GMM Estimation of SAR Models with Endogenous Regressors*

Xiaodong Liu
Department of Economics,
University of Colorado Boulder
E-mail: xiaodong.liu@colorado.edu

Paulo Saraiva
Department of Economics,
University of Colorado Boulder
E-mail: saraiva@colorado.edu

August 2015

Abstract

In this paper, we extend the GMM estimator in Lee (2007) to estimate SAR models with endogenous regressors. We propose a new set of quadratic moment conditions exploiting the correlation of the spatially lagged dependent variable with the disturbance term of the main regression equation and with the endogenous regressor. The proposed GMM estimator is more efficient than the IV-based linear estimators in the literature, and computationally simpler than the ML estimator. With carefully constructed quadratic moment equations, the GMM estimator can be asymptotically as efficient as the ML estimator under normality. Monte Carlo experiments show that the proposed GMM estimator performs well in finite samples.

Key Words: Spatial models, endogeneity, simultaneous equations, moment conditions, efficiency

JEL Classification: C31, C36, R15

*We would like to thank the Co-Editor and an anonymous referee for helpful comments and suggestions. The remaining errors are our own.

1 Introduction

In recent years, spatial econometric models play a vital role in empirical research on regional and urban economics. By expanding the notion of space from geographic space to "economic" space and "social" space, these models can be used to study cross-sectional interactions in much wider applications including education (e.g. Lin, 2010; Sacerdote, 2011; Carrell et al., 2013), crime (e.g. Patacchini and Zenou, 2012; Lindquist and Zenou, 2014), industrial organization (e.g. König et al., 2014), finance (e.g. Denbee et al., 2014), etc.

Among spatial econometric models, the spatial autoregressive (SAR) model introduced by Cliff and Ord (1973, 1981) has received the most attention. In this model, the cross-sectional dependence is modeled as the weighted average outcome of neighboring units, typically referred to as the spatially lagged dependent variable. As the spatially lagged dependent variable is endogenous, likelihood- and moment-based methods have been proposed to estimate the SAR model (e.g. Kelejian and Prucha, 1998; Lee, 2004; Lee, 2007; Lee and Liu, 2010). In particular, for the SAR model with exogenous regressors, Lee (2007) proposes a generalized method of moments (GMM) estimator that combines linear moment conditions, with the (estimated) mean of the spatially lagged dependent variable as the instrumental variable (IV), and quadratic moment conditions based on the covariance structure of the spatially lagged dependent variable and the model disturbance term. The GMM estimator improves estimation efficiency of IV-based linear estimators in Kelejian and Prucha (1998) and is computationally simple relative to the maximum likelihood (ML) estimator in Lee (2004). Furthermore, Lin and Lee (2010) show that a sub-class of the GMM estimators is consistent in the presence of an unknown form of heteroskedasticity in model disturbances, and thus more robust relative to the ML estimator.

For SAR models with endogenous regressors, Liu (2012) and Liu and Lee (2013) consider, respectively, the limited information maximum likelihood (LIML) and two stage least squares (2SLS) estimators, in the presence of many potential IVs. Liu and Lee (2013) also propose a criterion based on the approximate mean square error of the 2SLS estimator to select the optimal set of IVs. The SAR model with endogenous regressors can be considered as an equation in a system of simultaneous equations. For the full information estimation of the system, Kelejian and Prucha (2004) propose a three stage least squares (3SLS) estimator and, in a recent paper, Yang and Lee (2014) consider the quasi-maximum likelihood (QML) approach. The QML estimator is asymptotically more efficient than the 3SLS estimator under normality but can be computationally difficult to implement. The existing estimators for the SAR model with endogenous regressors are summarized in Table 1.

Table 1: Existing Estimators for SAR Models with Endogenous Regressors

	single-equation estimator	**system estimator**
IV-based linear estimator	Liu and Lee (2013)	Kelejian and Prucha (2004)
likelihood-based estimator	Liu (2012)	Yang and Lee (2014)

In this paper, we extend the GMM estimator in Lee (2007) to estimate SAR models with endogenous regressors. We propose a new set of quadratic moment equations exploiting (i) the covariance structure of the spatially lagged dependent variable and the disturbance term of the main regression equation and (ii) the covariance structure of the spatially lagged dependent variable and the endogenous regressor. We establish the identification, consistency and asymptotic normality of the proposed GMM estimator. The GMM estimator is more efficient than the 2SLS and 3SLS estimators, and computationally simpler than the ML estimator. With carefully constructed quadratic moment equations, the GMM estimator can be asymptotically as efficient as the ML estimator under normality. We also conduct a limited Monte Carlo experiment to show that the proposed GMM estimator performs well in finite samples.

The rest of the paper is organized as follows. In Section 2, we introduce the SAR model with endogenous regressors. In Section 3, we define the GMM estimator and discuss the identification of model parameters. In Section 4, we study the asymptotic properties of the GMM estimator and discuss the optimal moment conditions to use. Section 5 reports Monte Carlo experiment results. Section 6 briefly concludes. The proofs are collected in the appendix.

Throughout the paper, we adopt the following notation. For an $n \times n$ matrix $\mathbf{A} = [a_{ij}]_{i,j=1,\cdots,n}$, let $\mathbf{A}^{(s)} = \mathbf{A} + \mathbf{A}'$, $\mathrm{vec}_D(\mathbf{A}) = (a_{11}, \cdots, a_{nn})'$, and $\mathrm{diag}(\mathbf{A}) = \mathrm{diag}(a_{11}, \cdots, a_{nn})$. The row (or column) sums of $\mathbf{A}$ are uniformly bounded in absolute value if $\max_{i=1,\cdots,n} \sum_{j=1}^{n} |a_{ij}|$ (or $\max_{j=1,\cdots,n} \sum_{i=1}^{n} |a_{ij}|$) is bounded.

2 Model

Consider a SAR model with an endogenous regressor[1] given by

$$\mathbf{y}_1 = \lambda_0 \mathbf{W} \mathbf{y}_1 + \phi_0 \mathbf{y}_2 + \mathbf{X}_1 \boldsymbol{\beta}_0 + \mathbf{u}_1, \tag{1}$$

[1] In this paper, we focus on the model with a single endogenous regressor for exposition purpose. The model and proposed estimator can be easily generalized to accommodate any fixed number of endogenous regressors.

where $\mathbf{y}_1$ is an $n \times 1$ vector of observations on the dependent variable, $\mathbf{W}$ is an $n \times n$ nonstochastic spatial weights matrix with a zero diagonal, $\mathbf{y}_2$ is an $n \times 1$ vector of observations on an endogenous regressor, $\mathbf{X}_1$ is an $n \times K_1$ matrix of observations on K_1 nonstochastic exogenous regressors, and $\mathbf{u}_1$ is an $n \times 1$ vector of i.i.d. innovations.[2] $\mathbf{W}\mathbf{y}_1$ is usually referred to as the spatially lagged dependent variable. Let $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, where $\mathbf{X}_2$ is an $n \times K_2$ matrix of observations on K_2 excluded nonstochastic exogenous variables. The reduced form of the endogenous regressor $\mathbf{y}_2$ is assumed to be

$$\mathbf{y}_2 = \mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{u}_2, \tag{2}$$

where $\mathbf{u}_2$ is an $n \times 1$ vector of i.i.d. innovations. Let $\boldsymbol{\theta}_0 = (\boldsymbol{\delta}_0', \boldsymbol{\gamma}_0')'$, with $\boldsymbol{\delta}_0 = (\lambda_0, \phi_0, \boldsymbol{\beta}_0')'$, denote the vector of true parameter values in the data generating process (DGP). The following regularity conditions are common in the literature of SAR models (see, e.g., Lee, 2007; Kelejian and Prucha, 2010).

Assumption 1 *Let $u_{1,i}$ and $u_{2,i}$ denote, respectively, the i-th elements of $\mathbf{u}_1$ and $\mathbf{u}_2$.* (i) $(u_{1,i}, u_{2,i})' \sim$ i.i.d.$(\mathbf{0}, \boldsymbol{\Sigma})$*, where*

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}.$$

(ii) $\mathrm{E}|u_{k,i}u_{l,i}u_{r,i}u_{s,i}|^{1+\eta}$ *is bounded for $k, l, r, s = 1, 2$ and some small constant $\eta > 0$.*

Assumption 2 (i) *The elements of $\mathbf{X}$ are uniformly bounded constants.* (ii) $\mathbf{X}$ *has full column rank $K_X = K_1 + K_2$.* (iii) $\lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{X}$ *exists and is nonsingular.*

Assumption 3 (i) *All diagonal elements of the spatial weights matrix $\mathbf{W}$ are zero.* (ii) $\lambda_0 \in (-\underline{\lambda}, \overline{\lambda})$ *with $0 < \underline{\lambda}, \overline{\lambda} \leq c_\lambda < \infty$.* (iii) $\mathbf{S}(\lambda) = \mathbf{I}_n - \lambda\mathbf{W}$ *is nonsingular for all $\lambda \in (-\underline{\lambda}, \overline{\lambda})$.* (iv) *The row and column sums of $\mathbf{W}$ and $\mathbf{S}(\lambda_0)^{-1}$ are uniformly bounded in absolute value.*

Assumption 4 $\boldsymbol{\theta}_0$ *is in the interior of a compact parameter space Θ.*

[2] $\mathbf{y}_1, \mathbf{y}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{X}, \mathbf{W}$ are allowed to depend on the sample size n, i.e., to formulate triangular arrays as in Kelejian and Prucha (2010). Nevertheless, we suppress the subscript n to simplify the notation.

3 GMM Estimation

3.1 Estimator

Let $\mathbf{S} = \mathbf{S}(\lambda_0) = \mathbf{I}_n - \lambda_0 \mathbf{W}$ and $\mathbf{G} = \mathbf{W}\mathbf{S}^{-1}$. Under Assumption 3, model (1) has a reduced form

$$\mathbf{y}_1 = \mathbf{S}^{-1}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{S}^{-1}\mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{S}^{-1}\mathbf{u}_1 + \phi_0\mathbf{S}^{-1}\mathbf{u}_2, \tag{3}$$

which implies that

$$\mathbf{W}\mathbf{y}_1 = \mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{G}\mathbf{u}_1 + \phi_0\mathbf{G}\mathbf{u}_2. \tag{4}$$

As $\mathbf{W}\mathbf{y}_1$ and $\mathbf{y}_2$ are endogenous, consistent estimation of (1) requires IVs for $\mathbf{W}\mathbf{y}_1$ and $\mathbf{y}_2$. From (4), the deterministic part of $\mathbf{W}\mathbf{y}_1$ is a linear combination of the columns in $\mathbf{G}\mathbf{X} = [\mathbf{G}\mathbf{X}_1, \mathbf{G}\mathbf{X}_2]$. Therefore, $\mathbf{G}\mathbf{X}$ can be used as an IV matrix for $\mathbf{W}\mathbf{y}_1$.[3] From (2), $\mathbf{X}$ can be used as an IV matrix for $\mathbf{y}_2$. In general, let $\mathbf{Q}$ be an $n \times K_Q$ matrix of IVs such that $\mathrm{E}(\mathbf{Q}'\mathbf{u}_1) = \mathrm{E}(\mathbf{Q}'\mathbf{u}_2) = \mathbf{0}$. Let $\mathbf{u}_1(\boldsymbol{\delta}) = \mathbf{S}(\lambda)\mathbf{y}_1 - \phi\mathbf{y}_2 - \mathbf{X}_1\boldsymbol{\beta}$ and $\mathbf{u}_2(\boldsymbol{\gamma}) = \mathbf{y}_2 - \mathbf{X}\boldsymbol{\gamma}$, where $\boldsymbol{\delta} = (\lambda, \phi, \boldsymbol{\beta}')'$. The linear moment function for the GMM estimation is given by

$$\mathbf{g}_1(\boldsymbol{\theta}) = (\mathbf{I}_2 \otimes \mathbf{Q})'\mathbf{u}(\boldsymbol{\theta}),$$

where $\otimes$ denotes the Kronecker product, $\mathbf{u}(\boldsymbol{\theta}) = [\mathbf{u}_1(\boldsymbol{\delta})', \mathbf{u}_2(\boldsymbol{\gamma})']'$, and $\boldsymbol{\theta} = (\boldsymbol{\delta}', \boldsymbol{\gamma}')'$.[4]

Besides the linear moment functions, Lee (2007) proposes to use quadratic moment functions based on the covariance structure of the spatially lagged dependent variable and model disturbances to improve estimation efficiency. We generalize this idea to SAR models with endogenous regressors. Substitution of (2) into (1) leads to a "pseudo" reduced form

$$\mathbf{y}_1 = \lambda_0\mathbf{W}\mathbf{y}_1 + \phi_0\mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{X}_1\boldsymbol{\beta}_0 + \mathbf{u}_1 + \phi_0\mathbf{u}_2. \tag{5}$$

By exploiting the covariance structure of the spatially lagged dependent variable $\mathbf{W}\mathbf{y}_1$ and the

[3] The IV matrix $\mathbf{G}\mathbf{X}$ is not feasible as $\mathbf{G}$ involves the unknown parameter λ_0. Under Assumption 3, $\mathbf{G}\mathbf{X} = \mathbf{W}\mathbf{X} + \lambda_0\mathbf{W}^2\mathbf{X} + \lambda_0^2\mathbf{W}^3\mathbf{X} + \cdots$. Therefore, we can use the leading order terms $\mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}, \mathbf{W}^3\mathbf{X}$ of the series expansion as feasible IVs for $\mathbf{W}\mathbf{y}$.

[4] In practice, we could use two different IV matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ to construct linear moment functions $\mathbf{Q}_1'\mathbf{u}_1(\boldsymbol{\delta})$ and $\mathbf{Q}_2'\mathbf{u}_2(\boldsymbol{\gamma})$. The GMM estimator with $\mathbf{g}_1(\boldsymbol{\theta})$ is (asymptotically) no less efficient than that with $\mathbf{Q}_1'\mathbf{u}_1(\boldsymbol{\delta})$ and $\mathbf{Q}_2'\mathbf{u}_2(\boldsymbol{\gamma})$ if $\mathbf{Q}$ includes all linearly independent columns of $\mathbf{Q}_1$ and $\mathbf{Q}_2$.

disturbances of (5), we propose the following quadratic moment functions

$$\mathbf{g}_2(\boldsymbol{\theta}) = [\mathbf{g}_{2,11}(\boldsymbol{\delta})', \mathbf{g}_{2,12}(\boldsymbol{\theta})', \mathbf{g}_{2,21}(\boldsymbol{\theta})', \mathbf{g}_{2,22}(\gamma)']'$$

with

$$\begin{aligned}
\mathbf{g}_{2,11}(\boldsymbol{\delta}) &= [\boldsymbol{\Xi}_1'\mathbf{u}_1(\boldsymbol{\delta}), \cdots, \boldsymbol{\Xi}_m'\mathbf{u}_1(\boldsymbol{\delta})]'\mathbf{u}_1(\boldsymbol{\delta}) \\
\mathbf{g}_{2,12}(\boldsymbol{\theta}) &= [\boldsymbol{\Xi}_1'\mathbf{u}_1(\boldsymbol{\delta}), \cdots, \boldsymbol{\Xi}_m'\mathbf{u}_1(\boldsymbol{\delta})]'\mathbf{u}_2(\gamma) \\
\mathbf{g}_{2,21}(\boldsymbol{\theta}) &= [\boldsymbol{\Xi}_1'\mathbf{u}_2(\gamma), \cdots, \boldsymbol{\Xi}_m'\mathbf{u}_2(\gamma)]'\mathbf{u}_1(\boldsymbol{\delta}) \\
\mathbf{g}_{2,22}(\gamma) &= [\boldsymbol{\Xi}_1'\mathbf{u}_2(\gamma), \cdots, \boldsymbol{\Xi}_m'\mathbf{u}_2(\gamma)]'\mathbf{u}_2(\gamma)
\end{aligned}$$

where $\boldsymbol{\Xi}_j$ is an $n \times n$ constant matrix with $\mathrm{tr}(\boldsymbol{\Xi}_j) = 0$ for $j = 1, \cdots, m$.[5] Possible candidates for $\boldsymbol{\Xi}_j$ are $\mathbf{W}$, $\mathbf{W}^2 - n^{-1}\mathrm{tr}(\mathbf{W}^2)\mathbf{I}_n$, etc.[6] These quadratic moment functions are based on the moment conditions that $\mathrm{E}(\mathbf{u}_1'\boldsymbol{\Xi}_j\mathbf{u}_1) = \mathrm{E}(\mathbf{u}_1'\boldsymbol{\Xi}_j\mathbf{u}_2) = \mathrm{E}(\mathbf{u}_2'\boldsymbol{\Xi}_j\mathbf{u}_1) = \mathrm{E}(\mathbf{u}_2'\boldsymbol{\Xi}_j\mathbf{u}_2) = 0$ for $j = 1, \cdots, m$.

