Spring 1-1-2015

GMM Estimation of Spatial Autoregressive Models in a System of Simultaneous Equations

Paulo Quinderé Saraiva
University of Colorado Boulder, saraiva@colorado.edu

GMM Estimation of Spatial Autoregressive Models in a System of Simultaneous Equations

by

Paulo Quinderé Saraiva

B.A., Universidade Federal do Ceará, 2003

M.S., Universidade Federal do Ceará, 2005

M.S., Oregon State University, 2009

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Economics

2015

This thesis entitled:
GMM Estimation of Spatial Autoregressive Models in a System of Simultaneous Equations
written by Paulo Quinderé Saraiva
has been approved for the Department of Economics

Xiaodong Liu

Prof. Robert McNown

Date ____________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Saraiva, Paulo Quinderé (Ph.D., Economics)

GMM Estimation of Spatial Autoregressive Models in a System of Simultaneous Equations

Thesis directed by Prof. Xiaodong Liu

This dissertation proposes a generalized method of moments (GMM) estimation framework for the spatial autorregressive (SAR) model in a system of simultaneous equations with homoskedastic and heteroskedastic disturbances. It includes two chapters based on joint work with Prof. Xiaodong Liu.

The first chapter extends the GMM estimator in Lee (2007) to estimate SAR models with endogenous regressors and homoskedastic disturbances. We propose a new set of quadratic moment equations exploring 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 full information ML estimator. Monte Carlo experiment shows that the proposed GMM estimator performs well in finite samples.

The second chapter proposes a GMM estimator for the SAR model in a system of simultaneous equations with heteroskedastic disturbances. Besides linear moment conditions, the GMM estimator also utilizes quadratic moment conditions based on the covariance structure of model disturbances within and across equations. Compared with the QML approach considered in Yang and Lee (2014), the GMM estimator is easier to implement and robust under heteroskedasticity of an unknown form. We also derive a heteroskedasticity-robust estimator for the asymptotic covariance of the GMM estimator. Monte Carlo experiments show that the proposed GMM estimator performs well in finite samples.

Dedication

To my grandfather, Alber Garcia Quinderé.

Acknowledgements

The path to a Ph.D. is arduous. I would not have done it without help along the way. I have received both emotional and academic support throughout the years. It all started with my fifth grade teacher who said I was good at math and I was foolish enough to believe her.

My parents, siblings and grandparents never ceased to give me support. For example, my father, being an economics professor himself, taught me how to study. During my bachelors he advised me to develop the habit of studying multiple sources, especially when trying to understand a difficult concept.

In my two masters, my advisors where respectively Ivan Castelar and Carlos Martins-Filho. Both have invested immensely in my future. It was because of them that I chose to study Econometrics. They also guided me in obtain the required math background in my field.

Among the Economics department staff, Patricia Holcomb and Maria Olivares are my heroines. They have gone to great lengths to help me. However, I could never meet them before class, for I would always lose track of time when talking to them.

I am forever grateful to my advisor, Xiaodong Liu. He has taught me so much, including everything I know about Spatial Econometric Models. I hope to live up to all his teachings and expectations. He believed in me even during the difficult times when I thought I was not going to make it. That gave me the strength I needed to succeed.

I would also like to thank all my dissertation committee members; Jem Corcoran, Robert McNown, Scott Savage and Donald Waldman. Their enthusiasm about the subject, questions and suggestions where invaluable and helped make defending my dissertation quite enjoyable.

Contents

Tables

Table

Chapter 1

Efficient GMM Estimation of SAR Models with Endogenous Regressors

1.1 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; Kelejian and Prucha, 1999; 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 and is computational simple relative to the maximum likelihood (ML) estimator. 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.

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 exploring (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. The proposed GMM estimator is thus a "full information" estimator as it uses information across equations. Compare to other full information estimators for a system of simultaneous equations with spatial interdependence, the GMM estimator is more efficient than the three stage least squares (3SLS) estimator in Kelejian and Prucha (2004), and computationally simpler than the ML estimator in Yang and Lee (2014). With carefully constructed quadratic moment equations, the GMM estimator can be asymptotically as efficient as the full information ML estimator. 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 $\mathbf{A}^{(s)} = \mathbf{A} + \mathbf{A}'$, $\text{vec}_D(\mathbf{A}) = (a_{11}, \cdots, a_{nn})'$, and $\text{diag}(\mathbf{A}) = \text{diag}(a_{11}, \cdots, a_{nn})$. 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.

1.2 Model

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

$$\mathbf{y}_1 = \lambda_0 \mathbf{W} \mathbf{y}_1 + \phi_0 \mathbf{y}_2 + \mathbf{X}_1 \boldsymbol{\beta}_0 + \mathbf{u}_1, \tag{1.1}$$

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

$$\mathbf{y}_2 = \mathbf{X} \boldsymbol{\gamma}_0 + \mathbf{u}_2, \tag{1.2}$$

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 1.1 Let $u_{1,i}$ and $u_{2,i}$ denote, respectively, the i-th elements of $\mathbf{u}_1$ and $\mathbf{u}_2$. (i) $(u_{1,i}, u_{2,i})'$ is 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 $\eta > 0$.

[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.

[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.

Assumption 1.2 (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 1.3 (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 1.4 $\boldsymbol{\theta}_0$ is in the interior of a compact and convex parameter space Θ.

1.3 GMM Estimation

1.3.1 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 1.3, model (1.1) has a reduced form

$$\mathbf{y}_1 = \mathbf{S}^{-1}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{S}^{-1}\mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{S}^{-1}\mathbf{u}_1 + \phi_0\mathbf{S}^{-1}\mathbf{u}_2, \tag{1.3}$$

which implies that

$$\mathbf{W}\mathbf{y}_1 = \mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{G}\mathbf{u}_1 + \phi_0\mathbf{G}\mathbf{u}_2. \tag{1.4}$$

As $\mathbf{W}\mathbf{y}_1$ and $\mathbf{y}_2$ are endogenous, consistent estimation of (1.1) requires IVs for $\mathbf{W}\mathbf{y}_1$ and $\mathbf{y}_2$. From (2.8), 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 (1.2), $\mathbf{X}$ 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]

[3] The IV matrix $\mathbf{G}\mathbf{X}$ is not feasible as $\mathbf{G}$ involves the unknown parameter λ_0. Under Assumption 1.3, $\mathbf{G}\mathbf{X} = \mathbf{W}\mathbf{X} + \lambda_0\mathbf{W}^2\mathbf{X} + \lambda_0^2\mathbf{W}^3\mathbf{X} + \cdots$. 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 $\mathbf{Q}_2'\mathbf{u}_2(\boldsymbol{\delta})$. 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 $\mathbf{Q}_2'\mathbf{u}_2(\boldsymbol{\delta})$ if $\mathbf{Q}$ includes all linearly indepedent columns of $\mathbf{Q}_1$ and $\mathbf{Q}_2$.

