_{n2}\|_1 &\leq n^{(1-m)/2}\|\sqrt{n}(\hat{\alpha}_n - \alpha_0)\|_1^m \cdot \|z_{1n}'\|_1 \times \|K_{mn}(\hat{\alpha}_n)A_n\|_1 \cdot \frac{1}{n}\|\epsilon_n\|_1 \\ &\leq c_1c_2 n^{(1-m)/2}\|\sqrt{n}(\hat{\alpha}_n - \alpha_0)\|_1^m \cdot \left(\frac{1}{n}\sum_{i=1}^n |\epsilon_{ni}|\right).\end{aligned}$$

Hence $U_{n2} = o_p(1)$ for $m \geq 2$ because $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_p(1)$ and $\frac{1}{n}\sum_{i=1}^n|\epsilon_{ni}| = O_p(1)$ by the strong law of large numbers. These prove (b).

For (c), let $R_n = \frac{1}{n}\epsilon_n'B_{1n}'(C_n(\hat{\alpha}_n) - C_n(\alpha_0))B_{2n}\epsilon_n$. With (13), $R_n = R_{n1} + R_{n2}$, where $R_{n1} = \sum_{i=1}^{m-1}(\hat{\alpha}_n - \alpha_0)^{\langle i\rangle}\frac{1}{n}\epsilon_n'B_{1n}'K_{in}(\alpha_0)B_{2n}\epsilon_n = o_p(1)$ because $\frac{1}{n}\epsilon_n'B_{1n}'K_{in}(\alpha_0)B_{2n}\epsilon_n = O_p(1)$ by Lemma D.3, and $R_{n2} = (\hat{\alpha}_n - \alpha_0)^{\langle m\rangle}\frac{1}{n}\epsilon_n'B_{1n}'K_{mn}(\hat{\alpha}_n)B_{2n}\epsilon_n$. On the other hand,

$$\begin{aligned}\|R_{n2}\|_1 &\leq n^{-m/2}\|\sqrt{n}(\hat{\alpha}_n - \alpha_0)\|_1^m \cdot \frac{1}{n}\|\epsilon_n\|_1 \times \|\epsilon_n\|_1 \cdot \|B_{1n}'K_{mn}(\hat{\alpha}_n)B_{2n}\|_1 \\ &\leq cn^{1-m/2}\|\sqrt{n}(\hat{\alpha}_n - \alpha_0)\|_1^m \cdot \left(\frac{1}{n}\sum_{i=1}^n|\epsilon_{ni}|\right)^2,\end{aligned}$$

for some constant c. Hence $R_{n2} = o_p(1)$ for $m > 2$ because $\frac{1}{n}\sum_{i=1}^n|\epsilon_{ni}|$ converges in probability to the absolute first moment of ϵ_{ni} and $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_p(1)$. These prove (c).

For (d), let $V_n = \frac{1}{\sqrt{n}}\epsilon_n'(C_n(\hat{\alpha}_n) - C_n(\alpha_0))\epsilon_n$. Then, $V_n = V_{n1} + V_{n2}$, where $V_{n1} = \sum_{i=1}^{m-1}(\hat{\alpha}_n - \alpha_0)^{\langle i\rangle}\frac{1}{\sqrt{n}}\epsilon_n'K_{in}(\alpha_0)\epsilon_n = o_p(1)$ because $\frac{1}{\sqrt{n}}\epsilon_n'K_{in}(\alpha_0)\epsilon_n = O_p(1)$ by Lemma D.5, and $V_{n2} = \frac{1}{\sqrt{n}}(\hat{\alpha}_n - \alpha_0)^{\langle m\rangle} \times \epsilon_n'K_{mn}(\hat{\alpha}_n)\epsilon_n$. The term $V_{n2} = o_p(1)$ for $m > 3$ because $\|V_{n2}\|_1 \leq$

$cn^{(3-m)/2}\|\sqrt{n}(\hat{\alpha}_n - \alpha_0)\|_1^m \cdot \left(\frac{1}{n}\sum_{i=1}^n |\epsilon_{ni}|\right)^2$. The desired results follow.

On the other hand, as

$$\left[\sum_{i=1}^{m-1}(\hat{\alpha}_n - \alpha_0)^{\langle i\rangle}K_{in}(\alpha_0) + (\hat{\alpha}_n - \alpha_0)^{\langle m\rangle}K_{mn}(\hat{\alpha}_n)\right] \times \left[\sum_{j=1}^{m-1}(\hat{\gamma}_n - \gamma_0)^{\langle j\rangle}L_{jn}(\gamma_0) + (\hat{\gamma}_n - \gamma_0)^{\langle m\rangle}L_{mn}(\hat{\gamma}_n)\right]$$
$$= \sum_{i=1}^{m-1}\sum_{j=1}^{m-1}(\hat{\alpha}_n - \alpha_0)^{\langle i\rangle}(\hat{\gamma}_n - \gamma_0)^{\langle j\rangle}K_{in}(\alpha_0)L_{jn}(\gamma_0) + \sum_{j=1}^{m-1}(\hat{\gamma}_n - \gamma_0)^{\langle j\rangle}(\hat{\alpha}_n - \alpha_0)^{\langle m\rangle}K_{mn}(\hat{\alpha}_n)L_{jn}(\gamma_0) + \sum_{i=1}^{m-1}(\hat{\alpha}_n - \alpha_0)^{\langle i\rangle}(\hat{\gamma}_n - \gamma_0)^{\langle m\rangle}K_{in}(\alpha_0)L_{mn}(\hat{\gamma}_n) + (\hat{\alpha}_n - \alpha_0)^{\langle m\rangle}(\hat{\gamma}_n - \gamma_0)^{\langle m\rangle}K_{mn}(\hat{\alpha}_n)L_{mn}(\hat{\gamma}_n),$$

(a′)–(d′) hold by the same argument as above applied to $K_{in}(\alpha_0)L_{jn}(\gamma_0)$, $(\hat{\alpha}_n - \alpha_0)^{\langle m\rangle}K_{mn}(\hat{\alpha}_n)L_{jn}(\gamma_0)$, $(\hat{\gamma}_n - \gamma_0)^{\langle m\rangle}K_{in}(\alpha_0)L_{mn}(\hat{\gamma}_n)$, and $(\hat{\gamma}_n - \gamma_0)^{\langle m\rangle}K_{mn}(\hat{\alpha}_n)L_{mn}(\hat{\gamma}_n)$. □

Lemma D.11. *Suppose that z_{1n} and z_{2n} are n-dimensional column vectors of constants which are uniformly bounded, the sequence of $n \times n$ constant matrices $\{A_n\}$ is UBC, $\{B_{1n}\}$ and $\{B_{2n}\}$ are UB, and $\epsilon_{n1}, \ldots, \epsilon_{nn}$ are i.i.d. with zero mean and finite fourth moment. $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$. Let C_n be either H_n, $\bar{G}_n$, $D(\bar{G}_n\bar{X}_n\beta_0)$, or $D(\bar{X}_{nj}^*)$ for $j = 1, \ldots, k^*$, and let $\hat{C}_n$ be its empirical counterpart. Then, under Assumption 3, for $\Delta_n = \hat{C}_n - C_n$, we have* (a) $\frac{1}{n}z_{1n}'\Delta_n^{(l)}z_{2n} = o_p(1)$, $\frac{1}{\sqrt{n}}z_{1n}'\Delta_n^{(l)}A_n\epsilon_n = o_p(1)$, $\frac{1}{n}\epsilon_n'B_{1n}'\Delta_n^{(l)}B_{2n}\epsilon_n = o_p(1)$, $\frac{1}{\sqrt{n}}\epsilon_n' \times \Delta_n^{(t)}\epsilon_n = o_p(1)$; (b) $\frac{1}{n}\mathrm{vec}_D'(\Delta_n^{(l)})z_{2n} = o_p(1)$, $\frac{1}{\sqrt{n}}\mathrm{vec}_D'(\Delta_n^{(l)})A_n\epsilon_n = o_p(1)$, $\frac{1}{n}\mathrm{tr}(A_n'\Delta_n^{(l)}) = o_p(1)$. *In addition, if $\{D_n(\gamma)\}$ is UB uniformly in a small neighborhood of γ_0 that is in the interior of its parameter space, then* (c) $\frac{1}{n}\mathrm{tr}[D_n'(\hat{\gamma}_n)\Delta_n^{(l)}] = o_p(1)$, *where $\hat{\gamma}_n - \gamma_0 = o_p(1)$.*