Let

$$\mathbf{g}(\boldsymbol{\theta}) = [\mathbf{g}_1(\boldsymbol{\theta})', \mathbf{g}_2(\boldsymbol{\theta})']', \tag{6}$$

and $\boldsymbol{\Omega} = \mathrm{Var}[\mathbf{g}(\boldsymbol{\theta}_0)]$. The following assumption is from Lee (2007).

Assumption 5 (i) *The elements of* $\mathbf{Q}$ *are uniformly bounded constants.* (ii) $\boldsymbol{\Xi}_j$ *is an* $n\times n$ *constant matrix with* $\mathrm{tr}(\boldsymbol{\Xi}_j) = 0$ *for* $j = 1, \cdots, m$. *The row and column sums of* $\boldsymbol{\Xi}_j$ *are uniformly bounded in absolute value.* (iii) $\lim_{n\to\infty} n^{-1}\boldsymbol{\Omega}$ *exists and is nonsingular.*

Combining both linear and quadratic moment functions, the GMM estimator of $\boldsymbol{\theta}_0$ is given by

$$\widetilde{\boldsymbol{\theta}}_{gmm} = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbf{g}(\boldsymbol{\theta})'\mathbf{F}'\mathbf{F}\mathbf{g}(\boldsymbol{\theta}), \tag{7}$$

for some matrix $\mathbf{F}$ such that $\lim_{n\to\infty} \mathbf{F}$ exists and has full row rank greater than or equal to $\dim(\boldsymbol{\theta})$. In the GMM literature, $\mathbf{F}'\mathbf{F}$ is known as the GMM weighting matrix. For instance, one can use the identity matrix as the weighting matrix to implement the GMM. The asymptotic efficiency of the

[5] In practice, we could use different sets of weighting matrices $\{\boldsymbol{\Xi}_{11,j}\}_{j=1}^{m_{11}}$, $\{\boldsymbol{\Xi}_{12,j}\}_{j=1}^{m_{12}}$, $\{\boldsymbol{\Xi}_{21,j}\}_{j=1}^{m_{21}}$ and $\{\boldsymbol{\Xi}_{22,j}\}_{j=1}^{m_{22}}$ for the quadratic moment functions $\mathbf{g}_{2,11}(\boldsymbol{\theta})$, $\mathbf{g}_{2,12}(\boldsymbol{\theta})$, $\mathbf{g}_{2,21}(\boldsymbol{\theta})$ and $\mathbf{g}_{2,22}(\boldsymbol{\theta})$ respectively. The quadratic moment functions $\mathbf{g}_2(\boldsymbol{\theta})$ are (asymptotically) no less efficient than that with $\{\boldsymbol{\Xi}_{11,j}\}_{j=1}^{m_{11}}$, $\{\boldsymbol{\Xi}_{12,j}\}_{j=1}^{m_{12}}$, $\{\boldsymbol{\Xi}_{21,j}\}_{j=1}^{m_{21}}$ and $\{\boldsymbol{\Xi}_{22,j}\}_{j=1}^{m_{22}}$ if $\{\boldsymbol{\Xi}_1, \ldots, \boldsymbol{\Xi}_m\} = \{\boldsymbol{\Xi}_{11,j}\}_{j=1}^{m_{11}} \cup \{\boldsymbol{\Xi}_{12,j}\}_{j=1}^{m_{12}} \cup \{\boldsymbol{\Xi}_{21,j}\}_{j=1}^{m_{21}} \cup \{\boldsymbol{\Xi}_{22,j}\}_{j=1}^{m_{22}}$.

[6] We discuss the optimal $\mathbf{Q}$ and $\boldsymbol{\Xi}$ in Section 4.2.

GMM estimator depends on the choice of the weighting matrix as discussed in Section 4.1.

3.2 Identification

For $\boldsymbol{\theta}_0$ to be identified through the moment functions $\mathbf{g}(\boldsymbol{\theta})$, $\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})] = \mathbf{0}$ needs to have a unique solution at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (Hansen, 1982). As $\mathbf{S}(\lambda)\mathbf{S}^{-1} = \mathbf{I}_n + (\lambda_0 - \lambda)\mathbf{G}$, it follows from (2) and (3) that

$$\mathbf{u}_1(\boldsymbol{\delta}) = \mathbf{d}_1(\boldsymbol{\delta}) + [\mathbf{I}_n + (\lambda_0 - \lambda)\mathbf{G}]\mathbf{u}_1 + [(\phi_0 - \phi)\mathbf{I}_n + \phi_0(\lambda_0 - \lambda)\mathbf{G}]\mathbf{u}_2$$

and

$$\mathbf{u}_2(\boldsymbol{\gamma}) = \mathbf{d}_2(\boldsymbol{\gamma}) + \mathbf{u}_2,$$

where $\mathbf{d}_1(\boldsymbol{\delta}) = [\mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1](\boldsymbol{\delta}_0 - \boldsymbol{\delta})$ and $\mathbf{d}_2(\boldsymbol{\gamma}) = \mathbf{X}(\boldsymbol{\gamma}_0 - \boldsymbol{\gamma})$.

For the linear moment functions, we have

$$\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\delta})] = \lim_{n\to\infty} n^{-1}\mathbf{Q}'\mathbf{d}_1(\boldsymbol{\delta}) = \lim_{n\to\infty} n^{-1}\mathbf{Q}'[\mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1](\boldsymbol{\delta}_0 - \boldsymbol{\delta})$$

and

$$\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{Q}'\mathbf{u}_2(\boldsymbol{\gamma})] = \lim_{n\to\infty} n^{-1}\mathbf{Q}'\mathbf{d}_2(\boldsymbol{\gamma}) = \lim_{n\to\infty} n^{-1}\mathbf{Q}'\mathbf{X}(\boldsymbol{\gamma}_0 - \boldsymbol{\gamma})$$

Therefore, $\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{g}_1(\boldsymbol{\theta})] = \mathbf{0}$ has a unique solution at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, if $\mathbf{Q}'[\mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1]$ and $\mathbf{Q}'\mathbf{X}$ have full column rank for large enough n. This sufficient rank condition implies the necessary rank condition that $[\mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1]$ and $\mathbf{X}$ have full column rank and the rank of $\mathbf{Q}$ is at least $\max\{\dim(\boldsymbol{\delta}), K_X\}$, for large enough n.

Suppose $[\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1]$ has full column rank for large enough n.[7] The necessary rank condition for identification does not hold if $\mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0$ and $[\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1]$ are asymptotically linearly dependent.[8] $\mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0$ and $[\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1]$ are linearly dependent if there exist a constant scalar c_1 and a $K_1 \times 1$ constant vector $\mathbf{c}_2$ such that $\mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0 = c_1\mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{X}_1\mathbf{c}_2$, which implies that

$$\mathbf{d}_1(\boldsymbol{\delta}) = [(\lambda_0 - \lambda)c_1 + (\phi_0 - \phi)]\mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{X}_1[(\lambda_0 - \lambda)\mathbf{c}_2 + (\boldsymbol{\beta}_0 - \boldsymbol{\beta})].$$

Hence, the solutions of the linear moment equations $\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\delta})] = \mathbf{0}$ are characterized

[7] As $\mathbf{X}\boldsymbol{\gamma}_0 = \mathbf{X}_1\boldsymbol{\gamma}_{10} + \mathbf{X}_2\boldsymbol{\gamma}_{20}$, a necessary condition for $(\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1)$ to have full column rank is $\boldsymbol{\gamma}_{20} \neq \mathbf{0}$.

[8] A necessary condition for $\mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0$ and $[\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1]$ to be asymptotically linearly independent is $(\phi_0, \boldsymbol{\beta}_0')' \neq \mathbf{0}$.

by

$$\phi = \phi_0 + (\lambda_0 - \lambda)c_1 \quad \text{and} \quad \boldsymbol{\beta} = \boldsymbol{\beta}_0 + (\lambda_0 - \lambda)\mathbf{c}_2 \tag{8}$$

as long as $\mathbf{Q}'[\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1]$ has full column rank for large enough n. In this case, ϕ_0 and $\boldsymbol{\beta}_0$ can be identified if and only if λ_0 can be identified from the quadratic moment equations.

Given (8), we have

$$\begin{aligned} \mathrm{E}[\mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_j\mathbf{u}_1(\boldsymbol{\delta})] &= (\lambda_0 - \lambda)(\sigma_1^2 + \phi_0\sigma_{12})\mathrm{tr}(\boldsymbol{\Xi}_j^{(s)}\mathbf{G}) \\ &\quad +(\lambda_0 - \lambda)^2[(\sigma_1^2 + 2\phi_0\sigma_{12} + \phi_0^2\sigma_2^2)\mathrm{tr}(\mathbf{G}'\boldsymbol{\Xi}_j\mathbf{G}) - c_1(\sigma_{12} + \phi_0\sigma_2^2)\mathrm{tr}(\boldsymbol{\Xi}_j^{(s)}\mathbf{G})] \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}[\mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_j\mathbf{u}_2(\boldsymbol{\gamma})] &= (\lambda_0 - \lambda)(\sigma_{12} + \phi_0\sigma_2^2)\mathrm{tr}(\boldsymbol{\Xi}_j'\mathbf{G}) \\ \mathrm{E}[\mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}_j\mathbf{u}_1(\boldsymbol{\delta})] &= (\lambda_0 - \lambda)(\sigma_{12} + \phi_0\sigma_2^2)\mathrm{tr}(\boldsymbol{\Xi}_j\mathbf{G}) \end{aligned}$$

for $j = 1, \cdots, m$. If $(\sigma_1^2 + \phi_0\sigma_{12})\lim_{n\to\infty} n^{-1}\mathrm{tr}(\boldsymbol{\Xi}_j^{(s)}\mathbf{G}) \neq 0$ for some $j \in \{1, \cdots, m\}$, the quadratic moment equation

$$\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_j\mathbf{u}_1(\boldsymbol{\delta})] = 0$$

has two roots $\lambda = \lambda_0$ and

$$\lambda = \lambda_0 + \frac{(\sigma_1^2 + \phi_0\sigma_{12})}{(\sigma_1^2 + 2\phi_0\sigma_{12} + \phi_0^2\sigma_2^2)\lim_{n\to\infty}[\mathrm{tr}(\mathbf{G}'\boldsymbol{\Xi}_j\mathbf{G})/\mathrm{tr}(\boldsymbol{\Xi}_j^{(s)}\mathbf{G})] - c_1(\sigma_{12} + \phi_0\sigma_2^2)}.$$

As $(\sigma_1^2 + 2\phi_0\sigma_{12} + \phi_0^2\sigma_2^2) > 0$, if $\lim_{n\to\infty}[\mathrm{tr}(\mathbf{G}'\boldsymbol{\Xi}_j\mathbf{G})/\mathrm{tr}(\boldsymbol{\Xi}_j^{(s)}\mathbf{G})] \neq \lim_{n\to\infty}[\mathrm{tr}(\mathbf{G}'\boldsymbol{\Xi}_k\mathbf{G})/\mathrm{tr}(\boldsymbol{\Xi}_k^{(s)}\mathbf{G})]$ for some $j \neq k$, the moment equations

$$\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_j\mathbf{u}_1(\boldsymbol{\delta})] = 0 \qquad \text{and} \qquad \lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_k\mathbf{u}_1(\boldsymbol{\delta})] = 0$$

have a unique common root $\lambda = \lambda_0$. On the other hand, if $(\sigma_{12} + \phi_0\sigma_2^2)\lim_{n\to\infty} n^{-1}\mathrm{tr}(\boldsymbol{\Xi}_j'\mathbf{G}) \neq 0$ for some $j \in \{1, \cdots, m\}$, the quadratic moment equation

$$\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_j\mathbf{u}_2(\boldsymbol{\gamma})] = 0$$

has a unique root $\lambda = \lambda_0$; and if $(\sigma_{12} + \phi_0 \sigma_2^2) \lim_{n\to\infty} n^{-1}\mathrm{tr}(\mathbf{\Xi}_j \mathbf{G}) \neq 0$ for some $j \in \{1, \cdots, m\}$, the quadratic moment equation

$$\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{u}_2(\boldsymbol{\gamma})'\mathbf{\Xi}_j \mathbf{u}_1(\boldsymbol{\delta})] = 0$$

has a unique root $\lambda = \lambda_0$. To wrap up, the sufficient identification condition of $\boldsymbol{\theta}_0$ is summarized in the following assumption.

Assumption 6 $\lim_{n\to\infty} n^{-1}\mathbf{Q}'\mathbf{X}$ *and* $\lim_{n\to\infty} n^{-1}\mathbf{Q}'[\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1]$ *both have full column rank, and at least one of the following conditions is satisfied.* (i) $\lim_{n\to\infty} n^{-1}\mathbf{Q}'[\mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0 \mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}\boldsymbol{\gamma}_0, \mathbf{X}_1]$ *has full column rank.* (ii) $(\sigma_1^2 + \phi_0\sigma_{12}) \lim_{n\to\infty} n^{-1}\mathrm{tr}(\mathbf{\Xi}_j^{(s)}\mathbf{G}) \neq 0$ *for some* $j \in \{1, \cdots, m\}$*, and* $\lim_{n\to\infty} n^{-1}[\mathrm{tr}(\mathbf{\Xi}_1^{(s)}\mathbf{G}), \cdots, \mathrm{tr}(\mathbf{\Xi}_m^{(s)}\mathbf{G})]'$ *is linearly independent of* $\lim_{n\to\infty} n^{-1}[\mathrm{tr}(\mathbf{G}'\mathbf{\Xi}_1\mathbf{G}), \cdots, \mathrm{tr}(\mathbf{G}'\mathbf{\Xi}_m\mathbf{G})]'$. (iii) $(\sigma_{12} + \phi_0\sigma_2^2) \lim_{n\to\infty} n^{-1}\mathrm{tr}(\mathbf{\Xi}_j\mathbf{G}) \neq 0$ *or* $(\sigma_{12} + \phi_0\sigma_2^2) \lim_{n\to\infty} n^{-1}\mathrm{tr}(\mathbf{\Xi}_j'\mathbf{G}) \neq 0$ *for some* $j \in \{1, \cdots, m\}$.

4 Asymptotic Properties

4.1 Consistency and Asymptotic Normality

The GMM estimator defined in (7) falls into the class of Z-estimators (see Newey and McFadden, 1994). Therefore, to establish the consistency and asymptotic normality, it suffices to show that the GMM estimator satisfies the sufficient conditions for Z-estimators to be consistent and asymptotically normally distributed when properly normalized and centered. A similar argument has been adopted by Lee (2007) to establish the asymptotic normality of the GMM estimator for the SAR model with exogenous regressors.