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 (1.2) into (1.1) leads to a "pseudo" reduced form

$$\mathbf{y}_1 = \lambda_0 \mathbf{W}\mathbf{y}_1 + \phi_0 \mathbf{X}\boldsymbol{\gamma}_0 + \mathbf{X}_1\boldsymbol{\beta}_0 + \mathbf{u}_1 + \phi_0 \mathbf{u}_2. \tag{1.5}$$

By exploring the covariance structure of the spatially lagged dependent variable $\mathbf{W}\mathbf{y}_1$ and the disturbances of (1.5), 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 $\mathbf{W}$, $\mathbf{W}^2 - n^{-1}\mathrm{E}(\mathbf{W}^2)\mathbf{I}_n$, 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

$$\mathbf{g}(\boldsymbol{\theta}) = [\mathbf{g}_1(\boldsymbol{\theta})', \mathbf{g}_2(\boldsymbol{\theta})']', \tag{1.6}$$

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

Assumption 1.5 (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.

[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 1.4.2.

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

$$\tilde{\boldsymbol{\theta}}_{gmm} = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbf{g}(\boldsymbol{\theta})'\mathbf{F}'\mathbf{F}\mathbf{g}(\boldsymbol{\theta}), \tag{1.7}$$

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 GMM estimator depends on the choice of the weighting matrix as discussed in Section 1.4.1.

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 (1.2) and (2.7) 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 by

$$\phi = \phi_0 + (\lambda_0 - \lambda)c_1 \quad \text{and} \quad \boldsymbol{\beta} = \boldsymbol{\beta}_0 + (\lambda_0 - \lambda)\mathbf{c}_2 \tag{1.8}$$

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 (1.8), 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

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

[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 $(\phi_0, \boldsymbol{\beta}_0')' \neq \mathbf{0}$.

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

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

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

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

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

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

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 1.6 $\lim_{n\to\infty} n^{-1}\mathbf{Q}'\mathbf{X}$ 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) $(\sigma_1^2 + \phi_0\sigma_{12})\lim_{n\to\infty} n^{-1}\text{tr}(\mathbf{\Xi}_j^{(s)}\mathbf{G}) \neq 0$ for some $j \in \{1, \cdots, m\}$, and $\lim_{n\to\infty} n^{-1}[\text{tr}(\mathbf{\Xi}_1^{(s)}\mathbf{G}), \cdots, \text{tr}(\mathbf{\Xi}_m^{(s)}\mathbf{G})]'$ is linearly independent of

$$\lim_{n\to\infty} n^{-1}[\text{tr}(\mathbf{G}'\mathbf{\Xi}_1\mathbf{G}), \cdots, \text{tr}(\mathbf{G}'\mathbf{\Xi}_m\mathbf{G})]'.$$

(iii) $(\sigma_{12} + \phi_0\sigma_2^2)\lim_{n\to\infty} n^{-1}\text{tr}(\mathbf{\Xi}_j\mathbf{G}) \neq 0$ or $(\sigma_{12} + \phi_0\sigma_2^2)\lim_{n\to\infty} n^{-1}\text{tr}(\mathbf{\Xi}_j'\mathbf{G}) \neq 0$ for some $j \in \{1, \cdots, m\}$.

1.4 Asymptotic Properties

1.4.1 Consistency and Asymptotic Normality

The GMM estimator defined in (1.7) 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 1.1 and 1.2 in the Appendix, we have

$$\mathbf{\Omega} = \mathrm{Var}[\mathbf{g}(\boldsymbol{\theta}_0)] = \begin{bmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{12}' & \mathbf{\Omega}_{22} \end{bmatrix} \tag{1.9}$$

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

$$\mathbf{D} = -\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta}_0)] = [\mathbf{D}_1', \mathbf{D}_2']', \tag{1.10}$$

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 1.1 Suppose Assumptions 1.1-1.6 hold. Then $\tilde{\boldsymbol{\theta}}_{gmm}$ defined in (1.7) is a consistent estimator of $\boldsymbol{\theta}_0$ and has the following asymptotic distribution

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

where

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

with $\boldsymbol{\Omega}$ and $\mathbf{D}$ defined in (1.13) and (2.14) respectively.

Close inspection of $\mathrm{AsyVar}(\tilde{\boldsymbol{\theta}}_{gmm})$ 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 1.2 Suppose Assumptions 1.1-1.6 hold and $n^{-1}\hat{\mathbf{\Omega}}$ is a consistent estimator of $n^{-1}\mathbf{\Omega}$ defined in (1.13). Then,

$$\hat{\boldsymbol{\theta}}_{gmm} = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbf{g}(\boldsymbol{\theta})'\hat{\mathbf{\Omega}}^{-1}\mathbf{g}(\boldsymbol{\theta}) \tag{1.11}$$

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

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

where $\mathbf{D}$ is defined in (1.10). Furthermore

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

1.4.2 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 (1.11) reduces to the generalized spatial 3SLS in Kelejian and Prucha (2004) because

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

where $\hat{\mathbf{\Sigma}}$ is a consistent estimator of $\mathbf{\Sigma}$, $\mathbf{P} = \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2')'$, and

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

It follows from Proposition 1.2 that

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

As

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

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 $\mathbf{\Xi}_1, \cdots, \mathbf{\Xi}_m$. Following Lee (2007), our discussion on the asymptotic efficiency focuses on two cases: (i) $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')' \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}_n)$, and (ii) $\mathbf{\Xi}_j$ has a zero diagonal for all $j = 1, \cdots, m$. Let $\mathcal{P}$ be a subset of all $\mathbf{\Xi}$'s satisfying Assumption 1.5 such that $\mathrm{diag}(\mathbf{\Xi}) = \mathbf{0}$ for all $\mathbf{\Xi} \in \mathcal{P}$. 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

$$\mathbf{g}^*(\boldsymbol{\theta}) = [\mathbf{g}_1^*(\boldsymbol{\delta})', \mathbf{g}_2^*(\boldsymbol{\theta})']', \tag{1.12}$$

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),

$$\mathbf{\Omega}^* = \mathrm{Var}[\mathbf{g}^*(\boldsymbol{\theta}_0)] = \begin{bmatrix} \mathbf{\Sigma} \otimes (\mathbf{Q}^{*\prime}\mathbf{Q}^*) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_{22}^* \end{bmatrix} \tag{1.13}$$

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

$$\tilde{\boldsymbol{\theta}}_{bgmm} = \arg\min\nolimits_{\boldsymbol{\theta}\in\Theta} \mathbf{g}^*(\boldsymbol{\theta})'\mathbf{\Omega}^{*-1}\mathbf{g}^*(\boldsymbol{\theta}) \tag{1.14}$$

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

Proposition 1.3 Suppose Assumptions 1.1-1.6 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 (1.14) 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 (1.7).