Proof. As $S_n - S_n(\hat{\lambda}_n) = (\hat{\lambda}_n - \lambda_0)W_n$, it follows that $S_n^{-1}(\hat{\lambda}_n) - S_n^{-1} = S_n^{-1}(\hat{\lambda}_n)[S_n - S_n(\hat{\lambda}_n)]S_n^{-1} = S_n^{-1}(\hat{\lambda}_n)(\hat{\lambda}_n - \lambda_0)G_n$. By induction, $S_n^{-1}(\hat{\lambda}_n) - S_n^{-1} = \sum_{i=1}^{m-1}(\hat{\lambda}_n - \lambda_0)^iS_n^{-1}G_n^i + (\hat{\lambda}_n - \lambda_0)^mS_n^{-1}(\hat{\lambda}_n)G_n^m$ for any $m \geq 2$. Hence, for $\hat{G}_n = G_n(\hat{\lambda}_n)$, it follows that

$$(\hat{G}_n - G_n)^{(l)} = \sum_{i=1}^{m-1}(\hat{\lambda}_n - \lambda_0)^i(G_n^{i+1})^{(l)} + (\hat{\lambda}_n - \lambda_0)^m(\hat{G}_nG_n^m)^{(l)}, \quad (15)$$

which conforms to the expansion (13) with $K_{in}(\lambda_0) = (G_n^{i+1})^{(l)}$ and $K_{mn}(\hat{\lambda}_n) = (\hat{G}_nG_n^m)^{(l)}$. Analogously, for $\hat{R}_n = R_n(\hat{\rho}_n)$, we have,

$$\hat{R}_n^{-1} - R_n^{-1} = \sum_{i=1}^{m-1}(\hat{\rho}_n - \rho_0)^iR_n^{-1}H_n^i + (\hat{\rho}_n - \rho_0)^m\hat{R}_n^{-1}H_n^m, \quad (16)$$

for any $m \geq 2$, which implies that

$$(\hat{H}_n - H_n)^{(l)} = \sum_{i=1}^{m-1}(\hat{\rho}_n - \rho_0)^i(H_n^{i+1})^{(l)} + (\hat{\rho}_n - \rho_0)^m(\hat{H}_nH_n^m)^{(l)}, \quad (17)$$

where $\hat{H}_n = H_n(\hat{\rho}_n)$. (17) conforms to the expansion (13) with $K_{in}(\lambda_0) = (H_n^{i+1})^{(l)}$ and $K_{mn}(\hat{\lambda}_n) = (\hat{H}_nH_n^m)^{(l)}$. Note that when the transformation $\cdot^{(t)}$ is taken, the deterministic parts of the expansion $K_{in}(\lambda_0) = (H_n^{i+1})^{(t)}$ have a zero trace by construction. Hence, when $C_n = H_n$, (a) follows from Lemma D.10, where the uniform boundedness in a neighborhood of the true parameters of the relevant matrices in the remainder terms follow from D.9.

As $\bar{G}_n = R_nG_nR_n^{-1}$, we have $\hat{R}_n\hat{G}_n\hat{R}_n^{-1} - R_nG_nR_n^{-1} = (\hat{R}_n - R_n)\hat{G}_n\hat{R}_n^{-1} + R_n(\hat{G}_n - G_n)(\hat{R}_n^{-1} - R_n^{-1}) + R_nG_n(\hat{R}_n^{-1} - R_n^{-1}) + R_n(\hat{G}_n - G_n)R_n^{-1}$, where $(\hat{R}_n - R_n)\hat{G}_n\hat{R}_n^{-1} = (\rho_0 - \hat{\rho}_n)M_n\hat{G}_n\hat{R}_n^{-1}$. $\hat{G}_n - G_n$ and $\hat{R}_n^{-1} - R_n^{-1}$ can be expanded to the form of (13) by (15) and (16). Hence, it follows by the same argument as above that (a) holds when $C_n = \bar{G}_n$.

As $\bar{G}_n\bar{X}_n\beta_0 = R_nG_nX_n\beta_0$, we have $D(\hat{R}_n\hat{G}_nX_n\hat{\beta}_n) - D(R_nG_nX_n\beta_0) = D[\hat{R}_n\hat{G}_nX_n(\hat{\beta}_n - \beta_0)] - (\hat{\rho}_n - \rho_0)D(M_n\hat{G}_nX_n\beta_0) + D[R_n(\hat{G}_n - G_n)X_n\beta_0]$. Let e_{kj} be the jth unit vector in R^k, then $\frac{1}{n}z_{1n}'D'[\hat{R}_n\hat{G}_nX_n(\hat{\beta}_n - \beta_0)]z_{2n} = \frac{1}{n}\sum_{i=1}^n z_{1n,i}z_{2n,i}e_{ni}'\hat{R}_n\hat{G}_nX_n(\hat{\beta}_n - \beta_0) = o_p(1)$, because $\frac{1}{n}\sum_{i=1}^n z_{1n,i}z_{2n,i}e_{ni}'\hat{R}_n\hat{G}_nX_n = O_p(1)$ and $\hat{\beta}_n - \beta_0 = o_p(1)$. On the other hand, $\frac{1}{n}z_{1n}'D[R_n(\hat{G}_n - G_n)X_n\beta_0]z_{2n} = o_p(1)$ by Lemma D.10. Hence, $\frac{1}{n}z_{1n}'[D(\hat{R}_n\hat{G}_nX_n\hat{\beta}_n) - D(R_nG_nX_n\beta_0)]z_{2n} = o_p(1)$. With similar arguments and corresponding results in Lemma D.10, the other results in (a) follow when $C_n = D(\bar{G}_n\bar{X}_n\beta_0)$.

As $\bar{X}_{nj}^* = R_nX_{nj}^*$, $D(\hat{R}_nX_{nj}^*) - D(R_nX_{nj}^*) = -(\hat{\rho}_n - \rho_0)D(M_nX_{nj}^*)$. Because $\sqrt{n}(\hat{\rho}_n - \rho_0) = O_p(1)$, the four claims in (a) hold for $C_n = D(\bar{X}_{nj}^*)$ by Lemmas D.1, D.4, D.3 and D.5 respectively.

For (b), as $\mathrm{vec}_D'(\Delta_n^{(l)}) = l_n'D(\Delta_n^{(l)})$, $\frac{1}{n}\mathrm{vec}_D'(\Delta_n^{(l)})z_{2n} = o_p(1)$ and $\frac{1}{\sqrt{n}}\mathrm{vec}_D'(\Delta_n^{(l)})A_n\epsilon_n = o_p(1)$ follow by similar arguments in the proof of (a) via Lemma D.10. To prove $\frac{1}{n}\mathrm{tr}(A_n'\Delta_n^{(l)}) = o_p(1)$, first consider the case when $C_n = H_n$. As in the proof of (a), for $m = 2$, $\hat{C}_n - C_n = (\hat{\alpha}_n - \alpha_0)K_{1n}(\alpha_0) + (\hat{\alpha}_n - \alpha_0)^2K_{2n}(\hat{\alpha}_n)$. Hence, $\frac{1}{n}\mathrm{tr}(A_n'\Delta_n^{(l)}) = (\hat{\alpha}_n - \alpha_0)\frac{1}{n}\mathrm{tr}[A_n'K_{1n}^{(l)}(\alpha_0)] + (\hat{\alpha}_n - \alpha_0)^2\frac{1}{n}\mathrm{tr}[A_n'K_{2n}^{(l)}(\hat{\alpha}_n)] = o_p(1)$, because $\frac{1}{n}\mathrm{tr}[A_n'K_{1n}^{(l)}(\alpha_0)] = O(1)$, $\frac{1}{n}\mathrm{tr}[A_n'K_{2n}^{(l)}(\hat{\alpha}_n)] = O_p(1)$, and $\hat{\alpha}_n - \alpha_0 = o_p(1)$. When $C_n = \bar{G}_n$, $\frac{1}{n}\mathrm{tr}(A_n'\Delta_n^{(l)}) = o_p(1)$ follows similar arguments. When $C_n = D(\bar{G}_n\bar{X}_n\beta_0)$, we have $\frac{1}{n}\mathrm{tr}(A_n'\Delta_n^{(l)}) = \frac{1}{n}\mathrm{vec}_D'(A_n)[\hat{R}_n\hat{G}_nX_n(\hat{\beta}_n - \beta_0) - (\hat{\rho}_n - \rho_0)M_n\hat{G}_nX_n\beta_0 + R_n(\hat{G}_n - G_n)X_n\beta_0] = o_p(1)$. When $C_n = D(\bar{X}_{nj}^*)$, we have $\frac{1}{n}\mathrm{tr}(A_n'\Delta_n^{(l)}) = -(\hat{\rho}_n - \rho_0)\frac{1}{n}\mathrm{tr}[A_n'D(M_nX_{nj}^*)] = o_p(1)$.