Let $\mu_{r,s} = \mathrm{E}(u_{1,i}^r u_{2,i}^s)$ for $r + s = 3, 4$. By Lemmas B.1 and B.2 in the Appendix, we have

$$\mathbf{\Omega} = \mathrm{Var}[\mathbf{g}(\boldsymbol{\theta}_0)] = \begin{bmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{12}' & \mathbf{\Omega}_{22} \end{bmatrix} \tag{9}$$

with $\mathbf{\Omega}_{11} = \mathrm{Var}[\mathbf{g}_1(\boldsymbol{\theta}_0)] = \mathbf{\Sigma} \otimes (\mathbf{Q}'\mathbf{Q})$,

$$\mathbf{\Omega}_{12} = \mathrm{E}[\mathbf{g}_1(\boldsymbol{\theta}_0)\mathbf{g}_2(\boldsymbol{\theta}_0)'] = \begin{bmatrix} \mu_{3,0} & \mu_{2,1} & \mu_{2,1} & \mu_{1,2} \\ \mu_{2,1} & \mu_{1,2} & \mu_{1,2} & \mu_{0,3} \end{bmatrix} \otimes (\mathbf{Q}'\boldsymbol{\omega})$$

and

$$
\begin{aligned}
&\mathbf{\Omega}_{22} = \mathrm{Var}[\mathbf{g}_2(\boldsymbol{\theta}_0)] \\
= &\begin{bmatrix}
\mu_{4,0} - 3\sigma_1^4 & \mu_{3,1} - 3\sigma_1^2\sigma_{12} & \mu_{3,1} - 3\sigma_1^2\sigma_{12} & \mu_{2,2} - \sigma_1^2\sigma_2^2 - 2\sigma_{12}^2 \\
* & \mu_{2,2} - \sigma_1^2\sigma_2^2 - 2\sigma_{12}^2 & \mu_{2,2} - \sigma_1^2\sigma_2^2 - 2\sigma_{12}^2 & \mu_{1,3} - 3\sigma_{12}\sigma_2^2 \\
* & * & \mu_{2,2} - \sigma_1^2\sigma_2^2 - 2\sigma_{12}^2 & \mu_{1,3} - 3\sigma_{12}\sigma_2^2 \\
* & * & * & \mu_{0,4} - 3\sigma_2^4
\end{bmatrix} \otimes (\boldsymbol{\omega}'\boldsymbol{\omega}) \\
&+ \begin{bmatrix}
\sigma_1^4 & \sigma_1^2\sigma_{12} & \sigma_1^2\sigma_{12} & \sigma_{12}^2 \\
* & \sigma_{12}^2 & \sigma_1^2\sigma_2^2 & \sigma_{12}\sigma_2^2 \\
* & * & \sigma_{12}^2 & \sigma_{12}\sigma_2^2 \\
* & * & * & \sigma_2^4
\end{bmatrix} \otimes \mathbf{\Delta}_1 + \begin{bmatrix}
\sigma_1^4 & \sigma_1^2\sigma_{12} & \sigma_1^2\sigma_{12} & \sigma_{12}^2 \\
* & \sigma_1^2\sigma_2^2 & \sigma_{12}^2 & \sigma_{12}\sigma_2^2 \\
* & * & \sigma_1^2\sigma_2^2 & \sigma_{12}\sigma_2^2 \\
* & * & * & \sigma_2^4
\end{bmatrix} \otimes \mathbf{\Delta}_2,
\end{aligned}
$$

where $\boldsymbol{\omega} = [\mathrm{vec}_D(\mathbf{\Xi}_1), \cdots, \mathrm{vec}_D(\mathbf{\Xi}_m)]$ and

$$
\mathbf{\Delta}_1 = \begin{bmatrix}
\mathrm{tr}(\mathbf{\Xi}_1\mathbf{\Xi}_1) & \cdots & \mathrm{tr}(\mathbf{\Xi}_1\mathbf{\Xi}_m) \\
\vdots & \ddots & \vdots \\
\mathrm{tr}(\mathbf{\Xi}_m\mathbf{\Xi}_1) & \cdots & \mathrm{tr}(\mathbf{\Xi}_m\mathbf{\Xi}_m)
\end{bmatrix} \quad \text{and} \quad \mathbf{\Delta}_2 = \begin{bmatrix}
\mathrm{tr}(\mathbf{\Xi}_1'\mathbf{\Xi}_1) & \cdots & \mathrm{tr}(\mathbf{\Xi}_1'\mathbf{\Xi}_m) \\
\vdots & \ddots & \vdots \\
\mathrm{tr}(\mathbf{\Xi}_m'\mathbf{\Xi}_1) & \cdots & \mathrm{tr}(\mathbf{\Xi}_m'\mathbf{\Xi}_m)
\end{bmatrix}.
$$

Let

$$
\mathbf{D} = -\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta}_0)] = [\mathbf{D}_1', \mathbf{D}_2']', \tag{10}
$$

where

$$
\mathbf{D}_1 = -\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_1(\boldsymbol{\theta}_0)] = \begin{bmatrix}
\mathbf{Q}'(\mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0) & \mathbf{Q}'\mathbf{X}\boldsymbol{\gamma}_0 & \mathbf{Q}'\mathbf{X}_1 & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}'\mathbf{X}
\end{bmatrix}
$$

and

$$\mathbf{D}_2 = -\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_2(\boldsymbol{\theta}_0)] = \begin{bmatrix} (\sigma_1^2+\phi_0\sigma_{12})\mathrm{tr}(\boldsymbol{\Xi}_1^{(s)}\mathbf{G}) & \mathbf{0}_{1\times(K_X+K_1+1)} \\ \vdots & \vdots \\ (\sigma_1^2+\phi_0\sigma_{12})\mathrm{tr}(\boldsymbol{\Xi}_m^{(s)}\mathbf{G}) & \mathbf{0}_{1\times(K_X+K_1+1)} \\ (\sigma_{12}+\phi_0\sigma_2^2)\mathrm{tr}(\boldsymbol{\Xi}_1'\mathbf{G}) & \mathbf{0}_{1\times(K_X+K_1+1)} \\ \vdots & \vdots \\ (\sigma_{12}+\phi_0\sigma_2^2)\mathrm{tr}(\boldsymbol{\Xi}_m'\mathbf{G}) & \mathbf{0}_{1\times(K_X+K_1+1)} \\ (\sigma_{12}+\phi_0\sigma_2^2)\mathrm{tr}(\boldsymbol{\Xi}_1\mathbf{G}) & \mathbf{0}_{1\times(K_X+K_1+1)} \\ \vdots & \vdots \\ (\sigma_{12}+\phi_0\sigma_2^2)\mathrm{tr}(\boldsymbol{\Xi}_m\mathbf{G}) & \mathbf{0}_{1\times(K_X+K_1+1)} \\ \mathbf{0}_{m\times 1} & \mathbf{0}_{m\times(K_X+K_1+1)} \end{bmatrix}.$$

The following proposition establishes the consistency and asymptotic normality of the GMM estimator.

Proposition 1 *Suppose Assumptions 1-6 hold. Then $\widetilde{\boldsymbol{\theta}}_{gmm}$ defined in (7) is a consistent estimator of $\boldsymbol{\theta}_0$ and has the following asymptotic distribution*

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} \mathrm{N}(\mathbf{0}, \mathrm{AsyVar}(\widetilde{\boldsymbol{\theta}}_{gmm}))$$

where

$$\mathrm{AsyVar}(\widetilde{\boldsymbol{\theta}}_{gmm}) = \lim_{n\to\infty}[(n^{-1}\mathbf{D})'\mathbf{F}'\mathbf{F}(n^{-1}\mathbf{D})]^{-1}(n^{-1}\mathbf{D})'\mathbf{F}'\mathbf{F}(n^{-1}\boldsymbol{\Omega})\mathbf{F}'\mathbf{F}(n^{-1}\mathbf{D})[(n^{-1}\mathbf{D})'\mathbf{F}'\mathbf{F}(n^{-1}\mathbf{D})]^{-1}$$

with $\boldsymbol{\Omega}$ and $\mathbf{D}$ defined in (9) and (10) respectively.

Close inspection of $\mathrm{AsyVar}(\widetilde{\boldsymbol{\theta}}_{gmm})$ reveals that the optimal $\mathbf{F}'\mathbf{F}$ is $(n^{-1}\boldsymbol{\Omega})^{-1}$ by the generalized Schwarz inequality. The following proposition establishes the consistency and asymptotic normality of the GMM estimator with the estimated optimal weighting matrix. It also suggests a over-identifying restrictions (OIR) test based on the proposed GMM estimator.

Proposition 2 *Suppose Assumptions 1-6 hold and $n^{-1}\widehat{\boldsymbol{\Omega}}$ is a consistent estimator of $n^{-1}\boldsymbol{\Omega}$ defined in (9). Then,*

$$\widehat{\boldsymbol{\theta}}_{gmm} = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbf{g}(\boldsymbol{\theta})'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{g}(\boldsymbol{\theta}) \tag{11}$$

is a consistent estimator of $\boldsymbol{\theta}_0$ *and*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} \mathrm{N}(\mathbf{0}, [\lim_{n\to\infty} n^{-1}\mathbf{D}'\boldsymbol{\Omega}^{-1}\mathbf{D}]^{-1}),$$

where $\mathbf{D}$ *is defined in (10). Furthermore*

$$\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm})'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm}) \xrightarrow{\mathrm{d}} \chi^2_{\dim(\mathbf{g})-\dim(\boldsymbol{\theta})}.$$

4.2 Asymptotic Efficiency

When only the linear moment function $\mathbf{g}_1(\boldsymbol{\theta}_0)$ is used for the GMM estimation, the GMM estimator defined in (11) reduces to the generalized spatial 3SLS in Kelejian and Prucha (2004) because

$$\widehat{\boldsymbol{\theta}}_{3SLS} = \arg\min \mathbf{g}_1(\boldsymbol{\theta})'\widehat{\boldsymbol{\Omega}}_{11}^{-1}\mathbf{g}_1(\boldsymbol{\theta}) = \arg\min \mathbf{u}(\boldsymbol{\theta})'(\widehat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{P})\mathbf{u}(\boldsymbol{\theta}) = [\mathbf{Z}'(\widehat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{P})\mathbf{Z}]^{-1}\mathbf{Z}'(\widehat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{P})\mathbf{y},$$

where $\mathbf{P} = \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2')'$,

$$\mathbf{Z} = \begin{bmatrix} \mathbf{W}\mathbf{y}_1 & \mathbf{y}_2 & \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{X} \end{bmatrix},$$

and $\widehat{\boldsymbol{\Sigma}}$ is a consistent estimator of $\boldsymbol{\Sigma}$. It follows from Proposition 2 that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{3SLS} - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} \mathrm{N}(\mathbf{0}, [\lim_{n\to\infty} n^{-1}\mathbf{D}_1'\boldsymbol{\Omega}_{11}^{-1}\mathbf{D}_1]^{-1}).$$

As

$$\mathbf{D}'\boldsymbol{\Omega}^{-1}\mathbf{D} - \mathbf{D}_1'\boldsymbol{\Omega}_{11}^{-1}\mathbf{D}_1 = (\mathbf{D}_2 - \boldsymbol{\Omega}_{12}'\boldsymbol{\Omega}_{11}^{-1}\mathbf{D}_1)'(\boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{12}'\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12})^{-1}(\mathbf{D}_2 - \boldsymbol{\Omega}_{12}'\boldsymbol{\Omega}_{11}^{-1}\mathbf{D}_1),$$

which is positive semi-definite, the proposed GMM estimator is asymptotically more efficient than the 3SLS estimator.

The asymptotic efficiency of the proposed GMM estimator depends on the choices of $\mathbf{Q}$ and $\boldsymbol{\Xi}_1, \cdots, \boldsymbol{\Xi}_m$. Following Lee (2007), our discussion on the asymptotic efficiency focuses on two cases: (i) $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}\otimes\mathbf{I}_n)$, and (ii) $\boldsymbol{\Xi}_j$ has a zero diagonal for all $j = 1, \cdots, m$. Let $\mathcal{P}$ be a subset of all $\boldsymbol{\Xi}$'s satisfying Assumption 5 such that $\mathrm{diag}(\boldsymbol{\Xi}) = \mathbf{0}$ for all $\boldsymbol{\Xi} \in \mathcal{P}$. The sub-class of quadratic moment functions using $\mathbf{\Xi} \in \mathcal{P}$ is of a particular interest because these quadratic moment functions could be robust against unknown form of heteroskedasticity as shown in Lin and Lee (2010).

Let

$$\mathbf{g}^*(\boldsymbol{\theta}) = [\mathbf{g}_1^*(\boldsymbol{\delta})', \mathbf{g}_2^*(\boldsymbol{\theta})']', \tag{12}$$

where $\mathbf{g}_1^*(\boldsymbol{\delta}) = (\mathbf{I}_2 \otimes \mathbf{Q}^*)'\mathbf{u}(\boldsymbol{\theta})$ and

$$\mathbf{g}_2^*(\boldsymbol{\theta}) = [\mathbf{u}_1(\boldsymbol{\delta})'\mathbf{\Xi}^*\mathbf{u}_1(\boldsymbol{\delta}), \mathbf{u}_1(\boldsymbol{\delta})'\mathbf{\Xi}^*\mathbf{u}_2(\gamma), \mathbf{u}_2(\gamma)'\mathbf{\Xi}^*\mathbf{u}_1(\boldsymbol{\delta}), \mathbf{u}_2(\gamma)'\mathbf{\Xi}^*\mathbf{u}_2(\gamma)]'.$$

In cases (i) and (ii),

$$\mathbf{\Omega}^* = \mathrm{Var}[\mathbf{g}^*(\boldsymbol{\theta}_0)] = \begin{bmatrix} \mathbf{\Sigma} \otimes (\mathbf{Q}^{*\prime}\mathbf{Q}^*) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_{22}^* \end{bmatrix} \tag{13}$$

where

$$\mathbf{\Omega}_{22}^* = \begin{bmatrix} \sigma_1^4 & \sigma_1^2\sigma_{12} & \sigma_1^2\sigma_{12} & \sigma_{12}^2 \\ * & \sigma_{12}^2 & \sigma_1^2\sigma_2^2 & \sigma_{12}\sigma_2^2 \\ * & * & \sigma_{12}^2 & \sigma_{12}\sigma_2^2 \\ * & * & * & \sigma_2^4 \end{bmatrix} \otimes \mathrm{tr}(\mathbf{\Xi}^*\mathbf{\Xi}^*) + \begin{bmatrix} \sigma_1^4 & \sigma_1^2\sigma_{12} & \sigma_1^2\sigma_{12} & \sigma_{12}^2 \\ * & \sigma_1^2\sigma_2^2 & \sigma_{12}^2 & \sigma_{12}\sigma_2^2 \\ * & * & \sigma_1^2\sigma_2^2 & \sigma_{12}\sigma_2^2 \\ * & * & * & \sigma_2^4 \end{bmatrix} \otimes \mathrm{tr}(\mathbf{\Xi}^{*\prime}\mathbf{\Xi}^*).$$

The following proposition gives the infeasible best GMM (BGMM) estimator

$$\widetilde{\boldsymbol{\theta}}_{bgmm} = \arg\min\nolimits_{\boldsymbol{\theta}\in\Theta} \mathbf{g}^*(\boldsymbol{\theta})'\mathbf{\Omega}^{*-1}\mathbf{g}^*(\boldsymbol{\theta}) \tag{14}$$

with the optimal $\mathbf{Q}^*$ and $\mathbf{\Xi}^*$ in cases (i) and (ii) respectively.

Proposition 3 *Suppose Assumptions 1-6 hold. Let* $\mathbf{G} = \mathbf{W}\mathbf{S}^{-1}$.

(i) *Suppose* $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}_n)$. *The BGMM estimator defined in (14) with* $\mathbf{Q}^* = [\mathbf{G}\mathbf{X}, \mathbf{X}]$ *and* $\mathbf{\Xi}^* = \mathbf{G} - n^{-1}\mathrm{tr}(\mathbf{G})\mathbf{I}_n$ *is the most efficient one in the class of GMM estimators defined in (7).*

(ii) *Without the normality assumption on* $\mathbf{u}$, *the BGMM estimator defined in (14) with* $\mathbf{Q}^* = [\mathbf{G}\mathbf{X}, \mathbf{X}]$ *and* $\mathbf{\Xi}^* = \mathbf{G} - \mathrm{diag}(\mathbf{G})$ *is the most efficient one in the sub-class of GMM estimators defined in (7) with* $\mathbf{\Xi}_j \in \mathcal{P}$ *for all* $j = 1, \cdots, m$.