(ii) Without the normality assumption on $\mathbf{u}$, the BGMM estimator defined in (1.14) with $\mathbf{Q}^* = [\mathbf{GX}, \mathbf{X}]$ and $\boldsymbol{\Xi}^* = \mathbf{G} - \text{diag}(\mathbf{G})$ is the most efficient one in the sub-class of GMM estimators defined in (1.7) with $\boldsymbol{\Xi}_j \in \mathcal{P}$ 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 $\boldsymbol{\Xi}^*$ 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 $\mathbf{G}(\lambda) = \mathbf{WS}(\lambda)^{-1}$, 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]

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

with the first order partial derivatives

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

and

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

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

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

[9] The detailed derivation of the log likelihood function and its partial derivatives can be found in Appendix 1.1.

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 treated as linear combinations of the moment functions $\mathbf{Q}^*(\lambda)'\mathbf{u}_1(\boldsymbol{\delta})$, $\mathbf{Q}^*(\lambda)'\mathbf{u}_2(\boldsymbol{\gamma})$, $\mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}^*(\lambda)\mathbf{u}_1(\boldsymbol{\delta})$, $\mathbf{u}_1(\boldsymbol{\delta})'\boldsymbol{\Xi}^*(\lambda)\mathbf{u}_2(\boldsymbol{\gamma})$, $\mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^*(\lambda)\mathbf{u}_1(\boldsymbol{\delta})$, and $\mathbf{u}_2(\boldsymbol{\gamma})'\boldsymbol{\Xi}^*(\lambda)\mathbf{u}_2(\boldsymbol{\gamma})$ with $\mathbf{Q}^*(\lambda) = [\mathbf{G}(\lambda)\mathbf{X}, \mathbf{X}]$ and

$$\boldsymbol{\Xi}^*(\lambda) = \mathbf{G}(\lambda) - n^{-1}\text{tr}(\mathbf{G}(\lambda))\mathbf{I}_n.$$

The optimal $\mathbf{Q}^*$ and $\boldsymbol{\Xi}^*$ are not feasible as $\mathbf{G}$ involves the unknown parameter λ_0. Suppose there exists a $\sqrt{n}$-consistent preliminary estimator $\hat{\lambda}$ for λ_0 (say, the 2SLS estimator with IV matrix $\mathbf{Q} = [\mathbf{W}\mathbf{X}, \mathbf{X}]$). Then, the feasible optimal $\hat{\mathbf{Q}}^*$ and $\hat{\boldsymbol{\Xi}}^*$ can obtained by replacing λ_0 in $\mathbf{Q}^*$ and $\boldsymbol{\Xi}^*$ by $\hat{\lambda}$. Furthermore, suppose $\hat{\sigma}_1^2, \hat{\sigma}_{12}, \hat{\sigma}_2^2$ are consistent preliminary estimators for $\sigma_1^2, \sigma_{12}, \sigma_2^2$. Then, $n^{-1}\hat{\boldsymbol{\Omega}}^*$ is a consistent estimator of $n^{-1}\boldsymbol{\Omega}^*$ defined in (1.13) with the unknown parameters $\lambda, \sigma_1^2, \sigma_{12}, \sigma_2^2$ in $\boldsymbol{\Omega}^*$ replaced by $\hat{\lambda}, \hat{\sigma}_1^2, \hat{\sigma}_{12}, \hat{\sigma}_2^2$. Then, the feasible BGMM estimator is given by

$$\hat{\boldsymbol{\theta}}_{bgmm} = \arg\min_{\boldsymbol{\theta}\in\Theta} \hat{\mathbf{g}}^*(\boldsymbol{\theta})'\hat{\boldsymbol{\Omega}}^{*-1}\hat{\mathbf{g}}^*(\boldsymbol{\theta}), \tag{1.15}$$

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

Proposition 1.4 Suppose Assumptions 1.1-1.6 hold, $\hat{\lambda}$ is a $\sqrt{n}$-consistent estimator of λ_0, and $\hat{\boldsymbol{\Sigma}}$ is a consistent estimator of $\boldsymbol{\Sigma}$. The feasible BGMM estimator $\hat{\boldsymbol{\theta}}_{bgmm}$ defined in (1.15) is asymptotically equivalent to the corresponding infeasible BGMM estimator $\tilde{\boldsymbol{\theta}}_{bgmm}$.

Under Assumption 1.3, $\mathbf{G} = \mathbf{W}\mathbf{S}^{-1} = \mathbf{W} + \lambda_0\mathbf{W}^2 + \lambda_0^2\mathbf{W}^3 + \cdots$. 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 $\mathbf{Q} = [\mathbf{W}\mathbf{X}, \cdots, \mathbf{W}^m\mathbf{X}, \mathbf{X}]$ and $\boldsymbol{\Xi}_1 = \mathbf{W}, \boldsymbol{\Xi}_2 = \mathbf{W}^2 - n^{-1}\text{tr}(\mathbf{W}^2)\mathbf{I}_n, \cdots, \boldsymbol{\Xi}_m = \mathbf{W}^m - n^{-1}\text{tr}(\mathbf{W}^m)\mathbf{I}_n$, for some fixed m.

1.5 Monte Carlo Experiments

We conduct a small Monte Carlo simulation experiment to study the finite sample performance of the proposed GMM estimator. The DGP considered in the experiment follows equations (1.1) and (1.2) with $K_1 = K_2 = 1$. In the DGP, we set $\lambda_0 = 0.6$ and $\boldsymbol{\gamma}_0 = (0, 1)'$, and generate $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ 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 (2.8), $\mathrm{E}(\mathbf{W}\mathbf{y}_1) = \mathbf{G}\mathbf{X}_1\boldsymbol{\beta}_0 + \phi_0\mathbf{G}\mathbf{X}\boldsymbol{\gamma}_0$. Therefore, $\phi_0 = \beta_0 = 0.5$ corresponds to the case that the IVs based on $\mathrm{E}(\mathbf{W}\mathbf{y}_1)$ are relatively informative and $\phi_0 = \beta_0 = 0.2$ corresponds to the case that the IVs based on $\mathrm{E}(\mathbf{W}\mathbf{y}_1)$ 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 $\hat{\mathbf{G}} = \mathbf{W}(\mathbf{I}_n - \hat{\lambda}\mathbf{W})^{-1}$, where $\hat{\lambda}$ is the 2SLS estimator of λ_0 using the IV matrix $\mathbf{Q} = [\mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}, \mathbf{X}]$. Let $\hat{\mathbf{Q}} = [\hat{\mathbf{G}}\mathbf{X}, \mathbf{X}]$. In the experiment, we consider the following estimators.

(a) The 2SLS estimator of equation (1.1) with the linear moment function $\hat{\mathbf{Q}}'\mathbf{u}_1(\boldsymbol{\delta})$.

(b) The 3SLS estimator of equations (1.1) and (1.2) with the linear moment function $(\mathbf{I}_2 \otimes \hat{\mathbf{Q}})'\mathbf{u}(\boldsymbol{\theta})$.