For (c), As $\{D_n(\gamma)\}$ is UB 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 UB in probability. The remaining arguments will be similar to those of the part 2 of (b). □

Lemma D.12. *Suppose that z_n is an n-dimensional column vector of constants which are uniformly bounded, and $\{A_n\}$ is UBC. $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$. Let T_n be either $\bar{X}_n$, $\bar{G}_n\bar{X}_n\beta_0$, $\mathrm{vec}_D(H_n^{(t)})$ or $\mathrm{vec}_D(\bar{G}_n^{(t)})$, with $\hat{T}_n$ being its estimated counterparts. Then, under Assumptions 1–3, we have* (a) $\frac{1}{n}(\hat{T}_n - T_n)'z_n = o_p(1)$, $\frac{1}{\sqrt{n}}(\hat{T}_n - T_n)'A_n\epsilon_n = o_p(1)$. *Furthermore, let $D_n(\hat{\gamma}_n)$ be a stochastic matrix that can be expanded to the form of (14) for some $m > 2$. Then,* (b) $\frac{1}{n}(\hat{T}_n - T_n)'D_n(\hat{\gamma}_n) = o_p(1)$.

Proof. (a) holds by Lemma D.11 (b).

For (b), we shall illustrate the proof for the case that $T_n = \bar{G}_n\bar{X}_n\beta_0$ as the others are similar. Let $\hat{D}_n = D_n(\hat{\gamma}_n)$. We have $\frac{1}{n}(\hat{T}_n - T_n)'\hat{D}_n = \frac{1}{n}[R_n(\hat{G}_n - G_n)X_n\beta_0]'\hat{D}_n + \frac{1}{n}[\hat{R}_n\hat{G}_nX_n(\hat{\beta}_n - \beta_0) - (\hat{\rho}_n - \rho_0)M_n\hat{G}_nX_n\beta_0]'\hat{D}_n$. First, $\frac{1}{n}[R_n(\hat{G}_n - G_n)X_n\beta_0]'\hat{D}_n = \frac{1}{n}[R_n(\hat{G}_n - G_n)X_n\beta_0]'(\hat{D}_n - D_n) + \frac{1}{n}[R_n(\hat{G}_n - G_n)X_n\beta_0]'D_n = o_p(1)$ by Lemma D.10. The remaining term is also $o_p(1)$ because $\hat{\rho}_n - \rho_0 = o_p(1)$, $\hat{\beta}_n - \beta_0 = o_p(1)$, and $\frac{1}{n}(M_n\hat{G}_nX_n\beta_0)'\hat{D}_n = O_p(1)$, $\frac{1}{n}(\hat{R}_n\hat{G}_nX_n)'\hat{D}_n = O_p(1)$. Hence, the desired result follows. □

To show the proposed moment conditions are optimal, we show adding additional moment conditions to the optimal moment

$$\Omega_{21} = E(g_n(\theta_0)g_n^{*\prime}(\theta_0)) = \begin{bmatrix} \sigma_0^2 Q_n' Q_n^* & \mu_3 Q_n' \mathrm{vec}_D(P_{1n}^*) & \cdots & \mu_3 Q_n' \mathrm{vec}_D(P_{k^*+5,n}^*) \\ \mu_3 \mathrm{vec}_D'(P_{1n}) Q_n^* & \sigma_0^4 \mathrm{tr}(P_{1n}^{(s)} P_{1n}^*) & \cdots & \sigma_0^4 \mathrm{tr}(P_{1n}^{(s)} P_{k^*+5,n}^*) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_3 \mathrm{vec}_D'(P_{mn}) Q_n^* & \sigma_0^4 \mathrm{tr}(P_{mn}^{(s)} P_{1n}^*) & \cdots & \sigma_0^4 \mathrm{tr}(P_{mn}^{(s)} P_{k^*+5,n}^*) \end{bmatrix}$$
$$+ (\mu_4 - 3\sigma_0^4) \begin{bmatrix} 0_{q\times(k^*+4)} & 0_{q\times(k^*+5)} \\ 0_{1\times(k^*+4)} & \mathrm{vec}_D'(P_{1n})(\mathrm{vec}_D(P_{1n}^*), \ldots, \mathrm{vec}_D(P_{k^*+5,n}^*)) \\ \vdots & \vdots \\ 0_{1\times(k^*+4)} & \mathrm{vec}_D'(P_{mn})(\mathrm{vec}_D(P_{1n}^*), \ldots, \mathrm{vec}_D(P_{k^*+5,n}^*)) \end{bmatrix} \quad (18)$$

Box V.

$$\Delta_P' = \begin{bmatrix} I_n & -\frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2} I_n & -\frac{\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2} I_n & 0 & 0 & 0 \\ 0 & 0 & 0 & I_n & -\frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2} I_n & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{nk^*} \end{bmatrix}$$

Box VI.

$$\Delta_{Q1}' = \begin{bmatrix} \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2} I_{k^*} & 0 & -\frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}\left(\frac{1}{n} l_n' \bar{X}_n^*\right)' & 0 & 0 \\ 0 & \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2} & -\frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}\left(\frac{1}{n} l_n' \bar{G}_n \bar{X}_n \beta_0\right) & -\frac{2\sigma_0\eta_3}{(\eta_4-1)-\eta_3^2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Box VII.

conditions does not increase the asymptotic efficiency of the GMME using the conditions for redundancy in Breusch et al. (1999). The definition of redundancy is given as follows. "Let $\hat{\theta}$ be the optimal GMME 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 GMME based on the whole set of moment conditions $E[g(y,\theta)] \equiv E\left[g_1'(y,\theta), g_2'(y,\theta)\right]' = 0$. 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). For moment conditions $E[g(y,\theta)] \equiv E\left[g_1'(y,\theta), g_2'(y,\theta)\right]' = 0$, let $\Omega \equiv E\left[g(y,\theta)g'(y,\theta)\right] = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$, with $\Omega_{jl} = E\left[g_j(y,\theta)g_l'(y,\theta)\right]$ for $j, l = 1, 2$. And define $D_j = E\left[\partial g_j(y,\theta)/\partial\theta'\right]$ for $j = 1, 2$. Let the dimensions of $g_1(y,\theta)$, $g_2(y,\theta)$ and θ be k_1, k_2 and p.

Lemma D.13. *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* $D_2 = \Omega_{21}A$.

Lemma D.14. *Let the set of moment conditions to be considered be* $E[g(\theta)] \equiv E\left[g_1'(\theta), g_2'(\theta), g_3'(\theta)\right]' = 0$*, 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 *are redundant given* g_1.

Appendix E. Proofs

Proof of Proposition 1. Consider the moment conditions $E[g_n^{*\prime}(\theta_0), g_n'(\theta_0)]' = 0$, where $g_n(\theta)$ is a vector of arbitrary moment functions taken the form of (2). 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_{jn} $(j = 1, \ldots, m)$ and Q_n st. $D_2 = \Omega_{21}A_n$ according to Lemma D.13(c), where

$$D_2 = E\left(\frac{\partial}{\partial\theta'} g_n(\theta_0)\right) = -\begin{bmatrix} 0_{q\times 1} & Q_n'\bar{G}_n\bar{X}_n\beta_0 & Q_n'\bar{X}_n \\ \sigma_0^2 \mathrm{tr}(P_{1n}^{(s)}H_n) & \sigma_0^2 \mathrm{tr}(P_{1n}^{(s)}\bar{G}_n) & 0_{1\times k} \\ \vdots & \vdots & \vdots \\ \sigma_0^2 \mathrm{tr}(P_{mn}^{(s)}H_n) & \sigma_0^2 \mathrm{tr}(P_{mn}^{(s)}\bar{G}_n) & 0_{1\times k} \end{bmatrix},$$

Ω_{21} has an explicit form as in Eq. (18) given in Box V.