Under normality, the model can be efficiently estimated by the ML estimator. To get some intuition of the optimal $\mathbf{Q}^*$ and $\mathbf{\Xi}^*$ in case (i), we compare the linear and quadratic moment functions utilized by the GMM estimator with the first order partial derivatives of the log likelihood function. Let $\mathbf{G}(\lambda) = \mathbf{W}\mathbf{S}(\lambda)^{-1}$, where $\mathbf{S}(\lambda) = \mathbf{I}_n - \lambda\mathbf{W}$. The log likelihood function based on the joint normal distribution of $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2')'$ is[9]

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = -n\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}\otimes\mathbf{I}_n| + \ln|\mathbf{S}(\lambda)| - \frac{1}{2}\mathbf{u}(\boldsymbol{\theta})'(\boldsymbol{\Sigma}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \tag{15}$$

with the first order partial derivatives

$$\begin{aligned}
\frac{\partial}{\partial\lambda}L(\boldsymbol{\theta}, \widehat{\boldsymbol{\Sigma}}) \;=\;& [(\phi\mathbf{X}\boldsymbol{\gamma} + \mathbf{X}_1\boldsymbol{\beta})'\mathbf{G}(\lambda)', \mathbf{0}_{1\times n}]'(\widehat{\boldsymbol{\Sigma}}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \\
&+\frac{\widehat{\sigma}_2^2}{|\widehat{\boldsymbol{\Sigma}}|}\mathbf{u}_1(\boldsymbol{\delta})'[\mathbf{G}(\lambda) - n^{-1}\mathrm{tr}(\mathbf{G}(\lambda))\mathbf{I}_n]\mathbf{u}_1(\boldsymbol{\delta}) - \frac{\widehat{\sigma}_{12}^2}{|\widehat{\boldsymbol{\Sigma}}|}\mathbf{u}_2(\boldsymbol{\gamma})'[\mathbf{G}(\lambda) - n^{-1}\mathrm{tr}(\mathbf{G}(\lambda))\mathbf{I}_n]\mathbf{u}_1(\boldsymbol{\delta}) \\
&+\phi\frac{\widehat{\sigma}_2^2}{|\widehat{\boldsymbol{\Sigma}}|}\mathbf{u}_1(\boldsymbol{\delta})'[\mathbf{G}(\lambda) - n^{-1}\mathrm{tr}(\mathbf{G}(\lambda))\mathbf{I}_n]\mathbf{u}_2(\boldsymbol{\gamma}) - \phi\frac{\widehat{\sigma}_{12}}{|\widehat{\boldsymbol{\Sigma}}|}\mathbf{u}_2(\boldsymbol{\gamma})'[\mathbf{G}(\lambda) - n^{-1}\mathrm{tr}(\mathbf{G}(\lambda))\mathbf{I}_n]\mathbf{u}_2(\boldsymbol{\gamma})
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial\phi}L(\boldsymbol{\theta}, \widehat{\boldsymbol{\Sigma}}) &= [\boldsymbol{\gamma}'\mathbf{X}', \mathbf{0}_{1\times n}](\widehat{\boldsymbol{\Sigma}}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \\
\frac{\partial}{\partial\boldsymbol{\beta}}L(\boldsymbol{\theta}, \widehat{\boldsymbol{\Sigma}}) &= [\mathbf{X}_1', \mathbf{0}_{K_1\times n}](\widehat{\boldsymbol{\Sigma}}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \\
\frac{\partial}{\partial\boldsymbol{\gamma}}L(\boldsymbol{\theta}, \widehat{\boldsymbol{\Sigma}}) &= [\mathbf{0}_{K_X\times n}, \mathbf{X}'](\widehat{\boldsymbol{\Sigma}}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta})
\end{aligned}$$

where $\widehat{\boldsymbol{\Sigma}}$ is the ML estimator for $\boldsymbol{\Sigma}$ given by

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} \widehat{\sigma}_1^2 & \widehat{\sigma}_{12} \\ \widehat{\sigma}_{12} & \widehat{\sigma}_2^2 \end{bmatrix} = n^{-1}\begin{bmatrix} \mathbf{u}_1(\boldsymbol{\delta})'\mathbf{u}_1(\boldsymbol{\delta}) & \mathbf{u}_1(\boldsymbol{\delta})'\mathbf{u}_2(\boldsymbol{\gamma}) \\ \mathbf{u}_1(\boldsymbol{\delta})'\mathbf{u}_2(\boldsymbol{\gamma}) & \mathbf{u}_2(\boldsymbol{\gamma})'\mathbf{u}_2(\boldsymbol{\gamma}) \end{bmatrix}.$$

Close inspection reveals the similarity between the ML and BGMM estimators under normality, as the first order partial derivatives of the log likelihood function can be considered as linear combinations of the moment functions $\mathbf{Q}^*(\lambda)'\mathbf{u}_1(\boldsymbol{\delta})$, $\mathbf{Q}^*(\lambda)'\mathbf{u}_2(\boldsymbol{\gamma})$, $\mathbf{u}_1(\boldsymbol{\delta})'\mathbf{\Xi}^*(\lambda)\mathbf{u}_1(\boldsymbol{\delta})$, $\mathbf{u}_1(\boldsymbol{\delta})'\mathbf{\Xi}^*(\lambda)\mathbf{u}_2(\boldsymbol{\gamma})$, $\mathbf{u}_2(\boldsymbol{\gamma})'\mathbf{\Xi}^*(\lambda)\mathbf{u}_1(\boldsymbol{\delta})$, and $\mathbf{u}_2(\boldsymbol{\gamma})'\mathbf{\Xi}^*(\lambda)\mathbf{u}_2(\boldsymbol{\gamma})$ with $\mathbf{Q}^*(\lambda) = [\mathbf{G}(\lambda)\mathbf{X}, \mathbf{X}]$ and $\mathbf{\Xi}^*(\lambda) = \mathbf{G}(\lambda) - n^{-1}\mathrm{tr}(\mathbf{G}(\lambda))\mathbf{I}_n$.

The optimal $\mathbf{Q}^*$ and $\mathbf{\Xi}^*$ are not feasible as $\mathbf{G}$ involves the unknown parameter λ_0. Suppose there exists a $\sqrt{n}$-consistent preliminary estimator $\widehat{\lambda}$ for λ_0 (say, the 2SLS estimator with IV matrix

[9] The detailed derivation of the log likelihood function and its partial derivatives can be found in Appendix A.

$\mathbf{Q} = [\mathbf{WX}, \mathbf{X}]$). Then, the feasible optimal $\widehat{\mathbf{Q}}^*$ and $\widehat{\boldsymbol{\Xi}}^*$ can obtained by replacing λ_0 in $\mathbf{Q}^*$ and $\boldsymbol{\Xi}^*$ by $\widehat{\lambda}$. Furthermore, suppose $\widehat{\sigma}_1^2, \widehat{\sigma}_{12}, \widehat{\sigma}_2^2$ are consistent preliminary estimators for $\sigma_1^2, \sigma_{12}, \sigma_2^2$. Then, $n^{-1}\widehat{\boldsymbol{\Omega}}^*$ is a consistent estimator of $n^{-1}\boldsymbol{\Omega}^*$ defined in (13) with the unknown parameters $\lambda, \sigma_1^2, \sigma_{12}, \sigma_2^2$ in $\boldsymbol{\Omega}^*$ replaced by $\widehat{\lambda}, \widehat{\sigma}_1^2, \widehat{\sigma}_{12}, \widehat{\sigma}_2^2$. Then, the feasible BGMM estimator is given by

$$\widehat{\boldsymbol{\theta}}_{bgmm} = \arg\min\nolimits_{\boldsymbol{\theta}\in\Theta} \widehat{\mathbf{g}}^*(\boldsymbol{\theta})'\widehat{\boldsymbol{\Omega}}^{*-1}\widehat{\mathbf{g}}^*(\boldsymbol{\theta}), \tag{16}$$

where $\widehat{\mathbf{g}}^*(\boldsymbol{\theta})$ is obtained by replacing $\mathbf{Q}^*$ and $\boldsymbol{\Xi}^*$ in $\mathbf{g}^*(\boldsymbol{\theta})$ with $\widehat{\mathbf{Q}}^*$ and $\widehat{\boldsymbol{\Xi}}^*$. Following a similar argument in the proof of Proposition 3 in Lee (2007), the feasible BGMM estimator $\widehat{\boldsymbol{\theta}}_{bgmm}$ can be shown to have the same limiting distribution as its infeasible counterpart $\widetilde{\boldsymbol{\theta}}_{bgmm}$.

Proposition 4 *Suppose Assumptions 1-6 hold, $\widehat{\lambda}$ is a $\sqrt{n}$-consistent estimator of λ_0, and $\widehat{\boldsymbol{\Sigma}}$ is a consistent estimator of $\boldsymbol{\Sigma}$. The feasible BGMM estimator $\widehat{\boldsymbol{\theta}}_{bgmm}$ defined in (16) is asymptotically equivalent to the corresponding infeasible BGMM estimator $\widetilde{\boldsymbol{\theta}}_{bgmm}$.*

Under Assumption 3, $\mathbf{G} = \mathbf{WS}^{-1} = \mathbf{W} + \lambda_0\mathbf{W}^2 + \lambda_0^2\mathbf{W}^3 + \cdots$. Thus, $\mathbf{G}$ can be approximated by the leading order terms of the series expansion, i.e. $\mathbf{W}, \mathbf{W}^2, \mathbf{W}^3, \cdots$. Therefore, a convenient alternative to the BGMM estimator under normality for empirical researchers would be the GMM estimator with $\mathbf{Q} = [\mathbf{WX}, \cdots, \mathbf{W}^m\mathbf{X}, \mathbf{X}]$ and $\boldsymbol{\Xi}_1 = \mathbf{W}, \boldsymbol{\Xi}_2 = \mathbf{W}^2 - n^{-1}\mathrm{tr}(\mathbf{W}^2)\mathbf{I}_n, \cdots, \boldsymbol{\Xi}_m = \mathbf{W}^m - n^{-1}\mathrm{tr}(\mathbf{W}^m)\mathbf{I}_n$, for some fixed m.

5 Monte Carlo Experiments

We conduct a small Monte Carlo simulation experiment to study the finite sample performance of the proposed GMM estimator based on the following model

$$\begin{aligned} \mathbf{y}_1 &= \lambda_0\mathbf{W}\mathbf{y}_1 + \phi_0\mathbf{y}_2 + \beta_0\mathbf{x}_1 + \mathbf{u}_1 \\ \mathbf{y}_2 &= \gamma_0\mathbf{x}_2 + \mathbf{u}_2. \end{aligned}$$

In the DGP, we set $\lambda_0 = 0.6$ and $\gamma_0 = 1$, and generate $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$ as $\mathbf{x}_1 \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_n)$, $\mathbf{x}_2 \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_n)$, and $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$, where

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{bmatrix}.$$

We conduct 1000 replications in the simulation experiment for different specifications with $n \in \{245, 490\}$, $\sigma_{12} \in \{0.1, 0.5, 0.9\}$, and $(\phi_0, \beta_0) \in \{(0.5, 0.5), (0.2, 0.2)\}$. From the reduce form equation (4), $\mathrm{E}(\mathbf{Wy}_1) = \mathbf{Gx}_1\beta_0 + \phi_0\mathbf{Gx}_2\gamma_0$. Therefore, $\phi_0 = \beta_0 = 0.5$ corresponds to the case that the IVs based on $\mathrm{E}(\mathbf{Wy}_1)$ are informative and $\phi_0 = \beta_0 = 0.2$ corresponds to the case that the IVs based on $\mathrm{E}(\mathbf{Wy}_1)$ are less informative. Let $\mathbf{W}_0$ denote the spatial weights matrix for the study of crimes across 49 districts in Columbus, Ohio, in Anselin (1988). For $n = 245$, we set $\mathbf{W} = \mathbf{I}_5 \otimes \mathbf{W}_0$, and for $n = 490$, we set $\mathbf{W} = \mathbf{I}_{10} \otimes \mathbf{W}_0$. Let $\widehat{\mathbf{G}} = \mathbf{W}(\mathbf{I}_n - \widehat{\lambda}\mathbf{W})^{-1}$, where $\widehat{\lambda}$ is the 2SLS estimator of λ_0 using the IV matrix $\mathbf{Q} = [\mathbf{WX}, \mathbf{W}^2\mathbf{X}, \mathbf{X}]$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2]$. Let $\widehat{\mathbf{Q}} = [\widehat{\mathbf{G}}\mathbf{X}, \mathbf{X}]$. In the experiment, we consider the following estimators.

(a) The 2SLS estimator of equation (1) with the linear moment function $\widehat{\mathbf{Q}}'\mathbf{u}_1(\boldsymbol{\delta})$.

(b) The 3SLS estimator of equations (1) and (2) with the linear moment function $(\mathbf{I}_2 \otimes \widehat{\mathbf{Q}})'\mathbf{u}(\boldsymbol{\theta})$.

(c) The single-equation GMM (GMM-1) estimator of equation (1) with the linear moment function $\widehat{\mathbf{Q}}'\mathbf{u}_1(\boldsymbol{\delta})$ and the quadratic moment function $\mathbf{u}_1(\boldsymbol{\delta})'[\widehat{\mathbf{G}} - n^{-1}\mathrm{tr}(\widehat{\mathbf{G}})\mathbf{I}_n]\mathbf{u}_1(\boldsymbol{\delta})$.

(d) The system GMM (GMM-2) estimator of equations (1) and (2) with the linear moment function $(\mathbf{I}_2 \otimes \widehat{\mathbf{Q}})'\mathbf{u}(\boldsymbol{\theta})$ and the quadratic moment functions $\mathbf{u}_1(\boldsymbol{\delta})'[\widehat{\mathbf{G}} - n^{-1}\mathrm{tr}(\widehat{\mathbf{G}})\mathbf{I}_n]\mathbf{u}_1(\boldsymbol{\delta})$, $\mathbf{u}_1(\boldsymbol{\delta})'[\widehat{\mathbf{G}} - n^{-1}\mathrm{tr}(\widehat{\mathbf{G}})\mathbf{I}_n]\mathbf{u}_2(\gamma)$, $\mathbf{u}_2(\gamma)'[\widehat{\mathbf{G}} - n^{-1}\mathrm{tr}(\widehat{\mathbf{G}})\mathbf{I}_n]\mathbf{u}_1(\boldsymbol{\delta})$, and $\mathbf{u}_2(\gamma)'[\widehat{\mathbf{G}} - n^{-1}\mathrm{tr}(\widehat{\mathbf{G}})\mathbf{I}_n]\mathbf{u}_2(\gamma)$.

Although the 2SLS estimator and the single-equation GMM estimator only use "limited information" in equation (1) and thus may not be as efficient as their counterparts (i.e. the 3SLS estimator and the system GMM estimator respectively) that use "full information" in the whole system, these estimators require weaker assumptions on the reduced form equation (2) and thus may be desirable under certain circumstances. The estimation results are reported in Tables 2-5. We report the mean and standard deviation (SD) of the empirical distributions of the estimates. To facilitate the comparison of different estimators, we also report their root mean square errors (RMSE). The main observations from the experiment are summarized as follows.

[Insert Tables 2-5 here]

(i) The 2SLS and 3SLS estimators of λ_0 are upwards biased with large SDs when the IVs for $\mathbf{Wy}_1$ are less informative. For example, when $n = 245$ and $\sigma_{12} = 0.1$, the 2SLS and 3SLS estimates of λ_0 reported in Table 4 are upwards biased by about 10%. The biases and SDs reduce as sample size increases. The 3SLS estimators of λ_0 and β_0 perform better as σ_{12} increases.