(c) The single-equation GMM (GMM-1) estimator of equation (1.1) with the linear moment function $\hat{\mathbf{Q}}'\mathbf{u}_1(\boldsymbol{\delta})$ and the quadratic moment function $\mathbf{u}_1(\boldsymbol{\delta})'[\hat{\mathbf{G}} - n^{-1}\mathrm{tr}(\hat{\mathbf{G}})\mathbf{I}_n]\mathbf{u}_1(\boldsymbol{\delta})$.

(d) The system GMM (GMM-2) estimator of equations (1.1) and (1.2) with the linear moment function $(\mathbf{I}_2 \otimes \hat{\mathbf{Q}})'\mathbf{u}(\boldsymbol{\theta})$ and the quadratic moment functions $\mathbf{u}_1(\boldsymbol{\delta})'[\hat{\mathbf{G}} - n^{-1}\mathrm{tr}(\hat{\mathbf{G}})\mathbf{I}_n]\mathbf{u}_1(\boldsymbol{\delta})$, $\mathbf{u}_1(\boldsymbol{\delta})'[\hat{\mathbf{G}} - n^{-1}\mathrm{tr}(\hat{\mathbf{G}})\mathbf{I}_n]\mathbf{u}_2(\gamma)$, $\mathbf{u}_2(\gamma)'[\hat{\mathbf{G}} - n^{-1}\mathrm{tr}(\hat{\mathbf{G}})\mathbf{I}_n]\mathbf{u}_1(\boldsymbol{\delta})$, and

$$\mathbf{u}_2(\gamma)'[\hat{\mathbf{G}} - n^{-1}\mathrm{tr}(\hat{\mathbf{G}})\mathbf{I}_n]\mathbf{u}_2(\gamma).$$

Although the 2SLS estimator and the single-equation GMM estimator only use "limited information" in equation (1.1) 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 (1.2) and thus may be desirable under certain circumstances. The estimation results of equation (1.1) are reported in Tables 1.1-1.4. 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.

(i) The 2SLS and 3SLS estimators of λ_0 are upwards biased with large SDs when the IVs for $\mathbf{W}\mathbf{y}_1$ are less informative. For example, when $n = 245$ and $\sigma_{12} = 0.1$, the 2SLS and 3SLS estimates of λ_0 reported in Table 1.3 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 $\mathbf{W}\mathbf{y}_1$ are less informative. When $n = 245$ and $\sigma_{12} = 0.1$, the GMM-1 estimates of λ_0 reported in Table 1.3 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 $\mathbf{W}\mathbf{y}_1$ are less informative. In Table 1.1, when $\sigma_{12} = 0.1$, the GMM-1 estimator reduces the SD of the 2SLS estimator by about 60%. In Table 3, 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 $\mathbf{W}\mathbf{y}_1$ 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 $\mathbf{W}\mathbf{y}_1$ are less informative. For example, in Table 1.2, 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 1.4, 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%.

1.6 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 (1.1) and (1.2) 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 (1.1), 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 (1.2) 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.

Appendix 1.1 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 (1.2) and (2.7), $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2')' = \boldsymbol{\mu}_y(\boldsymbol{\theta}_0) + \mathbf{R}(\phi_0,\lambda_0)\mathbf{u}$ 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 (1.1) and (1.2) is given by

$$\begin{aligned} L(\boldsymbol{\theta},\boldsymbol{\Sigma}) &= -n\ln(2\pi) - \frac{1}{2}\ln|\mathbf{R}(\phi,\lambda)(\boldsymbol{\Sigma}\otimes\mathbf{I}_n)\mathbf{R}'(\phi,\lambda)| \\ &\quad -\frac{1}{2}[\mathbf{y}-\boldsymbol{\mu}_y(\boldsymbol{\theta})]'[\mathbf{R}(\phi_0,\lambda_0)(\boldsymbol{\Sigma}\otimes\mathbf{I}_n)\mathbf{R}(\phi_0,\lambda_0)']^{-1}[\mathbf{y}-\boldsymbol{\mu}_y(\boldsymbol{\theta})]. \end{aligned}$$

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

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

and

$$\frac{\partial}{\partial(\boldsymbol{\Sigma}\otimes\mathbf{I}_n)^{-1}}L(\boldsymbol{\theta},\boldsymbol{\Sigma}) = \frac{1}{2}(\boldsymbol{\Sigma}\otimes\mathbf{I}_n) - \frac{1}{2}\mathbf{u}(\boldsymbol{\theta})\mathbf{u}(\boldsymbol{\theta})', \tag{1.16}$$

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

$$\begin{aligned}\frac{\partial}{\partial \lambda} L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) &= -\text{tr}(\mathbf{G}(\lambda)) + [(\phi \mathbf{X}\boldsymbol{\gamma} + \mathbf{X}_1\boldsymbol{\beta})'\mathbf{G}(\lambda)', \mathbf{0}]'(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) \\ &\quad + [(\mathbf{u}_1(\boldsymbol{\delta}) + \phi \mathbf{u}_2(\boldsymbol{\gamma}))'\mathbf{G}(\lambda)', \mathbf{0}]'(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta})\end{aligned} \tag{1.17}$$

and

$$\frac{\partial}{\partial \phi} L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = [\boldsymbol{\gamma}'\mathbf{X}', \mathbf{0}](\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}) + [\mathbf{u}_2(\boldsymbol{\gamma})', \mathbf{0}](\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}\mathbf{u}(\boldsymbol{\theta}). \tag{1.18}$$

From (1.16), the ML estimator for $\boldsymbol{\Sigma}$ is given by

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

Substitution of $\hat{\boldsymbol{\Sigma}}$ into (1.17) and (1.18) gives

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

and

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

Appendix 1.2 Lemmas

For ease of reference, we list some useful results without proofs. Lemmas 1.1-1.6 can be found (or are straightforward extensions of the lemmas) in Lee (2007). Lemma 1.7 is a special case of Lemma 3 in Yang and Lee (2014). Lemmas 1.8 and 1.9 are from Breusch et al. (1999).

Lemma 1.1 Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ nonstochastic matrices such that $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{B}) = 0$. Then,

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

Lemma 1.2 Let $\mathbf{A}$ be an $n \times n$ nonstochastic matrix and $\mathbf{c}$ be an $n \times 1$ nonstochastic vector. Then,

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

Lemma 1.3 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 1.4 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 1.5 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 $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Gamma(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\theta}}^* = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Gamma^*(\boldsymbol{\theta})$. If $n^{-1}[\Gamma(\boldsymbol{\theta}) - \Gamma^*(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ then both $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^*$ 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 $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \sqrt{n}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0) = \mathrm{o_p}(1)$.