In the case that $\bar{X}_n$ does not have a column proportional to l_n so that $\bar{X}_n^* = \bar{X}_n$, let $P_{\lambda n}^* = P_{1n}^* - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2} P_{2n}^* - \frac{\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2} P_{3n}^*$, $P_{\rho n}^* = P_{4n}^* - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2} P_{5n}^*$, $P_{\beta nl}^* = P_{l+5,n}^*$ for $l = 1, \ldots, k^*$, $Q_{\beta n}^* = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2} Q_{1n}^* - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2} Q_{3n}^* \left(\frac{1}{n} l_n' \bar{X}_n^*\right)$, $Q_{\lambda n}^* = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2} Q_{2n}^* - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2} Q_{3n}^* \left(\frac{1}{n} l_n' \bar{G}_n \bar{X}_n \beta_0\right) - \frac{2\sigma_0\eta_3}{(\eta_4-1)-\eta_3^2} Q_{4n}^*$ and $Q_{\rho n}^* = Q_{5n}^*$. Note that $(P_{\lambda n}^*, P_{\rho n}^*, P_{\beta n1}^*, \ldots, P_{\beta nk^*}^*) = (P_{1n}^*, \ldots, P_{k^*+5,n}^*)\Delta_P$, where Δ_P' is as given in Box VI. and $(Q_{\beta n}^*, Q_{\lambda n}^*, Q_{\rho n}^*) = (Q_{1n}^*, \ldots, Q_{5n}^*)\Delta_{Q1}$, where Δ_{Q1}' is as given in Box VII. On the other hand, in the case that $\bar{X}_n$'s last column is given by $c(\rho_0)l_n$, let $Q_{\beta n}^* = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2} Q_{1n}^*(I_{k^*}, 0_{k^*\times 1}) + \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2} c(\rho_0) Q_{3n}^* e_{kk}' - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2} Q_{3n}^* \left(\frac{1}{n} l_n' \bar{X}_n\right)$, where e_{kj} is the jth unit vector in R^k, so that $(Q_{\beta n}^*, Q_{\lambda n}^*, Q_{\rho n}^*) = (Q_{1n}^*, \ldots, Q_{5n}^*)\Delta_{Q1}$,

$$\Delta'_{Q2} = \begin{bmatrix} \frac{\eta_4 - 1}{(\eta_4 - 1) - \eta_3^2}(I_{k^*}, 0_{k^* \times 1}) & 0 & \frac{\eta_4 - 1}{(\eta_4 - 1) - \eta_3^2} c(\rho_0) e_{kk} - \frac{\eta_3^2}{(\eta_4 - 1) - \eta_3^2}\left(\frac{1}{n} l'_n \bar{X}_n\right)' & 0 & 0 \\ 0 & \frac{\eta_4 - 1}{(\eta_4 - 1) - \eta_3^2} & -\frac{\eta_3^2}{(\eta_4 - 1) - \eta_3^2}\left(\frac{1}{n} l'_n \bar{G}_n \bar{X}_n \beta_0\right) & -\frac{2\sigma_0 \eta_3}{(\eta_4 - 1) - \eta_3^2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Box VIII.

where Δ'_{Q2} is as given in Box VIII. Let

$$B'_n = -\begin{bmatrix} 0 & 0 & -\frac{2\sigma_0^{-1}\eta_3}{(\eta_4 - 1) - \eta_3^2} & 0 & \sigma_0^{-2} & 0 \\ 0 & \sigma_0^{-2} & 0 & \sigma_0^{-2} & 0 & 0 \\ \sigma_0^{-2} I_k & 0 & 0 & 0 & 0 & b' \end{bmatrix},$$

where $b = (b'_1, \ldots, b'_{k^*})'$ with $b_l = -\frac{\sigma_0^{-3}\eta_3}{(\eta_4-1)-\eta_3^2} e'_{kl}$ for $l = 1, \ldots, k^*$. Let $A_n = \begin{bmatrix} \Delta_{Q1} & 0 \\ 0 & \Delta_p \end{bmatrix} B_n$ when $\bar{X}_n$ does not have a column proportional to l_n, and $A_n = \begin{bmatrix} \Delta_{Q2} & 0 \\ 0 & \Delta_p \end{bmatrix} B_n$when $\bar{X}_n$'s last column is $c(\rho_0) l_n$. Let $J_n = I_n - \frac{1}{n} l_n l'_n$. To check $D_2 = \Omega_{21} A_n$, the following identities are helpful. For $l = 1, \ldots, k^*$, (a) $\mathrm{vec}_D(P^*_{\lambda n}) = \frac{2}{(\eta_4-1)-\eta_3^2} \mathrm{vec}_D(\bar{G}_n^{(t)}) - \frac{\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2} J_n \bar{G}_n \bar{X}_n \beta_0$; (b) $\mathrm{vec}_D(P^*_{\rho n}) = \frac{2}{(\eta_4-1)-\eta_3^2} \times \mathrm{vec}_D(H_n^{(t)})$; (c) $\mathrm{vec}_D(P^*_{\beta nl}) = J_n \bar{X}^*_{nl}$; and (d) $\sum_{l=1}^{k^*} \mathrm{vec}_D(P^*_{\beta nl}) e'_{kl} = J_n \bar{X}_n$.

It follows from (a), (b) and (d), respectively, to have that (e) $\sigma_0^2 Q^*_{\lambda n} + \mu_3 \mathrm{vec}_D(P^*_{\lambda n}) = \sigma_0^2 \bar{G}_n \bar{X}_n \beta_0$; (f) $\frac{2}{(\eta_4-1)-\eta_3^2} Q^*_{\rho n} = \mathrm{vec}_D(P^*_{\rho n})$, and (g) $Q^*_{\beta n} - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2} \sum_{l=1}^{k^*} \mathrm{vec}_D(P^*_{\beta nl}) e'_{kl} = \bar{X}_n$.

For an arbitrary $n \times n$ matrix P_n with $\mathrm{tr}(P_n) = 0$, we have: (h) $\mathrm{vec}'_D(P_n) Q^*_{\beta n} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2} \mathrm{vec}'_D(P_n) \bar{X}_n$; (i) $\mu_3 \mathrm{vec}'_D(P_n) Q^*_{\lambda n} + \sigma_0^4 \mathrm{tr}(P_n^{(s)} P^*_{\lambda n}) + (\mu_4 - 3\sigma_0^4) \mathrm{vec}'_D(P_n) \mathrm{vec}_D(P^*_{\lambda n}) = \sigma_0^4 \mathrm{tr}(P_n^{(s)} \bar{G}_n)$; (j) $-\frac{2\sigma_0^{-1}\eta_3^2}{(\eta_4-1)-\eta_3^2} \mathrm{vec}'_D(P_n) Q^*_{\rho n} + \sigma_0^4 \mathrm{tr}(P_n^{(s)} P^*_{\rho n}) + (\mu_4 - 3\sigma_0^4) \mathrm{vec}'_D(P_n) \mathrm{vec}_D(P^*_{\rho n}) = \sigma_0^4 \mathrm{tr}(P_n^{(s)} H_n)$; and (k) $\sigma_0^4 \mathrm{tr}(P_n^{(s)} P^*_{\beta nl}) + (\mu_4 - 3\sigma_0^4) \mathrm{vec}'_D(P_n) \mathrm{vec}_D(P^*_{\beta nl}) = (\mu_4 - \sigma_0^4) \mathrm{vec}'_D(P_n) \mathrm{vec}_D(P^*_{\beta nl})$.