(ii) The single-equation GMM (GMM-1) estimator of λ_0 is upwards biased when the IVs for $\mathbf{Wy}_1$ are less informative. When $n = 245$ and $\sigma_{12} = 0.1$, the GMM-1 estimates of λ_0 reported in Table 4 are upwards biased by about 6%. The bias reduces as sample size increases. The GMM-1 estimator of λ_0 reduces the SD of the 2SLS estimator. The SD reduction is more significant when the IVs for $\mathbf{Wy}_1$ are less informative. In Table 2, when $\sigma_{12} = 0.1$, the GMM-1 estimator reduces the SD of the 2SLS estimator by about 60%. In Table 4, when $\sigma_{12} = 0.1$, the GMM-1 estimator reduces the SD of the 2SLS estimator by about 65%.

(iii) The system GMM (GMM-2) estimator of λ_0 is upwards biased when the sample size is moderate ($n = 245$) and the IVs for $\mathbf{Wy}_1$ are less informative. The bias reduces as σ_{12} increases. When $n = 490$, the GMM-2 estimator is essentially unbiased even if the IVs are weak. The GMM-2 estimators of λ_0 and β_0 have smaller SDs than the corresponding GMM-1 estimators. The reduction in the SD is more significant when the endogeneity problem is more severe (i.e. σ_{12} is larger) and/or the IVs for $\mathbf{Wy}_1$ are less informative. For example, in Table 3, when $\sigma_{12} = 0.9$, the GMM-2 estimator of λ_0 reduces the SD of the GMM-1 estimator by about 42%. In Table 5, when $\sigma_{12} = 0.9$, the GMM-2 estimator of λ_0 reduces the SD of the GMM-1 estimator by about 75%. In both cases, the GMM-2 estimator of β_0 reduces the SD of the corresponding GMM-1 estimator by about 56%.

6 Conclusion

In this paper, we propose a general GMM framework for the estimation of SAR models with endogenous regressors. We introduce a new set of quadratic moment conditions to construct the GMM estimator, based on the correlation structure of the spatially lagged dependent variable with the model disturbance term and with the endogenous regressor. We establish the consistency and asymptotic normality of the proposed GMM estimator and discuss the optimal choice of moment conditions. We also conduct a Monte Carlo experiment to show the GMM estimator works well in finite samples.

The proposed GMM estimator utilizes correlation across equations (1) and (2) to construct moment equations and thus can be considered as a "full information" estimator. If we only use the moment equations based on $\mathbf{u}_1(\boldsymbol{\delta})$, i.e., the residual function of equation (1), the proposed GMM estimator becomes a single-equation GMM estimator. Although the single-equation GMM estimator may not be as efficient as the "full information" GMM estimator, the single-equation GMM estimator requires weaker assumptions on the reduced form equation (2) and thus may be desirable under certain circumstances. The Monte Carlo experiment shows that the "full information" GMM estimator improves the efficiency of the single-equation GMM estimator when the endogeneity problem is severe and/or the IVs for the spatially lagged dependent variable are weak.

References

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

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

Carrell, S. E., Sacerdote, B. I. and West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation, *Econometrica* **81**: 855–882.

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

Cliff, A. and Ord, J. (1981). *Spatial Processes, Models and Applications*, Pion, London.

Denbee, E., Julliard, C., Li, Y. and Yuan, K. (2014). Network risk and key players: A structural analysis of interbank liquidity. AXA working paper series NO 12, FMG discussion paper 734.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica* **50**: 1029–1054.

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

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

Kelejian, H. H. and Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, *Journal of Econometrics* **157**: 53–67.

König, M., Liu, X. and Zenou, Y. (2014). R&D networks: Theory, empirics and policy implications. CEPR discussion paper no. 9872.

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.

Lee, L. F. and Liu, X. (2010). Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances, *Econometric Theory* **26**: 187–230.

Lin, X. (2010). Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables, *Journal of Labor Economics* **28**: 825–860.

Lin, X. and Lee, L. F. (2010). Gmm estimation of spatial autoregressive models with unknown heteroskedasticity, *Journal of Econometrics* **157**: 34–52.

Lindquist, M. J. and Zenou, Y. (2014). Key players in co-offending networks. Working paper, Stockholm University.

Liu, X. (2012). On the consistency of the LIML estimator of a spatial autoregressive model with many instruments, *Economics Letters* **116**: 472–475.

Liu, X. and Lee, L. F. (2013). Two stage least squares estimation of spatial autoregressive models with endogenous regressors and many instruments, *Econometric Reviews* **32**: 734–753.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing, *in* D. McFadden and R. F. Engle (eds), *Handbook of Econometrics*, Vol. IV, Elsevier, North-Holland, pp. 2111–2245.

Patacchini, E. and Zenou, Y. (2012). Juvenile delinquency and conformism, *Journal of Law, Economics, and Organization* **28**: 1–31.

Peracchi, F. (2001). *Econometrics*, John Wiley & Sons, New York.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far?, *in* E. A. Hanushek, S. Machin and L. Woessmann (eds), *Handbook of Economics of Education*, Vol. 3, Elsevier, pp. 249–277.

Yang, K. and Lee, L. F. (2014). Identification and QML estimation of multivariate and simultaneous spatial autoregressive models. Working paper, The Ohio State University.

A Likelihood Function of the SAR Model with Endogenous Regressors

Let

$$\boldsymbol{\mu}_y(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{S}^{-1}(\lambda)(\phi\mathbf{X}\boldsymbol{\gamma} + \mathbf{X}_1\boldsymbol{\beta}) \\ \mathbf{X}\boldsymbol{\gamma} \end{bmatrix} \quad \text{and} \quad \mathbf{R}(\phi, \lambda) = \begin{bmatrix} \mathbf{S}^{-1}(\lambda) & \phi\mathbf{S}^{-1}(\lambda) \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix},$$

where $\mathbf{S}(\lambda) = \mathbf{I}_n - \lambda\mathbf{W}$. From the reduced form equations (2) and (3), $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2')' = \boldsymbol{\mu}_y(\boldsymbol{\theta}_0) + \mathbf{R}(\phi_0, \lambda_0)\mathbf{u}$ where $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$. Under normality, $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$, and thus $\mathbf{y} \sim \mathrm{N}(\boldsymbol{\mu}_y, \mathbf{R}(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)\mathbf{R}')$, where $\boldsymbol{\mu}_y = \boldsymbol{\mu}_y(\boldsymbol{\theta}_0)$ and $\mathbf{R} = \mathbf{R}(\phi_0, \lambda_0)$. Hence, the log likelihood function of (1) and (2) is given by

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) &= -n\ln(2\pi) - \frac{1}{2}\ln|\mathbf{R}(\phi, \lambda)(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)\mathbf{R}(\phi, \lambda)'| \\ &\quad -\frac{1}{2}[\mathbf{y} - \boldsymbol{\mu}_y(\boldsymbol{\theta})]'[\mathbf{R}(\phi, \lambda)(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)\mathbf{R}(\phi, \lambda)']^{-1}[\mathbf{y} - \boldsymbol{\mu}_y(\boldsymbol{\theta})]. \end{aligned}$$

As $\mathbf{u}(\boldsymbol{\theta}) = \mathbf{R}^{-1}(\phi, \lambda)[\mathbf{y} - \boldsymbol{\mu}_y(\boldsymbol{\theta})]$ and $|\mathbf{R}^{-1}(\phi, \lambda)| = |\mathbf{S}(\lambda)|$. Then, the log likelihood function can be written as

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = -n\ln(2\pi) - \frac{1}{2}\ln|(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)| + \ln|\mathbf{S}(\lambda)| - \frac{1}{2}\mathbf{u}(\boldsymbol{\theta})'(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}).$$

The first order partial derivatives of the log likelihood function are

$$\begin{aligned} \frac{\partial}{\partial\lambda}L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) &= -\mathrm{tr}(\mathbf{G}(\lambda)) + [\mathbf{y}_1'\mathbf{W}', \mathbf{0}](\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \\ \frac{\partial}{\partial\phi}L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) &= [\mathbf{y}_2', \mathbf{0}](\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \\ \frac{\partial}{\partial\boldsymbol{\beta}}L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) &= [\mathbf{X}_1', \mathbf{0}](\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \\ \frac{\partial}{\partial\boldsymbol{\gamma}}L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) &= [\mathbf{0}, \mathbf{X}'](\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \end{aligned}$$

and

$$\frac{\partial}{\partial(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}}L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \frac{1}{2}(\boldsymbol{\Sigma} \otimes \mathbf{I}_n) - \frac{1}{2}\mathbf{u}(\boldsymbol{\theta})\mathbf{u}(\boldsymbol{\theta})', \tag{17}$$

where $\mathbf{G}(\lambda) = \mathbf{W}\mathbf{S}(\lambda)^{-1}$. Since $\mathbf{W}\mathbf{y}_1 = \mathbf{G}(\lambda)(\mathbf{u}_1(\delta) + \phi\mathbf{u}_2(\gamma)) + \mathbf{G}(\lambda)(\phi\mathbf{X}\boldsymbol{\gamma} + \mathbf{X}_1\boldsymbol{\beta})$ and $\mathbf{y}_2 =$

$\mathbf{X}\boldsymbol{\gamma} + \mathbf{u}_2(\boldsymbol{\gamma})$, then

$$\begin{aligned}\frac{\partial}{\partial\lambda}L(\boldsymbol{\theta},\boldsymbol{\Sigma}) &= -\mathrm{tr}(\mathbf{G}(\lambda)) + [(\phi\mathbf{X}\boldsymbol{\gamma} + \mathbf{X}_1\boldsymbol{\beta})'\mathbf{G}(\lambda)', \mathbf{0}]'(\boldsymbol{\Sigma}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \\ &\quad + [(\mathbf{u}_1(\boldsymbol{\delta}) + \phi\mathbf{u}_2(\boldsymbol{\gamma}))'\mathbf{G}(\lambda)', \mathbf{0}]'(\boldsymbol{\Sigma}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta})\end{aligned} \tag{18}$$

and

$$\frac{\partial}{\partial\phi}L(\boldsymbol{\theta},\boldsymbol{\Sigma}) = [\boldsymbol{\gamma}'\mathbf{X}', \mathbf{0}](\boldsymbol{\Sigma}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) + [\mathbf{u}_2(\boldsymbol{\gamma})', \mathbf{0}](\boldsymbol{\Sigma}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}). \tag{19}$$

From (17), the ML estimator for $\boldsymbol{\Sigma}$ is given by

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} \widehat{\sigma}_1^2 & \widehat{\sigma}_{12} \\ \widehat{\sigma}_{12} & \widehat{\sigma}_2^2 \end{bmatrix} = n^{-1}\begin{bmatrix} \mathbf{u}_1(\boldsymbol{\delta})'\mathbf{u}_1(\boldsymbol{\delta}) & \mathbf{u}_1(\boldsymbol{\delta})'\mathbf{u}_2(\boldsymbol{\gamma}) \\ \mathbf{u}_1(\boldsymbol{\delta})'\mathbf{u}_2(\boldsymbol{\gamma}) & \mathbf{u}_2(\boldsymbol{\gamma})'\mathbf{u}_2(\boldsymbol{\gamma}) \end{bmatrix}.$$

Substitution of $\widehat{\boldsymbol{\Sigma}}$ into (18) and (19) gives

$$\begin{aligned}\frac{\partial}{\partial\lambda}L(\boldsymbol{\theta},\widehat{\boldsymbol{\Sigma}}) &= [(\phi\mathbf{X}\boldsymbol{\gamma} + \mathbf{X}_1\boldsymbol{\beta})'\mathbf{G}(\lambda)', \mathbf{0}]'(\widehat{\boldsymbol{\Sigma}}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \\ &\quad + \frac{\widehat{\sigma}_2^2}{|\widehat{\boldsymbol{\Sigma}}|}\mathbf{u}_1(\boldsymbol{\delta})'[\mathbf{G}(\lambda) - n^{-1}\mathrm{tr}(\mathbf{G}(\lambda))\mathbf{I}_n]\mathbf{u}_1(\boldsymbol{\delta}) - \frac{\widehat{\sigma}_{12}^2}{|\widehat{\boldsymbol{\Sigma}}|}\mathbf{u}_2(\boldsymbol{\gamma})'[\mathbf{G}(\lambda) - n^{-1}\mathrm{tr}(\mathbf{G}(\lambda))\mathbf{I}_n]\mathbf{u}_1(\boldsymbol{\delta}) \\ &\quad + \phi\frac{\widehat{\sigma}_2^2}{|\widehat{\boldsymbol{\Sigma}}|}\mathbf{u}_1(\boldsymbol{\delta})'[\mathbf{G}(\lambda) - n^{-1}\mathrm{tr}(\mathbf{G}(\lambda))\mathbf{I}_n]\mathbf{u}_2(\boldsymbol{\gamma}) - \phi\frac{\widehat{\sigma}_{12}}{|\widehat{\boldsymbol{\Sigma}}|}\mathbf{u}_2(\boldsymbol{\gamma})'[\mathbf{G}(\lambda) - n^{-1}\mathrm{tr}(\mathbf{G}(\lambda))\mathbf{I}_n]\mathbf{u}_2(\boldsymbol{\gamma})\end{aligned}$$

and

$$\frac{\partial}{\partial\phi}L(\boldsymbol{\theta},\widehat{\boldsymbol{\Sigma}}) = [\boldsymbol{\gamma}'\mathbf{X}', \mathbf{0}](\widehat{\boldsymbol{\Sigma}}\otimes\mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}).$$

B Lemmas

For ease of reference, we list some useful results without proofs. Lemmas B.1-B.6 can be found (or are straightforward extensions of the lemmas) in Lee (2007). Lemma B.7 is a special case of Lemma 3 in Yang and Lee (2014). Lemmas B.8 and B.9 are from Breusch et al. (1999).