Lemma 1.6 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 $\hat{\mathbf{G}}^*$ is obtained by replacing λ_0 in $\mathbf{G}^*$ by its $\sqrt{n}$-consistent estimator $\hat{\lambda}$. Suppose Assumption 1.3 holds. Then, $n^{-1}\mathbf{c}_1'(\hat{\mathbf{G}}^* - \mathbf{G})\mathbf{c}_2 = \mathrm{o_p}(1)$, $n^{-1/2}\mathbf{c}_1'(\hat{\mathbf{G}}^* - \mathbf{G})\mathbf{A}\mathbf{u}_r = \mathrm{o_p}(1)$, $n^{-1}\mathbf{u}_r'\mathbf{A}'(\hat{\mathbf{G}}^* - \mathbf{G})\mathbf{B}\mathbf{u}_s = \mathrm{o_p}(1)$, and $n^{-1/2}\mathbf{u}_r'(\hat{\mathbf{G}}^* - \mathbf{G})\mathbf{u}_s = \mathrm{o_p}(1)$, for $r, s = 1, 2$.

Lemma 1.7 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 1.8 Consider the set of moment conditions $\mathrm{E}[\mathbf{g}(\boldsymbol{\theta}_0)] = \mathbf{0}$ with $\mathbf{g}(\boldsymbol{\theta}) = [\mathbf{g}_1(\boldsymbol{\theta})', \mathbf{g}_2(\boldsymbol{\theta})']'$. 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 1.9 Consider the set of moment conditions $\mathrm{E}[\mathbf{g}(\boldsymbol{\theta}_0)] = \mathbf{0}$ with

$$\mathbf{g}(\boldsymbol{\theta}) = [\mathbf{g}_1(\boldsymbol{\theta})', \mathbf{g}_2(\boldsymbol{\theta})', \mathbf{g}_3(\boldsymbol{\theta})']'.$$

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 $\mathbf{g}_1$.

Appendix 1.3 Proofs

Proof of Proposition 1.1: 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 $\mathbf{F}$

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

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 1.3 and 1.4 that $n^{-1}l_1(\boldsymbol{\delta}) = o_p(1)$ and $n^{-1}q_1(\boldsymbol{\delta}) - n^{-1}\mathrm{E}[q_1(\boldsymbol{\delta})] = o_p(1)$ uniformly in $\boldsymbol{\Theta}$, where

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

Hence, $n^{-1}\mathbf{u}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^m f_{1,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta}) - n^{-1}\mathrm{E}[\mathbf{u}_1(\boldsymbol{\delta})'\left(\sum_{j=1}^m f_{1,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta})] = o_p(1)$ 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 1.3 and 1.4 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,

$$\begin{aligned}
&n^{-1}\mathbf{u}_2(\boldsymbol{\gamma})'\left(\sum_{j=1}^m f_{3,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta}) - n^{-1}\mathrm{E}[\mathbf{u}_2(\boldsymbol{\gamma})'\left(\sum_{j=1}^m f_{3,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_1(\boldsymbol{\delta})] = \mathrm{o_p}(1),\\
&n^{-1}\mathbf{u}_2(\boldsymbol{\gamma})'\left(\sum_{j=1}^m f_{4,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2(\boldsymbol{\gamma}) - n^{-1}\mathrm{E}[\mathbf{u}_2(\boldsymbol{\gamma})'\left(\sum_{j=1}^m f_{4,ij}\boldsymbol{\Xi}_j\right)\mathbf{u}_2(\boldsymbol{\gamma})] = \mathrm{o_p}(1),\\
&n^{-1}\mathbf{f}_{1,i\cdot}\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\delta}) - n^{-1}\mathrm{E}[\mathbf{f}_{1,i\cdot}\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\delta})] = \mathrm{o_p}(1), \quad \text{and}\\
&n^{-1}\mathbf{f}_{2,i\cdot}\mathbf{Q}'\mathbf{u}_2(\boldsymbol{\gamma}) - n^{-1}\mathrm{E}[\mathbf{f}_{2,i\cdot}\mathbf{Q}'\mathbf{u}_2(\boldsymbol{\gamma})] = \mathrm{o_p}(1)
\end{aligned}$$

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,

$$\frac{1}{n^2}\mathbf{g}(\boldsymbol{\theta})'\mathbf{F}'\mathbf{F}\mathbf{g}(\boldsymbol{\theta})$$

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 1.4. 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 $\hat{\boldsymbol{\theta}}$ follows by Theorem 15.1 of Peracchi (2001).

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

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

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

$$-\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{Q}'\mathbf{W}\mathbf{y}_1 & \mathbf{Q}'\mathbf{y}_2 & \mathbf{Q}'\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}'\mathbf{X} \\ \mathbf{u}_1(\boldsymbol{\delta})'\Xi_1^{(s)}\mathbf{W}\mathbf{y}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\Xi_1^{(s)}\mathbf{y}_2 & \mathbf{u}_1(\boldsymbol{\delta})'\Xi_1^{(s)}\mathbf{X}_1 & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_1(\boldsymbol{\delta})'\Xi_m^{(s)}\mathbf{W}\mathbf{y}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\Xi_m^{(s)}\mathbf{y}_2 & \mathbf{u}_1(\boldsymbol{\delta})'\Xi_m^{(s)}\mathbf{X}_1 & \mathbf{0} \\ \mathbf{u}_2(\gamma)'\Xi_1'\mathbf{W}\mathbf{y}_1 & \mathbf{u}_2(\gamma)'\Xi_1'\mathbf{y}_2 & \mathbf{u}_2(\gamma)'\Xi_1'\mathbf{X}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\Xi_1\mathbf{X} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_2(\gamma)'\Xi_m'\mathbf{W}\mathbf{y}_1 & \mathbf{u}_2(\gamma)'\Xi_m'\mathbf{y}_2 & \mathbf{u}_2(\gamma)'\Xi_m'\mathbf{X}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\Xi_m\mathbf{X} \\ \mathbf{u}_2(\gamma)'\Xi_1\mathbf{W}\mathbf{y}_1 & \mathbf{u}_2(\gamma)'\Xi_1\mathbf{y}_2 & \mathbf{u}_2(\gamma)'\Xi_1\mathbf{X}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\Xi_1'\mathbf{X} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_2(\gamma)'\Xi_m\mathbf{W}\mathbf{y}_1 & \mathbf{u}_2(\gamma)'\Xi_m\mathbf{y}_2 & \mathbf{u}_2(\gamma)'\Xi_m\mathbf{X}_1 & \mathbf{u}_1(\boldsymbol{\delta})'\Xi_m'\mathbf{X} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{u}_2(\gamma)'\Xi_1^{(s)}\mathbf{X} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{u}_2(\gamma)'\Xi_m^{(s)}\mathbf{X} \end{bmatrix}.$$

Using Lemmas 1.3 and 1.4, it follows by a similar argument in the proof of Proposition 1 in Lee (2007) that $-n^{-1}\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\hat{\boldsymbol{\theta}}) - n^{-1}\mathbf{D} = \mathrm{o_p}(1)$ and $-n^{-1}\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\bar{\boldsymbol{\theta}}) - n^{-1}\mathbf{D} = \mathrm{o_p}(1)$ with $\mathbf{D}$ given by (1.10). By Lemma 1.7 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 (1.13). The desired result follows. ■

Proof of Proposition 1.2: Note that

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

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 1.1. On the other hand,

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

where $\|\cdot\|$ is the Euclidean norm for vectors and matrices. By a similar argument in the proof of Proposition 1.1, 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, $\left\| n^{-1}\mathbf{g}(\boldsymbol{\theta})'(\hat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\mathbf{g}(\boldsymbol{\theta}) \right\| = \mathrm{o_p}(1)$ uniformly in Θ. The consistency of $\hat{\boldsymbol{\theta}}_{gmm}$ follows.