It follows from identity (f) the (1, 1) block of $\Omega_{21} A_n$ is 0; from (e) that the (1, 2) block of $\Omega_{21} A_n$ is $-Q'_n \bar{G}_n \bar{X}_n \beta_0$; and from (g) that the (1, 3) block of $\Omega_{21} A_n$ is $-Q'_n \bar{X}_n$. Identity (j) implies that the $(j+1, 1)$ blocks of $\Omega_{21} A_n$ are $-\sigma_0^2 \mathrm{tr}(P_{jn}^{(s)} H_n)$ for $j = 1, \ldots, m$; (i) implies that the $(j+1, 2)$ blocks of $\Omega_{21} A_n$ are $-\sigma_0^2 \mathrm{tr}(P_{jn}^{(s)} \bar{G}_n)$ for $j = 1, \ldots, m$; and (d), (h) and (k) imply that the remaining blocks of $\Omega_{21} A_n$ are zeros. Therefore, $\Omega_{21} A_n = D_2$ and the desired result follows.

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, $D_1 = \Omega_{11} A_n$, and hence $\Omega_{11}^{-1} D_1 = A_n$, where $\Omega_{11} = \mathrm{var}(g^*_n(\theta_0))$ and $D_1 = E\left(\frac{\partial}{\partial\theta} g^*_n(\theta_0)\right)$. Hence $\Sigma_B = \lim_{n\to\infty} \frac{1}{n} D'_1 \Omega_{11}^{-1} D_1 = \lim_{n\to\infty} \frac{1}{n} D'_1 A_n$. Some tedious but straightforward algebra gives the explicit form of Σ_B in (3). □

Proof of Proposition 2. We shall show that the objective functions $F^*_n(\theta) = \hat{g}^{*\prime}_n(\theta) \hat{\Omega}_n^{*-1} \hat{g}^*_n(\theta)$ and $F_n(\theta) = g^{*\prime}_n(\theta) \Omega_n^{*-1} g^*_n(\theta)$ will satisfy the conditions in Lemma D.7. If so, the GMME from the minimization of $F^*_n(\theta)$ will have the same limiting distribution as that of the minimization of $F_n(\theta)$. The difference of $F^*_n(\theta)$ and $F_n(\theta)$ 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 $m^* = k^* + 5$. Explicitly,

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

The $\epsilon_n(\theta)$ is related to ϵ_n as $\epsilon_n(\theta) = (I_n + (\rho_0 - \rho)H_n)(I_n + (\lambda_0 - \lambda)\bar{G}_n)\epsilon_n + d_n(\theta)$, where $d_n(\theta) = (I_n + (\rho_0 - \rho)H_n)[(\lambda_0 - \lambda)\bar{G}_n \bar{X}_n \beta_0 + \bar{X}_n(\beta_0 - \beta)]$. It follows that $\frac{1}{n}(\hat{Q}^*_n - Q^*_n)'\epsilon_n(\theta) = \frac{1}{n}(\hat{Q}^*_n - Q^*_n)'(I_n + (\rho_0 - \rho)H_n)(I_n + (\lambda_0 - \lambda)\bar{G}_n)\epsilon_n + \frac{1}{n}(\hat{Q}^*_n - Q^*_n)' d_n(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$ by Lemma D.12. Similarly, it follows by Lemma D.11 that $\frac{1}{n}\epsilon'_n(\theta)(\hat{P}^*_{jn} - P^*_{jn})\epsilon_n(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$ for $j = 1, \ldots, m^*$. Hence, $\frac{1}{n}(\hat{g}^*_n(\theta) - g^*_n(\theta)) = o_p(1)$ uniformly in $\theta \in \Theta$.

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

$$\frac{\partial g^*_n(\theta)}{\partial\theta'} = \begin{bmatrix} Q^{*\prime}_n \frac{\partial\epsilon_n(\theta)}{\partial\theta'} \\ \epsilon'_n(\theta) P^{*(s)}_{1n} \frac{\partial\epsilon_n(\theta)}{\partial\theta'} \\ \vdots \\ \epsilon'_n(\theta) P^{*(s)}_{m^*n} \frac{\partial\epsilon_n(\theta)}{\partial\theta'} \end{bmatrix}, \quad \text{and}$$

$$\frac{\partial^2 g^*_n(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} Q^{*\prime}_n \frac{\partial^2\epsilon_n(\theta)}{\partial\theta\partial\theta'} \\ \frac{\partial\epsilon'_n(\theta)}{\partial\theta} P^{*(s)}_{1n} \frac{\partial\epsilon_n(\theta)}{\partial\theta'} + \epsilon'_n(\theta) P^{*(s)}_{1n} \frac{\partial^2\epsilon_n(\theta)}{\partial\theta\partial\theta'} \\ \vdots \\ \frac{\partial\epsilon'_n(\theta)}{\partial\theta} P^{*(s)}_{m^*n} \frac{\partial\epsilon_n(\theta)}{\partial\theta'} + \epsilon'_n(\theta) P^{*(s)}_{m^*n} \frac{\partial^2\epsilon_n(\theta)}{\partial\theta\partial\theta'} \end{bmatrix}.$$

$\frac{\partial\epsilon_n(\theta)}{\partial\theta'} = -[M_n(I_n - \lambda W_n)Y_n - M_n X_n \beta, R_n(\rho) W_n Y_n, R_n(\rho) X_n]$, where $Y_n = S_n^{-1} X_n \beta_0 + S_n^{-1} R_n^{-1} \epsilon_n$. $\frac{\partial^2\epsilon_n(\theta)}{\partial\rho\partial\theta'} = [0, M_n W_n Y_n, M_n X_n]$, $\frac{\partial^2\epsilon_n(\theta)}{\partial\lambda\partial\theta'} = [M_n W_n Y_n, 0, 0]$, and $\frac{\partial^2\epsilon_n(\theta)}{\partial\beta\partial\theta'} = [M_n X_n, 0, 0]$. It follows from Lemmas D.11 and D.12 that $\frac{1}{n}\left(\frac{\partial\hat{g}^*_n(\theta)}{\partial\theta} - \frac{\partial g^*_n(\theta)}{\partial\theta}\right) = o_p(1)$ and $\frac{1}{n}\left(\frac{\partial^2\hat{g}^*_n(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 g^*_n(\theta)}{\partial\theta\partial\theta'}\right) = o_p(1)$ uniformly in $\theta \in \Theta$.

Consider $\frac{1}{n}(\hat{\Omega}^*_n - \Omega^*_n)$, where

$$\Omega^*_n = E\left[g^*_n(\theta_0) g^{*\prime}_n(\theta_0)\right] = \begin{bmatrix} \sigma_0^2 Q^{*\prime}_n Q^*_n & \mu_3 Q^{*\prime}_n \omega^*_{m^*n} \\ \mu_3 \omega^{*\prime}_{m^*n} Q^*_n & \sigma_0^4 \Delta^*_{m^*n} + (\mu_4 - 3\sigma_0^4)\omega^{*\prime}_{m^*n}\omega^*_{m^*n} \end{bmatrix},$$

with $\omega^*_{m^*n} = [\mathrm{vec}_D(P^*_{1n}), \ldots, \mathrm{vec}_D(P^*_{k^*+2,n})]$ and $\Delta^*_{m^*} = \begin{bmatrix} \mathrm{tr}(P^{*(s)}_{1n} P^*_{1n}) & \cdots & \mathrm{tr}(P^{*(s)}_{1n} P^*_{m^*n}) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(P^{*(s)}_{m^*n} P^*_{1n}) & \cdots & \mathrm{tr}(P^{*(s)}_{m^*n} P^*_{m^*n}) \end{bmatrix}$. First, consider the block matrix