Lemma B.1 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $n \times n$ *nonstochastic matrices such that* $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{B}) = 0$. *Then,*

$$
\begin{aligned}
&\text{(i)} & \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_1\mathbf{u}_1'\mathbf{B}\mathbf{u}_1) &= (\mu_{4,0} - 3\sigma_1^4)\mathrm{vec}_D(\mathbf{A})'\mathrm{vec}_D(\mathbf{B}) + \sigma_1^4\mathrm{tr}(\mathbf{A}\mathbf{B}^{(s)}) \\
&\text{(ii)} & \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_1\mathbf{u}_1'\mathbf{B}\mathbf{u}_2) &= (\mu_{3,1} - 3\sigma_1^2\sigma_{12})\mathrm{vec}_D(\mathbf{A})\mathrm{vec}_D(\mathbf{B}) + \sigma_1^2\sigma_{12}\mathrm{tr}(\mathbf{A}\mathbf{B}^{(s)}) \\
&\text{(iii)} & \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_1\mathbf{u}_2'\mathbf{B}\mathbf{u}_2) &= (\mu_{2,2} - \sigma_1^2\sigma_2^2 - 2\sigma_{12}^2)\mathrm{vec}_D(\mathbf{A})'\mathrm{vec}_D(\mathbf{B}) + \sigma_{12}^2\mathrm{tr}(\mathbf{A}\mathbf{B}^{(s)}) \\
&\text{(iv)} & \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_2\mathbf{u}_1'\mathbf{B}\mathbf{u}_2) &= (\mu_{2,2} - \sigma_1^2\sigma_2^2 - 2\sigma_{12}^2)\mathrm{vec}_D(\mathbf{A})'\mathrm{vec}_D(\mathbf{B}) + \sigma_1^2\sigma_2^2\mathrm{tr}(\mathbf{A}\mathbf{B}') + \sigma_{12}^2\mathrm{tr}(\mathbf{A}\mathbf{B}) \\
&\text{(v)} & \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_2\mathbf{u}_2'\mathbf{B}\mathbf{u}_2) &= (\mu_{1,3} - 3\sigma_{12}\sigma_2^2)\mathrm{vec}_D(\mathbf{A})'\mathrm{vec}_D(\mathbf{B}) + \sigma_{12}\sigma_2^2\mathrm{tr}(\mathbf{A}\mathbf{B}^{(s)}) \\
&\text{(vi)} & \mathrm{E}(\mathbf{u}_2'\mathbf{A}\mathbf{u}_2\mathbf{u}_2'\mathbf{B}\mathbf{u}_2) &= (\mu_{0,4} - 3\sigma_2^4)\mathrm{vec}_D(\mathbf{A})'\mathrm{vec}_D(\mathbf{B}) + \sigma_2^4\mathrm{tr}(\mathbf{A}\mathbf{B}^{(s)})
\end{aligned}
$$

Lemma B.2 *Let* $\mathbf{A}$ *be an* $n\times n$ *nonstochastic matrix and* $\mathbf{c}$ *be an* $n\times 1$ *nonstochastic vector. Then,*

$$
\begin{aligned}
&\text{(i)} & \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_1\mathbf{u}_1'\mathbf{c}) &= \mu_{3,0}\mathrm{vec}_D(\mathbf{A})'\mathbf{c} \\
&\text{(ii)} & \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_1\mathbf{u}_2'\mathbf{c}) &= \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_2\mathbf{u}_1'\mathbf{c}) = \mu_{2,1}\mathrm{vec}_D(\mathbf{A})'\mathbf{c} \\
&\text{(iii)} & \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_2\mathbf{u}_2'\mathbf{c}) &= \mathrm{E}(\mathbf{u}_2'\mathbf{A}\mathbf{u}_2\mathbf{u}_1'\mathbf{c}) = \mu_{1,2}\mathrm{vec}_D(\mathbf{A})'\mathbf{c} \\
&\text{(iv)} & \mathrm{E}(\mathbf{u}_2'\mathbf{A}\mathbf{u}_2\mathbf{u}_2'\mathbf{c}) &= \mu_{0,3}\mathrm{vec}_D(\mathbf{A})'\mathbf{c}.
\end{aligned}
$$

Lemma B.3 *Let* $\mathbf{A}$ *be an* $n\times n$ *nonstochastic matrix with row and columns sums uniformly bounded in absolute value. Then,* (i) $n^{-1}\mathbf{u}_1'\mathbf{A}\mathbf{u}_1 = \mathrm{O_p}(1)$, $n^{-1}\mathbf{u}_1'\mathbf{A}\mathbf{u}_2 = \mathrm{O_p}(1)$*; and* (ii) $n^{-1}[\mathbf{u}_1'\mathbf{A}\mathbf{u}_1 - \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_1)] = \mathrm{o_p}(1)$, $n^{-1}[\mathbf{u}_1'\mathbf{A}\mathbf{u}_2 - \mathrm{E}(\mathbf{u}_1'\mathbf{A}\mathbf{u}_2)] = \mathrm{o_p}(1)$.

Lemma B.4 *Let* $\mathbf{A}$ *be an* $n\times n$ *nonstochastic matrix with row and columns sums uniformly bounded in absolute value. Let* $\mathbf{c}$ *be an* $n \times 1$ *nonstochastic vector with uniformly bounded elements. Then,* $n^{-1/2}\mathbf{c}'\mathbf{A}\mathbf{u}_r = \mathrm{O_p}(1)$ *and* $n^{-1}\mathbf{c}'\mathbf{A}\mathbf{u}_r = \mathrm{o_p}(1)$. *Furthermore, if* $\lim_{n\to\infty} n^{-1}\mathbf{c}'\mathbf{A}\mathbf{A}'\mathbf{c}$ *exists and is positive definite, then* $n^{-1/2}\mathbf{c}'\mathbf{A}\mathbf{u}_r \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma_r^2 \lim_{n\to\infty} n^{-1}\mathbf{c}'\mathbf{A}\mathbf{A}'\mathbf{c})$, *for* $r = 1, 2$.

Lemma B.5 *Suppose* $n^{-1}[\Gamma(\boldsymbol{\theta}) - \Gamma_0(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ *uniformly in* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, *where* $\Gamma_0(\boldsymbol{\theta})$ *is uniquely identified at* $\boldsymbol{\theta}_0$. *Define* $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Gamma(\boldsymbol{\theta})$ *and* $\widehat{\boldsymbol{\theta}}^* = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Gamma^*(\boldsymbol{\theta})$. *If* $n^{-1}[\Gamma(\boldsymbol{\theta}) - \Gamma^*(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ *uniformly in* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *then both* $\widehat{\boldsymbol{\theta}}$ *and* $\widehat{\boldsymbol{\theta}}^*$ *are consistent estimators of* $\boldsymbol{\theta}_0$. *Furthermore, assume that* $\frac{1}{n}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Gamma(\boldsymbol{\theta})$ *converges uniformly to a matrix which is nonsingular at* $\boldsymbol{\theta}_0$ *and* $\frac{1}{\sqrt{n}}\frac{\partial}{\partial\boldsymbol{\theta}'}\Gamma(\boldsymbol{\theta}) = \mathrm{O_p}(1)$. *If* $\frac{1}{n}[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Gamma^*(\boldsymbol{\theta}) - \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Gamma(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ *and* $\frac{1}{\sqrt{n}}[\frac{\partial}{\partial\boldsymbol{\theta}'}\Gamma^*(\boldsymbol{\theta}) - \frac{\partial}{\partial\boldsymbol{\theta}'}\Gamma(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ *uniformly in* $\boldsymbol{\Theta}$, *then* $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0) = \mathrm{o_p}(1)$.

Lemma B.6 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $n \times n$ *nonstochastic matrices with row and columns sums uniformly bounded in absolute value,* $\mathbf{c}_1$ *and* $\mathbf{c}_2$ *be* $n \times 1$ *nonstochastic vectors with uniformly bounded elements.* $\mathbf{G}^*$ *is either* $\mathbf{G}$, $\mathbf{G} - n^{-1}\mathrm{tr}(\mathbf{G})\mathbf{I}_n$ *or* $\mathbf{G} - \mathrm{diag}(\mathbf{G})$*, and* $\widehat{\mathbf{G}}^*$ *is obtained by replacing* λ_0 *in* $\mathbf{G}^*$ *by its* $\sqrt{n}$*-consistent estimator* $\widehat{\lambda}$*. Suppose Assumption 3 holds. Then,* $n^{-1}\mathbf{c}_1'(\widehat{\mathbf{G}}^* - \mathbf{G})\mathbf{c}_2 = \mathrm{o_p}(1)$, $n^{-1/2}\mathbf{c}_1'(\widehat{\mathbf{G}}^* - \mathbf{G})\mathbf{A}\mathbf{u}_r = \mathrm{o_p}(1)$, $n^{-1}\mathbf{u}_r'\mathbf{A}'(\widehat{\mathbf{G}}^* - \mathbf{G})\mathbf{B}\mathbf{u}_s = \mathrm{o_p}(1)$*, and* $n^{-1/2}\mathbf{u}_r'(\widehat{\mathbf{G}}^* - \mathbf{G})\mathbf{u}_s = \mathrm{o_p}(1)$*, for* $r, s = 1, 2$.

Lemma B.7 *Let* $\mathbf{A}_{r,s}$ *be an* $n\times n$ *nonstochastic matrix with row and column sums uniformly bounded in absolute value for* $r, s = 1, 2$*. Let* $\mathbf{c}_1$ *and* $\mathbf{c}_2$ *be* $n \times 1$ *nonstochastic vectors with uniformly bounded elements. Let* $\sigma^2 = \mathrm{Var}(\boldsymbol{\epsilon})$*, where* $\boldsymbol{\epsilon} = \sum_{r=1}^2 \mathbf{c}_r'\mathbf{u}_r + \sum_{s=1}^2\sum_{r=1}^2(\mathbf{u}_s'\mathbf{A}_{r,s}\mathbf{u}_r - \mathrm{E}[\mathbf{u}_s'\mathbf{A}_{r,s}\mathbf{u}_r])$*. Suppose* $\sigma^2 = \mathrm{O}(n)$ *and* $n^{-1}\sigma^2$ *is bounded away from zero. Then,* $\sigma^{-1}\boldsymbol{\epsilon} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1)$.

Lemma B.8 *Consider the set of moment conditions* $\mathbf{g}(\boldsymbol{\theta}) = [\mathbf{g}_1(\boldsymbol{\theta})', \mathbf{g}_2(\boldsymbol{\theta})']'$ *with* $\mathrm{E}[\mathbf{g}(\boldsymbol{\theta}_0)] = \mathbf{0}$*. Define* $\mathbf{D}_i = -\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}_i(\boldsymbol{\theta})]$ *and* $\boldsymbol{\Omega}_{ij} = \mathrm{E}[\mathbf{g}_i(\boldsymbol{\theta})\mathbf{g}_j(\boldsymbol{\theta})']$ *for* $i, j = 1, 2$*. The following statements are equivalent* (i) $\mathbf{g}_2$ *is redundant given* $\mathbf{g}_1$*;* (ii) $\mathbf{D}_2 = \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\mathbf{D}_1$ *and* (iii) *there exists a matrix* $\mathbf{A}$ *such that* $\mathbf{D}_2 = \boldsymbol{\Omega}_{21}\mathbf{A}$ *and* $\mathbf{D}_1 = \boldsymbol{\Omega}_{11}\mathbf{A}$.

Lemma B.9 *Consider the set of moment conditions* $\mathbf{g}(\boldsymbol{\theta}) = [\mathbf{g}_1(\boldsymbol{\theta})', \mathbf{g}_2(\boldsymbol{\theta})', \mathbf{g}_3(\boldsymbol{\theta})']'$ *with* $\mathrm{E}[\mathbf{g}(\boldsymbol{\theta}_0)] = \mathbf{0}$*. Then* $(\mathbf{g}_2', \mathbf{g}_3')'$ *is redundant given* $\mathbf{g}_1$ *if and only if* $\mathbf{g}_2$ *is redundant given* $\mathbf{g}_1$ *and* $\mathbf{g}_3$ *is redundant given* $\mathbf{g}_1$.

C Proofs

Proof of Proposition 1: To prove consistency, first we need to show the uniform convergence of $n^{-2}\mathbf{g}(\boldsymbol{\theta})'\mathbf{F}'\mathbf{F}\mathbf{g}(\boldsymbol{\theta})$ in probability. For some typical row $\mathbf{F}_{i\cdot}$ of $\mathbf{F}$

$$\begin{aligned}\mathbf{F}_{i\cdot}\mathbf{g}(\boldsymbol{\theta}) \quad = \quad & \mathbf{f}_{1,i\cdot}\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\delta}) + \mathbf{f}_{2,i\cdot}\mathbf{Q}'\mathbf{u}_2(\boldsymbol{\gamma}) + \mathbf{u}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^m f_{1,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta}) + \mathbf{u}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^m f_{2,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2(\boldsymbol{\gamma}) \\ & + \mathbf{u}_2(\boldsymbol{\gamma})'\left(\sum_{j=1}^m f_{3,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta}) + \mathbf{u}_2(\boldsymbol{\gamma})'\left(\sum_{j=1}^m f_{4,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2(\boldsymbol{\gamma})\end{aligned}$$

where $\mathbf{F}_{i\cdot} = (\mathbf{f}_{1,i\cdot}, \mathbf{f}_{2,i\cdot}, f_{1,i1}, \cdots, f_{1,im}, \cdots, f_{4,i1}, \cdots, f_{4,im})$ and $\mathbf{f}_{1,i\cdot}$ and $\mathbf{f}_{2,i\cdot}$ are row sub-vectors. As $\mathbf{u}_1(\boldsymbol{\delta}) = \mathbf{d}_1(\boldsymbol{\delta}) + \mathbf{r}_1(\boldsymbol{\delta})$, where $\mathbf{d}_1(\boldsymbol{\delta}) = (\lambda_0 - \lambda)\mathbf{G}(\phi_0\mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{X}_1\boldsymbol{\beta}_0) + (\phi_0 - \phi)\mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{X}_1(\boldsymbol{\beta}_0 - \boldsymbol{\beta})$ and

$\mathbf{r}_1(\boldsymbol{\delta}) = \mathbf{u}_1 + (\lambda_0 - \lambda)(\mathbf{G}\mathbf{u}_1 + \phi_0\mathbf{G}\mathbf{u}_2) + (\phi_0 - \phi)\mathbf{u}_2$, we have

$$\mathbf{u}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{1,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta}) = \mathbf{d}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{1,ij}\boldsymbol{\Xi}_j\right)\mathbf{d}_1(\boldsymbol{\delta}) + l_1(\boldsymbol{\delta}) + q_1(\boldsymbol{\delta})$$

where $l_1(\boldsymbol{\delta}) = \mathbf{d}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{1,ij}\boldsymbol{\Xi}_j^{(s)}\right)\mathbf{r}_1(\boldsymbol{\delta})$ and $q_1(\boldsymbol{\delta}) = \mathbf{r}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{1,ij}\boldsymbol{\Xi}_j\right)\mathbf{r}_1(\boldsymbol{\delta})$. It follows by Lemmas B.3 and B.4 that $n^{-1}l_1(\boldsymbol{\delta}) = \mathrm{o_p}(1)$ and $n^{-1}q_1(\boldsymbol{\delta}) - n^{-1}\mathrm{E}[q_1(\boldsymbol{\delta})] = \mathrm{o_p}(1)$ uniformly in Θ, where

$$\begin{aligned}
\mathrm{E}[q_1(\boldsymbol{\delta})] \quad = \quad & (\lambda_0 - \lambda)[\sigma_1^2 + \sigma_{12}(2\phi_0 - \phi) + \sigma_2^2\phi_0(\phi_0 - \phi)]\sum_{j=1}^{m} f_{1,ij}\mathrm{tr}(\mathbf{G}\boldsymbol{\Xi}_j^{(s)}) \\
& +(\lambda_0 - \lambda)^2(\sigma_2^2\phi_0^2 + 2\sigma_{12}\phi_0 + \sigma_1^2)\sum_{j=1}^{m} f_{1,ij}\mathrm{tr}(\mathbf{G}'\boldsymbol{\Xi}_j\mathbf{G}).
\end{aligned}$$

Hence, $n^{-1}\mathbf{u}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{1,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta}) - n^{-1}\mathrm{E}[\mathbf{u}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{1,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta})] = \mathrm{o_p}(1)$ uniformly in Θ, where $\mathrm{E}[\mathbf{u}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{1,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta})] = \mathbf{d}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{1,ij}\boldsymbol{\Xi}_j\right)\mathbf{d}_1(\boldsymbol{\delta}) + \mathrm{E}[q_1(\boldsymbol{\delta})]$. As $\mathbf{u}_2(\boldsymbol{\gamma}) = \mathbf{d}_2(\boldsymbol{\gamma}) + \mathbf{u}_2$, where $\mathbf{d}_2(\boldsymbol{\gamma}) = \mathbf{X}(\boldsymbol{\gamma}_0 - \boldsymbol{\gamma})$, we have

$$\mathbf{u}_1(\boldsymbol{\gamma})'\left(\sum_{j=1}^{m} f_{2,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2(\boldsymbol{\delta}) = \mathbf{d}_1(\boldsymbol{\gamma})'\left(\sum_{j=1}^{m} f_{2,ij}\boldsymbol{\Xi}_j\right)\mathbf{d}_2(\boldsymbol{\delta}) + l_2(\boldsymbol{\theta}) + q_2(\boldsymbol{\theta})$$

where $l_2(\boldsymbol{\theta}) = \mathbf{r}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{2,ij}\boldsymbol{\Xi}_j\right)\mathbf{d}_2(\boldsymbol{\gamma}) + \mathbf{d}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{2,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2$ and $q_2(\boldsymbol{\delta}) = \mathbf{r}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^{m} f_{2,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2$. It follows by Lemmas B.3 and B.4 that $n^{-1}l_2(\boldsymbol{\theta}) = \mathrm{o_p}(1)$ and $n^{-1}q_2(\boldsymbol{\theta}) - n^{-1}\mathrm{E}[q_2(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ uniformly in Θ, where

$$\mathrm{E}[q_2(\boldsymbol{\theta})] = (\lambda_0 - \lambda)(\sigma_{12} + \sigma_2^2\phi_0)\sum_{j=1}^{m} f_{2,ij}\mathrm{tr}(\mathbf{G}\boldsymbol{\Xi}_j).$$