For the asymptotic normality of $\sqrt{n}(\hat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0)$, note that from the proof of Proposition 1.1 we have $-n^{-1}\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\hat{\boldsymbol{\theta}}_{gmm}) - n^{-1}\mathbf{D} = \mathrm{o_p}(1)$, since $\hat{\boldsymbol{\theta}}_{gmm}$ is consistent. Let $\bar{\boldsymbol{\theta}} = t\hat{\boldsymbol{\theta}}_{gmm} + (1-t)\boldsymbol{\theta}_0$ for some $t \in [0,1]$, then by the mean value theorem,

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

which concludes the first part of the proof, since in the proof of Proposition 1.1 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 $\bar{\boldsymbol{\theta}} = t\hat{\boldsymbol{\theta}}_{gmm} + (1-t)\boldsymbol{\theta}_0$

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

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

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

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

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

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

Proof of Proposition 1.3: 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 (1.6) given $\mathbf{g}^*$ defined in (2.4) 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 (1.10). Based on Lemma 1.8 $\mathbf{g}$ is redundant given $\mathbf{g}^*$. Furthermore, Lemma 1.9 tells us that any subset of $\mathbf{g}$ is redundant given $\mathbf{g}^*$. ■

Proof of Proposition 1.4: To show the desired result, we only need to show $\hat{\mathbf{\Gamma}}(\boldsymbol{\theta}) = \hat{\mathbf{g}}^*(\boldsymbol{\theta})'\hat{\mathbf{\Omega}}^{*-1}\hat{\mathbf{g}}^*(\boldsymbol{\theta})$ and $\mathbf{\Gamma}(\boldsymbol{\theta}) = \mathbf{g}^*(\boldsymbol{\theta})'\mathbf{\Omega}^{*-1}\mathbf{g}^*(\boldsymbol{\theta})$ satisfy the conditions of Lemma 1.5. First, $n^{-1}[\hat{\mathbf{g}}_1^*(\boldsymbol{\theta}) - \mathbf{g}_1^*(\boldsymbol{\theta})] = n^{-1}[\mathbf{I}_2 \otimes (\hat{\mathbf{Q}}^* - \mathbf{Q}^*)]'\mathbf{u}(\boldsymbol{\theta})$, $n^{-1}[\hat{\mathbf{g}}_{2,rs}^*(\boldsymbol{\theta}) - \mathbf{g}_{2,rs}^*(\boldsymbol{\theta})] = n^{-1}\mathbf{u}_r(\boldsymbol{\theta})'(\hat{\mathbf{\Xi}}^* - \mathbf{\Xi}^*)\mathbf{u}_s(\boldsymbol{\theta})$,

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

and

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

where $\mathbf{Q}^* = [\mathbf{G}\mathbf{X}, \mathbf{X}]$, $\mathbf{\Xi}^*$ 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 $\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{u}_2(\gamma) = -[\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{X}]$. By inspection of each term of the above matrices, we conclude $n^{-1}[\hat{\mathbf{g}}^*(\boldsymbol{\theta}) - \mathbf{g}^*(\boldsymbol{\theta})] = o_p(1)$, $n^{-1}[\frac{\partial}{\partial \boldsymbol{\theta}'}\hat{\mathbf{g}}^*(\boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}^*(\boldsymbol{\theta})] = o_p(1)$ and $n^{-1}[\frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\hat{\mathbf{g}}^*(\boldsymbol{\theta}) - \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\mathbf{g}^*(\boldsymbol{\theta})] = o_p(1)$ uniformly in $\mathbf{\Theta}$ by Lemma 1.6. Second, as $\hat{\mathbf{G}} - \mathbf{G} = (\hat{\lambda} - \lambda_0)\mathbf{G}^2 + (\hat{\lambda} - \lambda_0)^2\hat{\mathbf{G}}\mathbf{G}^2$, we have $n^{-1}\mathrm{tr}(\hat{\mathbf{\Xi}}^*\hat{\mathbf{\Xi}}^*) - n^{-1}\mathrm{tr}(\mathbf{\Xi}^*\mathbf{\Xi}^*) = o_p(1)$ and $n^{-1}\mathrm{tr}(\hat{\mathbf{\Xi}}^{*\prime}\hat{\mathbf{\Xi}}^*) - n^{-1}\mathrm{tr}(\mathbf{\Xi}^{*\prime}\mathbf{\Xi}^*) = o_p(1)$. Therefore, as $\hat{\mathbf{\Sigma}}$ is a consistent estimator of $\mathbf{\Sigma}$, we have $n^{-1}(\hat{\mathbf{\Omega}} - \mathbf{\Omega}^*) = o_p(1)$. Hence, we can conclude that $n^{-1}[\hat{\Gamma}(\boldsymbol{\theta}) - \Gamma(\boldsymbol{\theta})] = o_p(1)$ and $n^{-1}[\frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\hat{\Gamma}(\boldsymbol{\theta}) - \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\Gamma(\boldsymbol{\theta})] = o_p(1)$ uniformly in $\mathbf{\Theta}$. Finally, since $n^{-1/2}\mathbf{g}^*(\boldsymbol{\theta}_0) = O_p(1)$ by a similar argument in the proof of Proposition 1.1 and $n^{-1/2}[\hat{\mathbf{g}}^*(\boldsymbol{\theta}_0) -$

$\mathbf{g}^*(\boldsymbol{\theta}_0)] = \mathrm{o_p}(1)$ by Lemma 1.6,

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

The desired result follows. ■

Appendix 1.4 Tables

Table 1.1: 2SLS, 3SLS and GMM Estimation ($n = 245$)

	$\lambda_0 = 0.3$	$\phi_0 = 1.0$	$\beta_0 = 0.5$
$\sigma_{12} = 0.1$			
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]
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]
$\sigma_{12} = 0.5$			
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]
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]
$\sigma_{12} = 0.9$			
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]
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]

Mean(SD)[RMSE]

Table 1.2: 2SLS, 3SLS and GMM Estimation ($n = 490$)

	$\lambda_0 = 0.3$	$\phi_0 = 1.0$	$\beta_0 = 0.5$
$\sigma_{12} = 0.1$			
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]
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]
$\sigma_{12} = 0.5$			
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]
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]
$\sigma_{12} = 0.9$			
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]
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]

Mean(SD)[RMSE]

Table 1.3: 2SLS, 3SLS and GMM Estimation ($n = 245$)

	$\lambda_0 = 0.3$	$\phi_0 = 1.0$	$\beta_0 = 0.2$
$\sigma_{12} = 0.1$			
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]
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]
$\sigma_{12} = 0.5$			
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]
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]
$\sigma_{12} = 0.9$			
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]
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]