$\sigma_0^4\Delta^*_{m^*n} + (\mu_4 - 3\sigma_0^4)\omega^{*\prime}_{m^*n}\omega^*_{m^*n}$. As $\{\hat{P}^*_{jn}\}$ is UBC in probability, it follows from Lemma D.11 that $\frac{1}{n}\mathrm{tr}(\hat{P}^{*(s)}_{in}\hat{P}^*_{jn}) - \frac{1}{n}\mathrm{tr}(P^{*(s)}_{in}P^*_{jn}) = \frac{1}{n}\mathrm{tr}[(\hat{P}^{*(s)}_{in} - P^{*(s)}_{in})\hat{P}^*_{jn} + P^{*(s)}_{in}(\hat{P}^*_{jn} - P^*_{jn})] = o_p(1)$, and $\frac{1}{n}\mathrm{vec}'_D(\hat{P}^*_{in})\mathrm{vec}_D(\hat{P}^*_{jn}) - \frac{1}{n}\mathrm{vec}'_D(P^*_{in})\mathrm{vec}_D(P^*_{jn}) = \frac{1}{n}\mathrm{vec}'_D(\hat{P}^*_{in})\mathrm{vec}_D(\hat{P}^*_{jn} - P^*_{jn}) + \frac{1}{n}\mathrm{vec}'_D(\hat{P}^*_{in} - P^*_{in})\mathrm{vec}_D(P^*_{jn}) = o_p(1)$ for $i, j = 1, \ldots, m^*$. Hence, $\frac{1}{n}(\hat{\sigma}^2_n)^2\mathrm{tr}(\hat{P}^{*(s)}_{in}\hat{P}^*_{jn}) - \frac{1}{n}\sigma_0^4\mathrm{tr}(P^{*(s)}_{in}P^*_{jn}) = o_p(1)$ and $\frac{1}{n}(\hat{\mu}_4 - 3(\hat{\sigma}^2_n)^2)\mathrm{vec}'_D(\hat{P}^*_{in})\mathrm{vec}_D(\hat{P}^*_{jn}) - \frac{1}{n}(\mu_4 - 3\sigma_0^4)\mathrm{vec}'_D(P^*_{in})\mathrm{vec}_D(P^*_{jn}) = o_p(1)$ for $i, j = 1, \ldots, m^*$, as $\hat{\sigma}^2_n - \sigma^2_0 = o_p(1)$ and $\hat{\mu}_4 - \mu_4 = o_p(1)$.

Next consider $\mu_3 Q^{*\prime}_n\omega^*_{m^*n}$. As elements of $\hat{Q}^*_n$ are uniformly bounded in probability for all n, it follows from Lemmas D.11 and D.12 that $\frac{1}{n}\hat{Q}^{*\prime}_n\mathrm{vec}_D(\hat{P}^*_{jn}) - \frac{1}{n}Q^{*\prime}_n\mathrm{vec}_D(P^*_{jn}) = \frac{1}{n}\hat{Q}^{*\prime}_n\mathrm{vec}_D(\hat{P}^*_{jn} - P^*_{jn}) + \frac{1}{n}(\hat{Q}^*_n - Q^*_n)'\mathrm{vec}_D(P^*_{jn}) = o_p(1)$ for $j = 1, \ldots, m^*$. Hence, $\frac{1}{n}\hat{\mu}_3\hat{Q}^{*\prime}_n\mathrm{vec}_D(\hat{P}^*_{jn}) - \frac{1}{n}\mu_3 Q^{*\prime}_n\mathrm{vec}_D(P^*_{jn}) = o_p(1)$ for $j = 1, \ldots, m^*$, as $\hat{\mu}_3 - \mu_3 = o_p(1)$.

Lastly, consider $\sigma_0^2 Q^{*\prime}_n Q^*_n$. As elements of $\hat{Q}^*_n$ are uniformly bounded in probability for all n, Lemma D.12 implies that $\frac{1}{n}(\hat{Q}^{*\prime}_{in}\hat{Q}^*_{jn} - Q^{*\prime}_{in}Q^*_{jn}) = \frac{1}{n}[\hat{Q}^{*\prime}_{in}(\hat{Q}^*_{jn} - Q^*_{jn}) + (\hat{Q}^*_{in} - Q^*_{in})'Q^*_{jn}] = o_p(1)$ for $i, j = 1, \ldots, 5$. Therefore, $\frac{1}{n}(\hat{\sigma}^2_n\hat{Q}^{*\prime}_n\hat{Q}^*_n - \sigma_0^2 Q^{*\prime}_n Q^*_n) = \hat{\sigma}^2_n\frac{1}{n}(\hat{Q}^{*\prime}_n\hat{Q}^*_n - Q^{*\prime}_n Q^*_n) + (\hat{\sigma}^2_n - \sigma_0^2)\frac{1}{n}Q^{*\prime}_n Q^*_n = o_p(1)$. 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, $\left(\frac{1}{n}\hat{\Omega}^*_n\right)^{-1} - \left(\frac{1}{n}\Omega^*_n\right)^{-1} = o_p(1)$ by the continuous mapping theorem.

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 $\sup_{\theta\in\Theta}\frac{1}{n}|E(g^*_n(\theta))| = O(1)$ (Lee, 2007, p. 21), $\frac{1}{n}g^*_n(\theta)$ 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 $\mathcal{F}^*_n(\theta)$ and $\mathcal{F}_n(\theta)$ can be investigated. By expansion, $\frac{1}{n}(\mathcal{F}^*_n(\theta) - \mathcal{F}_n(\theta)) = \frac{1}{n}\hat{g}^{*\prime}_n(\theta)\hat{\Omega}^{*-1}_n(\hat{g}^*_n(\theta) - g^*_n(\theta)) + \frac{1}{n}\hat{g}^{*\prime}_n(\theta)(\hat{\Omega}^{*-1}_n - \Omega^{*-1}_n)\hat{g}^*_n(\theta) + \frac{1}{n}g^{*\prime}_n(\theta)\Omega^{*-1}_n(\hat{g}^*_n(\theta) - g^*_n(\theta)) = o_p(1)$, uniformly in $\theta \in \Theta$. Similarly, for each component θ_l of θ, $\frac{1}{n}\frac{\partial^2\mathcal{F}^*_n(\theta)}{\partial\theta_l\partial\theta'} - \frac{1}{n}\frac{\partial^2\mathcal{F}_n(\theta)}{\partial\theta_l\partial\theta'} = \frac{2}{n}\left[\frac{\partial\hat{g}^{*\prime}_n(\theta)}{\partial\theta_l}\hat{\Omega}^{*-1}_n\frac{\partial\hat{g}^*_n(\theta)}{\partial\theta'} + \hat{g}^{*\prime}_n(\theta)\hat{\Omega}^{*-1}_n\frac{\partial^2\hat{g}^*_n(\theta)}{\partial\theta_l\partial\theta'} - \left(\frac{\partial g^{*\prime}_n(\theta)}{\partial\theta_l}\Omega^{*-1}_n\frac{\partial g^*_n(\theta)}{\partial\theta'} + g^{*\prime}_n(\theta)\Omega^{*-1}_n\frac{\partial^2 g^*_n(\theta)}{\partial\theta_l\partial\theta'}\right)\right] = o_p(1)$.

Finally, because $\left(\frac{\partial\hat{g}^{*\prime}_n(\theta_0)}{\partial\theta}\hat{\Omega}^{*-1}_n - \frac{\partial g^{*\prime}_n(\theta_0)}{\partial\theta}\Omega^{*-1}_n\right) = o_p(1)$ as above, and $\frac{1}{\sqrt{n}}g^*_n(\theta_0) = O_p(1)$ by the central limit theorems in Lemmas D.4 and D.5,

$$\frac{1}{\sqrt{n}}\left(\frac{\partial\mathcal{F}^*_n(\theta_0)}{\partial\theta} - \frac{\partial\mathcal{F}_n(\theta_0)}{\partial\theta}\right) = 2\left\{\frac{\partial\hat{g}^{*\prime}_n(\theta_0)}{\partial\theta}\hat{\Omega}^{*-1}_n\frac{1}{\sqrt{n}}(\hat{g}^*_n(\theta_0) - g^*_n(\theta_0)) + \left(\frac{\partial\hat{g}^{*\prime}_n(\theta_0)}{\partial\theta}\hat{\Omega}^{*-1}_n - \frac{\partial g^{*\prime}_n(\theta_0)}{\partial\theta}\Omega^{*-1}_n\right)\frac{1}{\sqrt{n}}g^*_n(\theta_0)\right\}$$
$$= 2\frac{\partial\hat{g}^{*\prime}_n(\theta_0)}{\partial\theta}\hat{\Omega}^{*-1}_n\frac{1}{\sqrt{n}}(\hat{g}^*_n(\theta_0) - g^*_n(\theta_0)) + o_p(1).$$