Hence, $n^{-1}\mathbf{u}_1(\boldsymbol{\gamma})'\left(\sum_{j=1}^{m} f_{2,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2(\boldsymbol{\delta}) - n^{-1}\mathrm{E}[\mathbf{u}_1(\boldsymbol{\gamma})'\left(\sum_{j=1}^{m} f_{2,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2(\boldsymbol{\delta})] = \mathrm{o_p}(1)$ uniformly in Θ, where $\mathrm{E}[\mathbf{u}_1(\boldsymbol{\gamma})'\left(\sum_{j=1}^{m} f_{2,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2(\boldsymbol{\delta})] = \mathbf{d}_1(\boldsymbol{\gamma})'\left(\sum_{j=1}^{m} f_{2,ij}\boldsymbol{\Xi}_j\right)\mathbf{d}_2(\boldsymbol{\delta}) + \mathrm{E}[q_2(\boldsymbol{\theta})]$. Similarly, $n^{-1}\mathbf{u}_2(\boldsymbol{\gamma})'\left(\sum_{j=1}^{m} f_{3,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta}) - n^{-1}\mathrm{E}[\mathbf{u}_2(\boldsymbol{\gamma})'\left(\sum_{j=1}^{m} f_{3,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta})] = \mathrm{o_p}(1)$, $n^{-1}\mathbf{u}_2(\boldsymbol{\gamma})'\left(\sum_{j=1}^{m} f_{4,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2(\boldsymbol{\gamma}) - n^{-1}\mathrm{E}[\mathbf{u}_2(\boldsymbol{\gamma})'\left(\sum_{j=1}^{m} f_{4,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2(\boldsymbol{\gamma})] = \mathrm{o_p}(1)$, $n^{-1}\mathbf{f}_{1,i\cdot}\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\delta}) - n^{-1}\mathrm{E}[\mathbf{f}_{1,i\cdot}\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\delta})] = \mathrm{o_p}(1)$, and $n^{-1}\mathbf{f}_{2,i\cdot}\mathbf{Q}'\mathbf{u}_2(\boldsymbol{\gamma}) - n^{-1}\mathrm{E}[\mathbf{f}_{2,i\cdot}\mathbf{Q}'\mathbf{u}_2(\boldsymbol{\gamma})] = \mathrm{o_p}(1)$ uniformly in Θ. Therefore, $n^{-1}\mathbf{F}\mathbf{g}(\boldsymbol{\theta}) - n^{-1}\mathbf{F}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ uniformly in Θ, and hence, $n^{-2}\mathbf{g}(\boldsymbol{\theta})'\mathbf{F}'\mathbf{F}\mathbf{g}(\boldsymbol{\theta})$ converges in probability to a well defined limit uniformly in Θ. As $\mathbf{g}(\boldsymbol{\theta})$ is a quadratic function of $\boldsymbol{\theta}$, $n^{-1}\mathbf{F}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})]$ is uniformly equicontinuous on Θ by Assumption 4. The identification condition and the uniform equicontinuity of $n^{-1}\mathbf{F}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})]$

imply that the identification uniqueness condition for $n^{-2}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})']\mathbf{F}'\mathbf{F}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})]$ must be satisfied. The consistency of $\widehat{\boldsymbol{\theta}}$ follows by Theorem 15.1 of Peracchi (2001).

For the asymptotic normality of $\widetilde{\boldsymbol{\theta}}_{gmm}$, by the mean value theorem,

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}}_{gmm}-\boldsymbol{\theta}_0) = -\left[n^{-1}\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{g}(\widetilde{\boldsymbol{\theta}}_{gmm})'\mathbf{F}'n^{-1}\mathbf{F}\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\overline{\boldsymbol{\theta}})\right]^{-1} n^{-1}\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{g}(\widetilde{\boldsymbol{\theta}}_{gmm})'\mathbf{F}'n^{-1/2}\mathbf{F}\mathbf{g}(\boldsymbol{\theta}_0)$$

where $\overline{\boldsymbol{\theta}} = t\widetilde{\boldsymbol{\theta}}_{gmm} + (1-t)\boldsymbol{\theta}_0$ for some $t \in [0,1]$ and

$$-\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{Q}'\mathbf{W}\mathbf{y}_1 & \mathbf{Q}'\mathbf{y}_2 & \mathbf{Q}'\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}'\mathbf{X} \\ \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_1^{(s)}\mathbf{W}\mathbf{y}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_1^{(s)}\mathbf{y}_2 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_1^{(s)}\mathbf{X}_1 & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_m^{(s)}\mathbf{W}\mathbf{y}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_m^{(s)}\mathbf{y}_2 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_m^{(s)}\mathbf{X}_1 & \mathbf{0} \\ \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_1'\mathbf{W}\mathbf{y}_1 & \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_1'\mathbf{y}_2 & \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_1'\mathbf{X}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_1\mathbf{X} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_m'\mathbf{W}\mathbf{y}_1 & \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_m'\mathbf{y}_2 & \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_m'\mathbf{X}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_m\mathbf{X} \\ \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_1\mathbf{W}\mathbf{y}_1 & \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_1\mathbf{y}_2 & \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_1\mathbf{X}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_1'\mathbf{X} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_m\mathbf{W}\mathbf{y}_1 & \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_m\mathbf{y}_2 & \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_m\mathbf{X}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}_m'\mathbf{X} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_1^{(s)}\mathbf{X} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{u}_2(\gamma)'\boldsymbol{\Xi}_m^{(s)}\mathbf{X} \end{bmatrix}.$$

Using Lemmas B.3 and B.4, it follows by a similar argument in the proof of Proposition 1 in Lee (2007) that $-n^{-1}\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\widehat{\boldsymbol{\theta}}) - n^{-1}\mathbf{D} = \mathrm{o_p}(1)$ and $-n^{-1}\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\overline{\boldsymbol{\theta}}) - n^{-1}\mathbf{D} = \mathrm{o_p}(1)$ with $\mathbf{D}$ given by (10). By Lemma B.7 and the Cramer-Wald device, we have $n^{-1/2}\mathbf{F}\mathbf{g}(\boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \lim_{n\to\infty} n^{-1}\mathbf{F}\boldsymbol{\Omega}\mathbf{F}')$ with $\boldsymbol{\Omega}$ given by (9). The desired result follows. ∎

Proof of Proposition 2: Note that

$$n^{-1}\mathbf{g}(\boldsymbol{\theta})'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{g}(\boldsymbol{\theta}) = n^{-1}\mathbf{g}(\boldsymbol{\theta})'\boldsymbol{\Omega}^{-1}\mathbf{g}(\boldsymbol{\theta}) + n^{-1}\mathbf{g}(\boldsymbol{\theta})'(\widehat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\mathbf{g}(\boldsymbol{\theta}).$$

With $\mathbf{F} = (n^{-1}\boldsymbol{\Omega})^{-1/2}$, uniform convergence of $n^{-1}\mathbf{g}(\boldsymbol{\theta})'\boldsymbol{\Omega}^{-1}\mathbf{g}(\boldsymbol{\theta})$ in probability follows by a similar argument in the proof of Proposition 1. On the other hand,

$$\left\| n^{-1}\mathbf{g}(\boldsymbol{\theta})'(\widehat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\mathbf{g}(\boldsymbol{\theta}) \right\| \leq \left(n^{-1}\|\mathbf{g}(\boldsymbol{\theta})\| \right)^2 \left\| (n^{-1}\widehat{\boldsymbol{\Omega}})^{-1} - (n^{-1}\boldsymbol{\Omega})^{-1} \right\|$$

where $\|\cdot\|$ is the Euclidean norm for vectors and matrices. By a similar argument in the proof of Proposition 1, we have $n^{-1}\mathbf{g}(\boldsymbol{\theta}) - n^{-1}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ and $n^{-1}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})] = \mathrm{O}(1)$ uniformly in Θ, which in turn implies that $n^{-1}\|\mathbf{g}(\boldsymbol{\theta})\| = \mathrm{O_p}(1)$ uniformly in Θ. Therefore, $\left\| n^{-1}\mathbf{g}(\boldsymbol{\theta})'(\widehat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\mathbf{g}(\boldsymbol{\theta}) \right\| = \mathrm{o_p}(1)$ uniformly in Θ. The consistency of $\widehat{\boldsymbol{\theta}}_{gmm}$ follows.

For the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0)$, note that from the proof of Proposition 1 we have $-n^{-1}\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm}) - n^{-1}\mathbf{D} = \mathrm{o_p}(1)$, since $\widehat{\boldsymbol{\theta}}_{gmm}$ is consistent. Let $\overline{\boldsymbol{\theta}} = t\widehat{\boldsymbol{\theta}}_{gmm} + (1-t)\boldsymbol{\theta}_0$ for some $t \in [0,1]$, then by the mean value theorem,

$$\begin{aligned}
& \sqrt{n}(\widehat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0) \\
= & -\left[n^{-1}\frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm})'(n^{-1}\widehat{\boldsymbol{\Omega}})^{-1} n^{-1}\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\overline{\boldsymbol{\theta}}) \right]^{-1} n^{-1}\frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm})'(n^{-1}\widehat{\boldsymbol{\Omega}})^{-1} n^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0) \\
= & \left[n^{-1}\mathbf{D}' \left(n^{-1}\boldsymbol{\Omega}\right)^{-1} n^{-1}\mathbf{D} \right]^{-1} n^{-1}\mathbf{D}' \left(n^{-1}\boldsymbol{\Omega}\right)^{-1} n^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0) + \mathrm{o_p}(1)
\end{aligned}$$

which concludes the first part of the proof, since in the proof of Proposition 1 it is established that $n^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0)$ converges in distribution.

For the overidentification test, by the mean value theorem, for some $t \in [0,1]$ and $\overline{\boldsymbol{\theta}} = t\widehat{\boldsymbol{\theta}}_{gmm} + (1-t)\boldsymbol{\theta}_0$

$$\begin{aligned}
n^{-1/2}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm}) &= n^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0) + n^{-1/2}\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\overline{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0) \\
&= n^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0) - n^{-1}\mathbf{D}\sqrt{n}(\widehat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0) + \mathrm{o_p}(1) \\
&= \mathbf{A}n^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0) + \mathrm{o_p}(1)
\end{aligned}$$

where $\mathbf{A} = \mathbf{I}_{\dim(\mathbf{g})} - n^{-1}\mathbf{D}\left[n^{-1}\mathbf{D}' \left(n^{-1}\boldsymbol{\Omega}\right)^{-1} n^{-1}\mathbf{D} \right]^{-1} n^{-1}\mathbf{D}' \left(n^{-1}\boldsymbol{\Omega}\right)^{-1}$. Therefore

$$\begin{aligned}
\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm})'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm}) &= n^{-1/2}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm})'(n^{-1}\boldsymbol{\Omega})^{-1}n^{-1/2}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm}) + \mathrm{o_p}(1) \\
&= n^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0)'\mathbf{A}'(n^{-1}\boldsymbol{\Omega})^{-1}\mathbf{A}n^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0) + \mathrm{o_p}(1) \\
&= [\left(n^{-1}\boldsymbol{\Omega}\right)^{-1/2} n^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0)]'\mathbf{B}[\left(n^{-1}\boldsymbol{\Omega}\right)^{-1/2} n^{-1/2}\mathbf{g}(\boldsymbol{\theta}_0)] + \mathrm{o_p}(1)
\end{aligned}$$

where $\mathbf{B} = \mathbf{I}_{\dim(\mathbf{g})} - \left(n^{-1}\boldsymbol{\Omega}\right)^{-1/2} n^{-1}\mathbf{D}\left[n^{-1}\mathbf{D}'\left(n^{-1}\boldsymbol{\Omega}\right)^{-1} n^{-1}\mathbf{D}\right]^{-1} n^{-1}\mathbf{D}'\left(n^{-1}\boldsymbol{\Omega}\right)^{-1/2}$. Therefore

$$\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm})'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm}) \xrightarrow{\mathrm{d}} \chi^2_{\mathrm{tr}(\mathbf{B})},$$

where $\mathrm{tr}(\mathbf{B}) = \dim(\mathbf{g}) - \dim(\boldsymbol{\theta})$. ■

Proof of Proposition 3: To establish the asymptotic efficiency, we use an argument by Breusch et al. (1999) to show that any additional moment conditions $\mathbf{g}$ defined in (6) given $\mathbf{g}^*$ defined in (12) will be redundant. Following Breusch et al. (1999), $\mathbf{g}$ is redundant given $\mathbf{g}^*$ if the asymptotic variance of an estimator based on moment equations $\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})] = \mathbf{0}$ and $\mathrm{E}[\mathbf{g}^*(\boldsymbol{\theta})] = \mathbf{0}$ is the same as an estimator based on $\mathrm{E}[\mathbf{g}^*(\boldsymbol{\theta})] = \mathbf{0}$. In cases (i) and (ii),

$$\boldsymbol{\Omega}^{\#} = \mathrm{E}[\mathbf{g}(\boldsymbol{\theta}_0)\mathbf{g}^*(\boldsymbol{\theta}_0)'] = \begin{bmatrix} (\boldsymbol{\Sigma} \otimes \mathbf{Q}'\mathbf{Q}^*) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}^{\#}_{22} \end{bmatrix}$$

where

$$\begin{aligned} \boldsymbol{\Omega}^{\#}_{22} &= \begin{bmatrix} \sigma_1^4 & \sigma_1^2\sigma_{12} & \sigma_1^2\sigma_{12} & \sigma_{12}^2 \\ \sigma_1^2\sigma_{12} & \sigma_{12}^2 & \sigma_1^2\sigma_2^2 & \sigma_2^2\sigma_{12} \\ \sigma_1^2\sigma_{12} & \sigma_1^2\sigma_2^2 & \sigma_{12}^2 & \sigma_2^2\sigma_{12} \\ \sigma_{12}^2 & \sigma_2^2\sigma_{12} & \sigma_2^2\sigma_{12} & \sigma_2^4 \end{bmatrix} \otimes \begin{bmatrix} \mathrm{tr}(\boldsymbol{\Xi}_1\boldsymbol{\Xi}^*) \\ \vdots \\ \mathrm{tr}(\boldsymbol{\Xi}_m\boldsymbol{\Xi}^*) \end{bmatrix} \\ &+ \begin{bmatrix} \sigma_1^4 & \sigma_1^2\sigma_{12} & \sigma_1^2\sigma_{12} & \sigma_{12}^2 \\ \sigma_1^2\sigma_{12} & \sigma_1^2\sigma_2^2 & \sigma_{12}^2 & \sigma_2^2\sigma_{12} \\ \sigma_1^2\sigma_{12} & \sigma_{12}^2 & \sigma_1^2\sigma_2^2 & \sigma_2^2\sigma_{12} \\ \sigma_{12}^2 & \sigma_2^2\sigma_{12} & \sigma_2^2\sigma_{12} & \sigma_2^4 \end{bmatrix} \otimes \begin{bmatrix} \mathrm{tr}(\boldsymbol{\Xi}_1'\boldsymbol{\Xi}^*) \\ \vdots \\ \mathrm{tr}(\boldsymbol{\Xi}_m'\boldsymbol{\Xi}^*) \end{bmatrix}. \end{aligned}$$

Let

$$\mathbf{A} = \frac{1}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}\begin{bmatrix} \sigma_2^2(\mathbf{C}\boldsymbol{\beta}_0 + \phi_0\boldsymbol{\gamma}_0) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\boldsymbol{\gamma}_0 & \sigma_2^2\mathbf{C} & -\sigma_{12}\mathbf{I}_{K_X} \\ -\sigma_{12}(\mathbf{C}\boldsymbol{\beta}_0 + \phi_0\boldsymbol{\gamma}_0) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\sigma_{12}\boldsymbol{\gamma}_0 & -\sigma_{12}\mathbf{C} & \sigma_1^2\mathbf{I}_{K_X} \\ \sigma_2^2 & 0 & \mathbf{0} & \mathbf{0} \\ \phi_0\sigma_2^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\sigma_{12} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\phi_0\sigma_{12} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where $\mathbf{C} = [\mathbf{I}_{K_1}, \mathbf{0}]'$ and $\mathbf{X}_1 = \mathbf{X}\mathbf{C}$. Then $\mathbf{D} = \boldsymbol{\Omega}^{\#}\mathbf{A}$, where $\mathbf{D}$ is defined in (10). Based on Lemma B.8 $\mathbf{g}$ is redundant given $\mathbf{g}^*$. Furthermore, Lemma B.9 tells us that any subset of $\mathbf{g}$ is redundant given $\mathbf{g}^*$. ■