Mean(SD)[RMSE]

Table 1.4: 2SLS, 3SLS and GMM Estimation ($n = 490$)

	$\lambda_0 = 0.3$	$\phi_0 = 1.0$	$\beta_0 = 0.2$
$\sigma_{12} = 0.1$			
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]
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]
$\sigma_{12} = 0.5$			
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]
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]
$\sigma_{12} = 0.9$			
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]
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]

Mean(SD)[RMSE]

Chapter 2

GMM Estimation of SAR Simultaneous Equation Models with Unknown Heteroskedasticity

2.1 Introduction

The spatial autoregressive (SAR) model introduced by Cliff and Ord (1973, 1981) has recently received considerable attention in different fields of economics as it provides a convenient framework to model the interaction between economic agents. However, with a few exceptions (e.g., Kelejian and Prucha, 2004; Baltagi and Pirotte, 2011; Yang and Lee, 2014), most theoretical works in the spatial econometrics literature focus on the single-equation SAR model, which assumes that an economic agent's choice (or outcome) in a certain activity is isolated from her and other agents' choices (or outcomes) in related activities. This restrictive assumption potentially limits the usefulness of the SAR model in many contexts.

To incorporate the interdependence of economic agents' choices and outcomes across different activities, Kelejian and Prucha (2004) extends the single-equation SAR model to the simultaneous-equation SAR model. They propose both limited information two stage least squares (2SLS) and full information three stage least squares (3SLS) estimators for the estimation of model parameters and establish the asymptotic properties of the estimators. In a recent paper, Yang and Lee (2014) study the identification and estimation of the simultanenous-equation SAR model by the full information quasi-maximum likelihood (QML) approach. They give identification conditions for the simultanenous-equation SAR model that are analogous to the rank and order conditions for the classical simultanenous-equation model and derive asymptotic properties of the QML estimator. The QML estimator is asymptotically more effecient than the 3SLS estimator as the former implicitly uses additional information on the covariance sturcture of model disturbances.

In this paper, we propose a generalized method of moments (GMM) estimator for the identification and estimation of simultaneous-equation SAR models with heteroskedastic disturbances. Similar to the GMM estimator proposed by Lee (2007) and Lin and Lee (2010) for single-equation SAR models, the GMM estimator utilizes both **linear moment conditions** based on the orthogonality condition between the instrumental variable (IV) and model disturbances and **quadratic moment conditions** based on the covariance structure of model disturbances. While the single-equation GMM estimator can be considered as an equation-by-equation limited information estimator for a system of simultaneous equations,[1] the simultaneous-equation GMM estimator proposed in this paper exploits the correlation structure of disturbances within and across equations and thus is a full information estimator. We study the identification of model parameters under the GMM framework and derive asymptotic properties of the GMM estimator under heteroskedasticity of an unknown form. Furthermore, we propose a heteroskedascity-robust estimator for the asymptotic covariance of the GMM estimator in the spirit of White (1980). The GMM estimator is asymptotically more efficient than the 3SLS estimator. Compared with the QML estimator considered in Yang and Lee (2014), the GMM estimator is easier to implement and robust under heteroskedasticity. Monte Carlo experiments show that the proposed GMM estimator performs well in finite samples.

The remaining of this paper is organized as follows. In Section 2.2, we describe the model and give the moment conditions used to construct the GMM estimator. In Section 2.3, we establish the identification for the model under the GMM framework. We derive the asymptotic properties of the GMM estimator in Section 2.4. Results of Monte Carlo simulation experiments are reported in Section 2.5. Section 2.6 briefly concludes. Proofs are collected in the Appendix.

Throughout the paper, we adopt the following notation. For an $n \times n$ matrix $\mathbf{A} = [a_{ij}]_{i,j=1,\cdots,n}$,

[1] To apply the single-equation GMM approach in Lee (2007) and Lin and Lee (2010) to estimate an equation in the simultaneous-equation SAR model, both the optimal GMM weighting matrix and the estimator for the asymptotic covariance of the GMM estimator need to be adjusted for the additional endogenous regressors in the equation.

let $\mathbf{A}^{(s)} = \mathbf{A} + \mathbf{A}'$ and diag($\mathbf{A}$) denote an $n \times n$ diagonal matrix with the i-th diagonal element being a_{ii}, i.e., $\text{diag}(\mathbf{A}) = \text{diag}(a_{11}, \cdots, a_{nn})$. For an $n \times m$ matrix $\mathbf{B} = [b_{ij}]$, the vectorization of $\mathbf{B}$ is denoted by $\text{vec}(\mathbf{B}) = (b_{11}, \cdots, b_{n1}, b_{12}, \cdots, b_{nm})'$.[2] Let $\mathbf{I}_n$ denote the $n \times n$ identity matrix and $\mathbf{i}_{n,k}$ denote the k-th column of $\mathbf{I}_n$.

2.2 Model and Moment Conditions

The model considered in this paper is described by a system of m simultaneous equations for n cross sectional units,

$$\mathbf{Y} = \mathbf{Y}\boldsymbol{\Gamma}_0 + \mathbf{W}\mathbf{Y}\boldsymbol{\Lambda}_0 + \mathbf{X}\mathbf{B}_0 + \mathbf{U}, \tag{2.1}$$

where $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_m]$ is an $n \times m$ matrix of endogenous variables, $\mathbf{X}$ is an $n \times K_X$ matrix of exogenous variables, and $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_m]$ is an $n \times m$ matrix of disturbances.[3] $\mathbf{W}$ is an $n \times n$ nonstochastic matrix of spatial weights, with its (i, j)-th element represents the proximity between cross sectional units i and j.[4] The diagonal elements of $\mathbf{W}$ are normalized to be zeros. In the literature, $\mathbf{WY}$ is usually referred to as the spatial lag. $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Lambda}_0$ and $\mathbf{B}_0$ are, respectively, $m \times m$, $m \times m$ and $K_X \times m$ matrices of true parameters in the data generating process (DGP). The diagonal elements of $\boldsymbol{\Gamma}_0$ are normalized to be zeros.

In general, the identification of simultaneous-equation models needs exclusion restrictions. Let $\boldsymbol{\gamma}_{k,0}$, $\boldsymbol{\lambda}_{k,0}$ and $\boldsymbol{\beta}_{k,0}$ denote vectors of nonzero elements of the k-th columns of $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Lambda}_0$ and $\mathbf{B}_0$ respectively under some exclusion restrictions. Let $\mathbf{Y}_k$, $\bar{\mathbf{Y}}_k$ and $\mathbf{X}_k$ denote the corresponding matrices containing columns of $\mathbf{Y}$ (except $\mathbf{y}_k$), $\bar{\mathbf{Y}} = \mathbf{WY}$ and $\mathbf{X}$ that appear in the k-th equation. Then, the k-th equation of model (2.1) is

$$\mathbf{y}_k = \mathbf{Y}_k\boldsymbol{\gamma}_{k,0} + \bar{\mathbf{Y}}_k\boldsymbol{\lambda}_{k,0} + \mathbf{X}_k\boldsymbol{\beta}_{k,0} + \mathbf{u}_k.$$

Throughout the paper, we maintain the following assumptions regarding the DGP.