As $\frac{1}{\sqrt{n}}(\hat{g}^*_n(\theta_0) - g^*_n(\theta_0)) = o_p(1)$ by Lemmas D.11 and D.12, $\frac{1}{\sqrt{n}}\left(\frac{\partial\mathcal{F}^*_n(\theta_0)}{\partial\theta} - \frac{\partial\mathcal{F}_n(\theta_0)}{\partial\theta}\right) = o_p(1)$. The desired result follows from Lemma D.7. □

Proof of Corollary 3. Let $P^\dagger_{\rho n} = P^\dagger_{1n} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2}P^\dagger_{2n}$, and $P^\dagger_{\beta nj} = P^\dagger_{j+2,n}$ for $j = 1, \ldots, k^*$. Let $Q^\dagger_{\beta n} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}\bar{X}_n - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2} \times Q^\dagger_{2n}\left(\frac{1}{n}l'_n\bar{X}_n\right)$ and $Q^\dagger_{\rho n} = Q^\dagger_{3n}$. Note that $P^\dagger_{\rho n}$ and $Q^\dagger_{\beta n}$ are linear combinations of $P^\dagger_{1n}$, $P^\dagger_{2n}$ and $Q^\dagger_n$. Hence, it is sufficient to show that the optimal GMME with $\bar{g}^\dagger_{\rho n}(\rho, \beta) = (Q^\dagger_{\beta n}, Q^\dagger_{\rho n}, P^\dagger_{\rho n}\epsilon_{\rho n}(\rho, \beta), P^\dagger_{\beta n1}\epsilon_{\rho n}(\rho, \beta), \ldots, P^\dagger_{\beta nk^*}\epsilon_{\rho n}(\rho, \beta))'\epsilon_{\rho n}(\rho, \beta)$ is the most efficient within $\mathcal{M}_{\rho n}$. Similar to the proof of Proposition 1, it is sufficient to show that there exists an A_n invariant with P_{jn} ($j = 1, \ldots, m$) and Q_n st. $D_2 = \Omega_{21}A_n$, where

$$D_2 = \left[E\left(\frac{\partial}{\partial\rho}g_{\rho n}\right), E\left(\frac{\partial}{\partial\beta'}g_{\rho n}\right)\right]\Bigg|_{\rho_0,\beta_0} = -\begin{bmatrix} 0 & \sigma_0^2\mathrm{tr}(P^{(s)}_{1n}H_n) & \cdots & \sigma_0^2\mathrm{tr}(P^{(s)}_{mn}H_n) \\ \bar{X}'_n Q_n & 0 & \cdots & 0 \end{bmatrix}',$$

and $\Omega_{21} = E(g_{\rho n}\bar{g}^{\dagger\prime}_{\rho n})|_{\rho_0,\beta_0}$ in the form of (18). Let

$$A_n = -\begin{bmatrix} 0 & \sigma_0^{-2} & -\dfrac{2\sigma_0^{-1}\eta_3}{(\eta_4 - 1) - \eta_3^2} & 0 \\ \sigma_0^{-2}I_k & 0 & 0 & b' \end{bmatrix}',$$

where $b = (b'_1, \ldots, b'_{k^*})'$ with $b_l = -\frac{\sigma_0^{-3}\eta_3}{(\eta_4-1)-\eta_3^2}e'_{kl}$ for $l = 1, \ldots, k^*$. With some simplified identities of those in the proof of Proposition 1, we have $\Omega_{21}A_n = D_2$.

Furthermore, as $\bar{g}^\dagger_{\rho n}$ is a special case of $g_{\rho n}$, $\Omega_{11}^{-1}D_1 = A_n$, where $\Omega_{11} = \mathrm{var}(\bar{g}^\dagger_{\rho n})$ and $D_1 = \left[E\left(\frac{\partial}{\partial\rho}\bar{g}^\dagger_{\rho n}\right), E\left(\frac{\partial}{\partial\beta'}\bar{g}^\dagger_{\rho n}\right)\right]\Big|_{\rho_0,\beta_0}$. Hence, the desired result follows by $\Sigma_{B\rho} = \lim_{n\to\infty}\frac{1}{n}D'_1A_n$. □

Proof of Corollary 4. Let $P^\star_{\lambda n} = P^\star_{1n} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2}P^\star_{2n} - \frac{\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2}P^\star_{3n}$, and $P^\star_{\beta nj} = P^\star_{j+3,n}$ for $j = 1, \ldots, k^*$. Let $Q^\star_{\beta n} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}X_n - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}Q^\star_{3n}\left(\frac{1}{n}l'_nX_n\right)$ and $Q^\star_{\lambda n} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}Q^\star_{2n} - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}Q^\star_{3n}\left(\frac{1}{n}l'_nG_nX_n\beta_0\right) - \frac{2\sigma_0\eta_3}{(\eta_4-1)-\eta_3^2}Q^\star_{4n}$. It is sufficient to show that the optimal GMME with

$$\bar{g}^\star_{\lambda n}(\rho, \beta) = (Q^\star_{\beta n}, Q^\star_{\lambda n}, P^\star_{\lambda n}\epsilon_{\lambda n}(\rho, \beta), P^\star_{\beta n1}\epsilon_{\lambda n}(\rho, \beta), \ldots, P^\star_{\beta nk^*}\epsilon_{\lambda n}(\rho, \beta))'\epsilon_{\lambda n}(\rho, \beta)$$

is the most efficient within $\mathcal{M}_{\lambda n}$. For

$$D_2 = \left[E\left(\frac{\partial}{\partial\lambda}g_{\lambda n}\right), E\left(\frac{\partial}{\partial\beta'}g_{\lambda n}\right)\right]\Bigg|_{\lambda_0,\beta_0} = -\begin{bmatrix} (G_nX_n\beta_0)'Q_n & \sigma_0^2\mathrm{tr}(P^{(s)}_{1n}G_n) & \cdots & \sigma_0^2\mathrm{tr}(P^{(s)}_{mn}G_n) \\ X'_nQ_n & 0 & \cdots & 0 \end{bmatrix}',$$

and $\Omega_{21} = E(g_{\lambda n}\bar{g}^{\star\prime}_{\lambda n})|_{\rho_0,\beta_0}$ in the form of (18), the desirable invariant matrix is

$$A_n = -\begin{bmatrix} 0 & \sigma_0^{-2} & \sigma_0^{-2} & 0 \\ \sigma_0^{-2}I_k & 0 & 0 & b' \end{bmatrix}',$$

where $b = (b'_1, \ldots, b'_{k^*})'$ with $b_l = -\frac{\sigma_0^{-3}\eta_3}{(\eta_4-1)-\eta_3^2}e'_{kl}$ for $l = 1, \ldots, k^*$. With some simplified identities of those in the proof of Proposition 1, we have $\Omega_{21}A_n = D_2$. Finally, $\Sigma_{B\lambda} = \lim_{n\to\infty}\frac{1}{n}D'_1A_n$, with $D_1 = \left[E\left(\frac{\partial}{\partial\lambda}\bar{g}^\star_{\lambda n}\right), E\left(\frac{\partial}{\partial\beta'}\bar{g}^\star_{\lambda n}\right)\right]\Big|_{\lambda_0,\beta_0}$. □

Proof of Proposition 5. We shall derive the best moment function P^*_n analytically. With m quadratic moments in $g_n(\rho)$, $\mathrm{var}(g_n(\rho_0)) =$

$\sigma_0^4\Omega_n$, where $\Omega_n = (\eta_4-3)\omega_m'\omega_m+V_n$, with $\omega_m = [\mathrm{vec}_D(P_{1n}), \ldots, \mathrm{vec}_D(P_{mn})]$ and

$$V_n = \frac{1}{2}(\mathrm{vec}(P_{1n}^{(s)}), \ldots, \mathrm{vec}(P_{mn}^{(s)}))'(\mathrm{vec}(P_{1n}^{(s)}), \ldots, \mathrm{vec}(P_{mn}^{(s)}))$$
$$= \begin{bmatrix} \mathrm{tr}(P_{1n}^{(s)}P_{1n}) & \cdots & \mathrm{tr}(P_{1n}^{(s)}P_{mn}) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(P_{mn}^{(s)}P_{1n}) & \cdots & \mathrm{tr}(P_{mn}^{(s)}P_{mn}) \end{bmatrix}. \quad (19)$$