Proof of Proposition 4: To show the desired result, we only need to show $\widehat{\boldsymbol{\Gamma}}(\boldsymbol{\theta}) = \widehat{\mathbf{g}}^*(\boldsymbol{\theta})'\widehat{\boldsymbol{\Omega}}^{*-1}\widehat{\mathbf{g}}^*(\boldsymbol{\theta})$ and $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \mathbf{g}^*(\boldsymbol{\theta})'\boldsymbol{\Omega}^{*-1}\mathbf{g}^*(\boldsymbol{\theta})$ satisfy the conditions of Lemma B.5. First, $n^{-1}[\widehat{\mathbf{g}}_1^*(\boldsymbol{\theta}) - \mathbf{g}_1^*(\boldsymbol{\theta})] = n^{-1}[\mathbf{I}_2 \otimes (\widehat{\mathbf{Q}}^* - \mathbf{Q}^*)]'\mathbf{u}(\boldsymbol{\theta})$, $n^{-1}[\widehat{\mathbf{g}}_{2,rs}^*(\boldsymbol{\theta}) - \mathbf{g}_{2,rs}^*(\boldsymbol{\theta})] = n^{-1}\mathbf{u}_r(\boldsymbol{\theta})'(\widehat{\boldsymbol{\Xi}}^* - \boldsymbol{\Xi}^*)\mathbf{u}_s(\boldsymbol{\theta})$,

$$\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}^*(\boldsymbol{\theta}) = -\begin{bmatrix} \mathbf{Q}^{*\prime}\mathbf{W}\mathbf{y}_1 & \mathbf{Q}^{*\prime}\mathbf{y}_2 & \mathbf{Q}^{*\prime}\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}^{*\prime}\mathbf{X} \\ \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}^{*(s)}\mathbf{W}\mathbf{y}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}^{*(s)}\mathbf{y}_2 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}^{*(s)}\mathbf{X}_1 & \mathbf{0} \\ \mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^{*\prime}\mathbf{W}\mathbf{y}_1 & \mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^{*\prime}\mathbf{y}_2 & \mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^{*\prime}\mathbf{X}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}^{*}\mathbf{X} \\ \mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^{*}\mathbf{W}\mathbf{y}_1 & \mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^{*}\mathbf{y}_2 & \mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^{*}\mathbf{X}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}^{*\prime}\mathbf{X} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^{*(s)}\mathbf{X} \end{bmatrix},$$

and

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\mathbf{g}^*(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{0} \\ \frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}^*\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{u}_1(\boldsymbol{\delta}) \\ \frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}^*\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{u}_2(\boldsymbol{\gamma}) \\ \frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^*\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{u}_1(\boldsymbol{\delta}) \\ \frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^*\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{u}_2(\boldsymbol{\gamma}) \end{bmatrix},$$

where $\mathbf{Q}^* = [\mathbf{G}\mathbf{X}, \mathbf{X}]$, $\boldsymbol{\Xi}^*$ is either $\mathbf{G} - n^{-1}\mathrm{tr}(\mathbf{G})\mathbf{I}_n$ or $\mathbf{G} - \mathrm{diag}(\mathbf{G})$, $\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{u}_1(\boldsymbol{\delta}) = -[\mathbf{W}\mathbf{y}_1, \mathbf{y}_2, \mathbf{X}_1, \mathbf{0}]$, and $\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{u}_2(\boldsymbol{\gamma}) = -[\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{X}]$. By inspection of each term of the above matrices, we conclude

$n^{-1}[\widehat{\mathbf{g}}^*(\boldsymbol{\theta})-\mathbf{g}^*(\boldsymbol{\theta})]=\mathrm{o_p}(1)$, $n^{-1}[\frac{\partial}{\partial\boldsymbol{\theta}'}\widehat{\mathbf{g}}^*(\boldsymbol{\theta})-\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}^*(\boldsymbol{\theta})]=\mathrm{o_p}(1)$ and $n^{-1}[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\widehat{\mathbf{g}}^*(\boldsymbol{\theta})-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\mathbf{g}^*(\boldsymbol{\theta})]=\mathrm{o_p}(1)$ uniformly in $\boldsymbol{\Theta}$ by Lemma B.6. Second, as $\widehat{\mathbf{G}}-\mathbf{G}=(\widehat{\lambda}-\lambda_0)\mathbf{G}^2+(\widehat{\lambda}-\lambda_0)^2\widehat{\mathbf{G}}\mathbf{G}^2$, we have $n^{-1}\mathrm{tr}(\widehat{\boldsymbol{\Xi}}^*\widehat{\boldsymbol{\Xi}}^*)-n^{-1}\mathrm{tr}(\boldsymbol{\Xi}^*\boldsymbol{\Xi}^*)=\mathrm{o_p}(1)$ and $n^{-1}\mathrm{tr}(\widehat{\boldsymbol{\Xi}}^{*\prime}\widehat{\boldsymbol{\Xi}}^*)-n^{-1}\mathrm{tr}(\boldsymbol{\Xi}^{*\prime}\boldsymbol{\Xi}^*)=\mathrm{o_p}(1)$. Therefore, as $\widehat{\boldsymbol{\Sigma}}$ is a consistent estimator of $\boldsymbol{\Sigma}$, we have $n^{-1}(\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}^*)=\mathrm{o_p}(1)$. Hence, we can conclude that $n^{-1}[\widehat{\boldsymbol{\Gamma}}(\boldsymbol{\theta})-\Gamma(\boldsymbol{\theta})]=\mathrm{o_p}(1)$ and $n^{-1}[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\widehat{\boldsymbol{\Gamma}}(\boldsymbol{\theta})-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Gamma(\boldsymbol{\theta})]=\mathrm{o_p}(1)$ uniformly in $\boldsymbol{\Theta}$. Finally, since $n^{-1/2}\mathbf{g}^*(\boldsymbol{\theta}_0)=\mathrm{O_p}(1)$ by a similar argument in the proof of Proposition 1 and $n^{-1/2}[\widehat{\mathbf{g}}^*(\boldsymbol{\theta}_0)-\mathbf{g}^*(\boldsymbol{\theta}_0)]=\mathrm{o_p}(1)$ by Lemma B.6,

$$
\begin{aligned}
& n^{-1/2}[\frac{\partial}{\partial\boldsymbol{\theta}'}\widehat{\boldsymbol{\Gamma}}(\boldsymbol{\theta}_0)-\frac{\partial}{\partial\boldsymbol{\theta}'}\Gamma(\boldsymbol{\theta}_0)] \\
= \;& 2\frac{\partial}{\partial\boldsymbol{\theta}}\widehat{\mathbf{g}}^*(\boldsymbol{\theta}_0)'\widehat{\boldsymbol{\Omega}}^{-1}n^{-1/2}[\widehat{\mathbf{g}}^*(\boldsymbol{\theta}_0)-\mathbf{g}^*(\boldsymbol{\theta}_0)]+2[\frac{\partial}{\partial\boldsymbol{\theta}}\widehat{\mathbf{g}}^*(\boldsymbol{\theta}_0)'\widehat{\boldsymbol{\Omega}}^{-1}-\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{g}^*(\boldsymbol{\theta}_0)'\boldsymbol{\Omega}^{-1}]n^{-1/2}\mathbf{g}^*(\boldsymbol{\theta}_0) \\
= \;& \mathrm{o_p}(1).
\end{aligned}
$$

The desired result follows. ■

Table 2: 2SLS, 3SLS and GMM Estimation ($n = 245$)

	$\lambda_0 = 0.6$	$\phi_0 = 0.5$	$\beta_0 = 0.5$	$\gamma_0 = 1$
$\sigma_{12} = 0.1$				
2SLS	0.601(0.128)[0.128]	0.497(0.068)[0.068]	0.496(0.066)[0.066]	-
3SLS	0.601(0.126)[0.126]	0.497(0.068)[0.068]	0.496(0.066)[0.066]	0.999(0.066)[0.066]
GMM-1	0.602(0.052)[0.052]	0.499(0.066)[0.066]	0.498(0.065)[0.065]	-
GMM-2	0.607(0.051)[0.051]	0.497(0.067)[0.068]	0.498(0.065)[0.065]	0.999(0.066)[0.066]
$\sigma_{12} = 0.5$				
2SLS	0.601(0.138)[0.138]	0.496(0.068)[0.068]	0.495(0.066)[0.066]	-
3SLS	0.602(0.111)[0.111]	0.496(0.068)[0.068]	0.498(0.057)[0.057]	1.000(0.066)[0.066]
GMM-1	0.602(0.046)[0.046]	0.501(0.066)[0.066]	0.498(0.065)[0.065]	-
GMM-2	0.604(0.045)[0.045]	0.495(0.068)[0.068]	0.499(0.057)[0.057]	0.999(0.066)[0.066]
$\sigma_{12} = 0.9$				
2SLS	0.601(0.170)[0.170]	0.495(0.070)[0.070]	0.494(0.067)[0.067]	-
3SLS	0.603(0.059)[0.059]	0.497(0.068)[0.068]	0.500(0.029)[0.029]	1.001(0.066)[0.066]
GMM-1	0.603(0.050)[0.050]	0.503(0.066)[0.066]	0.498(0.065)[0.065]	-
GMM-2	0.601(0.023)[0.023]	0.495(0.070)[0.070]	0.500(0.028)[0.028]	0.999(0.067)[0.067]

Mean(SD)[RMSE]

Table 3: 2SLS, 3SLS and GMM Estimation ($n = 490$)

	$\lambda_0 = 0.6$	$\phi_0 = 0.5$	$\beta_0 = 0.5$	$\gamma_0 = 1$
$\sigma_{12} = 0.1$				
2SLS	0.600(0.080)[0.080]	0.497(0.047)[0.047]	0.497(0.046)[0.046]	-
3SLS	0.599(0.079)[0.079]	0.497(0.047)[0.047]	0.497(0.046)[0.046]	1.000(0.045)[0.045]
GMM-1	0.600(0.035)[0.035]	0.498(0.047)[0.047]	0.498(0.046)[0.046]	-
GMM-2	0.602(0.034)[0.034]	0.497(0.047)[0.047]	0.498(0.045)[0.046]	1.000(0.046)[0.046]
$\sigma_{12} = 0.5$				
2SLS	0.600(0.081)[0.081]	0.496(0.048)[0.048]	0.497(0.046)[0.046]	-
3SLS	0.600(0.068)[0.068]	0.496(0.047)[0.047]	0.499(0.040)[0.040]	0.999(0.046)[0.046]
GMM-1	0.600(0.030)[0.030]	0.499(0.047)[0.047]	0.498(0.046)[0.046]	-
GMM-2	0.601(0.029)[0.029]	0.496(0.047)[0.048]	0.499(0.040)[0.040]	0.999(0.046)[0.046]
$\sigma_{12} = 0.9$				
2SLS	0.601(0.082)[0.082]	0.496(0.048)[0.048]	0.496(0.046)[0.046]	-
3SLS	0.601(0.034)[0.034]	0.496(0.047)[0.048]	0.500(0.020)[0.020]	0.999(0.047)[0.047]
GMM-1	0.600(0.026)[0.026]	0.500(0.047)[0.047]	0.498(0.045)[0.045]	-
GMM-2	0.600(0.015)[0.015]	0.494(0.048)[0.049]	0.500(0.020)[0.020]	0.997(0.048)[0.048]

Mean(SD)[RMSE]

Table 4: 2SLS, 3SLS and GMM Estimation ($n = 245$)

	$\lambda_0 = 0.6$	$\phi_0 = 0.2$	$\beta_0 = 0.2$	$\gamma_0 = 1$
$\sigma_{12} = 0.1$				
2SLS	0.667(0.464)[0.469]	0.196(0.075)[0.076]	0.194(0.070)[0.071]	-
3SLS	0.660(0.482)[0.486]	0.195(0.076)[0.076]	0.195(0.070)[0.070]	0.999(0.066)[0.066]
GMM-1	0.637(0.163)[0.167]	0.201(0.067)[0.067]	0.198(0.066)[0.066]	-
GMM-2	0.640(0.145)[0.150]	0.199(0.068)[0.068]	0.198(0.065)[0.065]	0.999(0.066)[0.066]
$\sigma_{12} = 0.5$				
2SLS	0.678(0.439)[0.446]	0.195(0.070)[0.070]	0.194(0.068)[0.068]	-
3SLS	0.653(0.357)[0.361]	0.195(0.069)[0.069]	0.197(0.058)[0.059]	1.000(0.066)[0.066]
GMM-1	0.648(0.189)[0.195]	0.202(0.067)[0.067]	0.198(0.066)[0.066]	-
GMM-2	0.624(0.109)[0.112]	0.196(0.068)[0.068]	0.199(0.057)[0.057]	0.999(0.066)[0.066]
$\sigma_{12} = 0.9$				
2SLS	0.688(0.389)[0.399]	0.194(0.070)[0.070]	0.194(0.068)[0.068]	-
3SLS	0.627(0.168)[0.170]	0.196(0.067)[0.068]	0.199(0.029)[0.029]	1.001(0.066)[0.066]
GMM-1	0.646(0.178)[0.184]	0.204(0.067)[0.067]	0.198(0.065)[0.065]	-
GMM-2	0.608(0.052)[0.053]	0.196(0.068)[0.069]	0.200(0.029)[0.029]	0.999(0.067)[0.067]

Mean(SD)[RMSE]

Table 5: 2SLS, 3SLS and GMM Estimation ($n = 490$)

	$\lambda_0 = 0.6$	$\phi_0 = 0.2$	$\beta_0 = 0.2$	$\gamma_0 = 1$
$\sigma_{12} = 0.1$				
2SLS	0.625(0.251)[0.253]	0.195(0.047)[0.048]	0.195(0.046)[0.047]	-
3SLS	0.624(0.252)[0.253]	0.195(0.047)[0.048]	0.195(0.046)[0.046]	1.000(0.045)[0.045]
GMM-1	0.610(0.094)[0.094]	0.198(0.047)[0.047]	0.198(0.046)[0.046]	-
GMM-2	0.610(0.071)[0.072]	0.197(0.047)[0.047]	0.198(0.045)[0.045]	1.000(0.046)[0.046]
$\sigma_{12} = 0.5$				
2SLS	0.633(0.227)[0.230]	0.195(0.048)[0.048]	0.195(0.046)[0.047]	-
3SLS	0.620(0.195)[0.196]	0.195(0.047)[0.048]	0.197(0.040)[0.040]	0.999(0.046)[0.046]
GMM-1	0.611(0.092)[0.092]	0.199(0.047)[0.047]	0.198(0.046)[0.046]	-
GMM-2	0.604(0.043)[0.043]	0.196(0.047)[0.047]	0.199(0.040)[0.040]	0.999(0.046)[0.046]
$\sigma_{12} = 0.9$				
2SLS	0.628(0.280)[0.282]	0.195(0.048)[0.048]	0.194(0.047)[0.047]	-
3SLS	0.607(0.100)[0.100]	0.196(0.047)[0.048]	0.200(0.020)[0.020]	0.999(0.047)[0.047]
GMM-1	0.612(0.097)[0.098]	0.200(0.047)[0.047]	0.198(0.046)[0.046]	-
GMM-2	0.602(0.024)[0.024]	0.195(0.048)[0.048]	0.200(0.020)[0.020]	0.997(0.047)[0.047]

Mean(SD)[RMSE]