The two terms in Ω_n can be combined into a unified one as follows. First, because

$$\mathrm{tr}(P_{jn}^{(s)}P_{ln}) - \mathrm{vec}(P_{jn}-D(P_{jn}))^{(s)}\mathrm{vec}(P_{jn}-D(P_{jn}))$$
$$= \mathrm{tr}(P_{jn}^{(s)}P_{ln}) - \mathrm{tr}[(P_{jn}-D(P_{jn}))^{(s)}(P_{jn}-D(P_{jn}))]$$
$$= \mathrm{tr}(P_{jn}^{(s)}P_{ln}) - \mathrm{tr}[(P_{jn}-D(P_{jn}))^{(s)}P_{ln}]$$
$$= 2\mathrm{tr}[D(P_{jn})P_{ln}] = 2\mathrm{tr}[D(P_{jn})D(P_{ln})]$$
$$= 2\mathrm{vec}_D'(P_{jn})\mathrm{vec}_D(P_{ln}),$$

for any j and l, we have

$$\begin{bmatrix} \mathrm{tr}(P_{1n}^{(s)}P_{1n}) & \cdots & \mathrm{tr}(P_{1n}^{(s)}P_{mn}) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(P_{mn}^{(s)}P_{1n}) & \cdots & \mathrm{tr}(P_{mn}^{(s)}P_{mn}) \end{bmatrix} - 2\omega_m'\omega_m = \frac{1}{2}\varpi_m'\varpi_m,$$

where $\varpi_m = [\mathrm{vec}(P_{1n} - D(P_{1n}))^{(s)}, \ldots, \mathrm{vec}(P_{mn} - D(P_{mn}))^{(s)}]$. Therefore, $\Omega_n = \frac{1}{2}\left[2(\eta_4-1)\omega_m'\omega_m + \varpi_m'\varpi_m\right]$. Define the modified matrices $P_{jn}^{+} = P_{jn} - D(P_{jn}) + \sqrt{\frac{\eta_4-1}{2}}D(P_{jn})$ for $j = 1, \ldots, m$. As

$$\mathrm{vec}'(P_{jn}^{+(s)})\mathrm{vec}(P_{kn}^{+(s)}) = \mathrm{tr}(P_{jn}^{+(s)}P_{kn}^{+(s)})$$
$$= \mathrm{tr}\{[P_{jn}^{(s)} - D(P_{jn}^{(s)})][P_{kn}^{(s)} - D(P_{kn}^{(s)})]\} + 2(\eta_4-1)\mathrm{tr}[D(P_{jn})D(P_{kn})]$$
$$= \mathrm{vec}'[(P_{jn} - D(P_{jn}))^{(s)}] \times \mathrm{vec}[(P_{kn} - D(P_{kn}))^{(s)}] + 2(\eta_4-1)\mathrm{vec}_D'(P_{jn})\mathrm{vec}_D(P_{kn}),$$

it follows that $\Omega_n = \frac{1}{2}(\mathrm{vec}(P_{1n}^{+(s)}), \ldots, \mathrm{vec}(P_{kn}^{+(s)}))'(\mathrm{vec}(P_{1n}^{+(s)}), \ldots, \mathrm{vec}(P_{mn}^{+(s)}))$.

Consider now $\mathrm{tr}(P_{jn}^{(s)}H_n) = \mathrm{tr}(P_{jn}^{(s)}H_n^{(t)})$. We would like to find a matrix A_n such that $\mathrm{tr}(P_{jn}^{(s)}H_n^{(t)}) = \mathrm{tr}(P_{jn}^{+(s)}(H_n^{(t)} + A_n))$ holds for all j. By taking A_n to be a diagonal matrix, the solution is $A_n = \left(\sqrt{\frac{2}{\eta_4-1}} - 1\right)D(H_n^{(t)})$, which is invariant with any P_{jn}. Denote $H_n^{-} = H_n^{(t)} + A_n = H_n^{(t)} + \left(\sqrt{\frac{2}{\eta_4-1}} - 1\right)D(H_n^{(t)})$, which has zero trace. Therefore, $\mathrm{tr}(P_{jn}^{(s)}H_n) = \mathrm{tr}(P_{jn}^{+(s)}H_n^{-})$.

Following Lee (2001a), the limit variance of the GMME with $P_{jn}, j = 1, \ldots, m$, is $\Sigma_P^{-1} = \left(\lim_{n\to\infty}\frac{1}{n}\Sigma_{P,n}\right)^{-1}$, where $\Sigma_{P,n} = (\mathrm{tr}(P_{1n}^{(s)}H_n), \ldots, \mathrm{tr}(P_{mn}^{(s)}H_n))\Omega_n^{-1}(\mathrm{tr}(P_{1n}^{(s)}H_n), \ldots, \mathrm{tr}(P_{mn}^{(s)}H_n))'$. With the above manipulation, $\Sigma_{P,n}$ can be rewritten as $\Sigma_{P,n} = \frac{1}{2}\mathrm{vec}'(H_n^{-(s)})\tilde{\omega}_m(\tilde{\omega}_m'\tilde{\omega}_m)^{-1}\tilde{\omega}_m'\mathrm{vec}(H_n^{-(s)})$ with $\tilde{\omega}_m = [\mathrm{vec}(P_{1n}^{+(s)}), \ldots, \mathrm{vec}(P_{mn}^{+(s)})]$.

By the generalized Schwarz inequality, $\Sigma_{P,n} \le \frac{1}{2}\mathrm{vec}'(H_n^{-(s)})\mathrm{vec}(H_n^{-(s)})$, which provides a bound for the precision matrix $\Sigma_{P,n}$ for any GMME with a finite number of quadratic moments. This bound can be obtained with a corresponding optimum $P_n^{+*} = H_n^{(t)} + \left(\sqrt{\frac{2}{\eta_4-1}} - 1\right)D(H_n^{(t)})$. With P_n^{+} transformed back to the P_n, the best P_n^* is $P_n^* = P_n^{+*} - D(P_n^{+*}) + \sqrt{\frac{2}{\eta_4-1}}D(P_n^{+*}) = H_n^{(t)} - \frac{\eta_4-3}{\eta_4-1}D(H_n^{(t)})$.

Furthermore, as $\Sigma_B = \lim_{n\to\infty}\frac{1}{n}D_n^{*\prime}\Omega_n^{*-1}D_n^*$ (Lee, 2001a), where $\Omega_n^* = \mathrm{var}(g_n^*(\rho_0)) = \sigma_0^4\mathrm{tr}(P_n^{*(s)}H_n)$ and $D_n^* = E\left(\frac{\partial}{\partial\rho}g_n^*(\rho_0)\right) = -\sigma_0^2\mathrm{tr}(P_n^{*(s)}H_n)$, it follows that $\Sigma_B = \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(P_n^{*(s)}H_n)$. □

References

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge.

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

Anselin, L., 1990. Some robust approaches to testing and estimation in spatial econometrics. Regional Science and Urban Economics 20, 141–163.

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

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

Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.

Cressie, N., 1993. Statistics for Spatial Data. John Wiley and Sons, New York.

Han, C., Phillips, P.C., 2006. GMM with many moment conditions. Econometrica 74, 147–192.

Horn, R., Johnson, C., 1985. Matrix Analysis. Cambridge University Press, Cambridge.

Kelejian, H.H., Prucha, I.R., 1997. Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem. International Regional Science Review 20, 103–111.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. Journal of Real Estate Finance and Economics 17, 99–121.

Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.

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

Kelejian, H.H., Robinson, D.P., 1993. A suggested method of estimation for spatial interdependence models with autocorrelated errors, and an application to a county expenditure model. Papers in Regional Science 72, 297–312.

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

Lee, L.F., 2001b. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. 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., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 137, 489–514.

Lin, X., Lee, L.F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics 157, 34–52.

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

Robinson, P.M., 2010. Efficient estimation of the semiparametric spatial autoregressive model. Journal of Econometrics 157, 6–17.

Smirnov, O., Anselin, L., 2001. Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. Computational Statistics and Data Analysis 35, 301–319.