[2] If $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are conformable matrices, then $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}(\mathbf{B})$, where $\otimes$ denotes the Kronecker product.

[3] In this paper, all variables are allowed to depend on the sample size, i.e., are allowed to formulate triangular arrays as in Kelejian and Prucha (2010). Nevertheless, we suppress the subscript n to simplify the notation.

[4] For SAR models, the notion of proximity is not limited to the geographical sense. It can be economic proximity, technology proximity, or social proximity. Hence the SAR has a broad range of applications.

Assumption 2.1 Let u_{ik} denote the (i,k)-th element of $\mathbf{U}$ and $\mathbf{u}$ denote the vectorization of $\mathbf{U}$, i.e., $\mathbf{u} = \mathrm{vec}(\mathbf{U})$. (i) $(u_{i1}, \cdots, u_{im})$ are independently distributed across i with zero mean. (ii)

$$\boldsymbol{\Sigma} \equiv \mathrm{E}(\mathbf{u}\mathbf{u}') = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1m} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{m1} & \cdots & \boldsymbol{\Sigma}_{mm} \end{bmatrix}$$

is nonsingular, with $\boldsymbol{\Sigma}_{kl} = \boldsymbol{\Sigma}_{lk} = \mathrm{diag}(\sigma_{1,kl}, \cdots, \sigma_{n,kl})$. (iii) $\mathrm{E}|u_{ik}u_{il}u_{is}u_{it}|^{1+\eta}$ is bounded for any $i = 1, \cdots, n$ and $k, l, s, t = 1, \cdots, m$, for some positive constant η.

Assumption 2.2 The elements of $\mathbf{X}$ are uniformly bounded constants. $\mathbf{X}$ has full column rank K_X. $\lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{X}$ exists and is nonsingular.

Assumption 2.3 $\boldsymbol{\Gamma}_0$ is nonsingular with a zero diagonal. $\rho(\boldsymbol{\Lambda}_0(\mathbf{I}_m - \boldsymbol{\Gamma}_0)^{-1}) < 1/\rho(\mathbf{W})$ where $\rho(\cdot)$ denotes the spectral radius of a square matrix.

Assumption 2.4 $\mathbf{W}$ has a zero diagonal. The row and column sums of $\mathbf{W}$ and $(\mathbf{I}_{mn} - \boldsymbol{\Gamma}_0' \otimes \mathbf{I}_n - \boldsymbol{\Lambda}_0' \otimes \mathbf{W})^{-1}$ are uniformly bounded in absolute value.

Assumption 2.5 $\boldsymbol{\theta}_{k,0} = (\boldsymbol{\gamma}_{k,0}', \boldsymbol{\lambda}_{k,0}', \boldsymbol{\beta}_{k,0}')'$ is in the interior of a compact and convex parameter space for $k = 1, \cdots, m$.

The above assumptions are based on some standard assumptions in the literature of SAR models (see, e.g., Kelejian and Prucha, 2004; Lee, 2007; Lin and Lee, 2010). In particular, Assumption 2.3 is from Yang and Lee (2014). Under this assumption, $\mathbf{I}_{mn} - (\boldsymbol{\Gamma}_0' \otimes \mathbf{I}_n) - (\boldsymbol{\Lambda}_0' \otimes \mathbf{W})$ is nonsingular, and hence the simultaneous-equation SAR model (2.1) has a well defined reduced form

$$\mathbf{y} = [\mathbf{I}_{mn} - (\boldsymbol{\Gamma}_0' \otimes \mathbf{I}_n) - (\boldsymbol{\Lambda}_0' \otimes \mathbf{W})]^{-1}[(\mathbf{B}_0' \otimes \mathbf{I}_n)\mathbf{x} + \mathbf{u}], \tag{2.2}$$

where $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$ and $\mathbf{x} = \mathrm{vec}(\mathbf{X})$. Note that, when $m = 1$, we have $\boldsymbol{\Gamma}_0 = 0$ and $\boldsymbol{\Lambda}_0 = \lambda_{11,0}$. Then, $\rho(\boldsymbol{\Lambda}_0(\mathbf{I}_m - \boldsymbol{\Gamma}_0)^{-1}) < 1/\rho(\mathbf{W})$ becomes the familiar parameter space constraint $|\lambda_{11,0}| < 1/\rho(\mathbf{W})$ for the single-equation SAR model.

Following Lee (2007) and Lin and Lee (2010), for the estimation of the simultaneous-equation SAR model (2.1), we consider both linear moment conditions

$$\mathrm{E}(\mathbf{Q}'\mathbf{u}_k) = \mathbf{0}, \tag{2.3}$$

where $\mathbf{Q}$ is an $n \times K_Q$ matrix of IVs, and quadratic moment conditions

$$\mathrm{E}(\mathbf{u}_k'\mathbf{\Xi}_r\mathbf{u}_l) = \mathrm{tr}(\mathbf{\Xi}_r\mathbf{\Sigma}_{kl}), \quad \text{for } r = 1, \cdots, p,$$

where $\mathbf{\Xi}_r$'s are $n \times n$ constant matrices. Note that, if the diagonal elements of $\mathbf{\Xi}_r$'s are zero, then the quadratic moment conditions become

$$\mathrm{E}(\mathbf{u}_k'\mathbf{\Xi}_r\mathbf{u}_l) = 0, \quad \text{for } r = 1, \cdots, p. \tag{2.4}$$

As an example, we could use $\mathbf{Q} = [\mathbf{WX}, \cdots, \mathbf{W}^p\mathbf{X}]$ and $\mathbf{\Xi}_1 = \mathbf{W}, \mathbf{\Xi}_2 = \mathbf{W}^2 - \mathrm{diag}(\mathbf{W}^2), \cdots, \mathbf{\Xi}_p = \mathbf{W}^p - \mathrm{diag}(\mathbf{W}^p)$, where p is some predetermined positive integer, to construct the linear and quadratic moment conditions. The quadratic moment condition (2.4) exploits the covariance structure of model disturbances both within and across equations, and hence is more general than the quadratic moment condition used by the single-equation GMM estimator in Lee (2007) and Lin and Lee (2010).

Let the residual function for the k-th equation be

$$\mathbf{u}_k(\boldsymbol{\theta}_k) = \mathbf{y}_k - \mathbf{Y}_k\boldsymbol{\gamma}_{k,0} - \bar{\mathbf{Y}}_k\boldsymbol{\lambda}_{k,0} - \mathbf{X}_k\boldsymbol{\beta}_{k,0}.$$

Then, the empirical linear moment function based on (2.3) can be written as

$$\mathbf{g}_{1,k} \equiv \mathbf{g}_{1,k}(\boldsymbol{\theta}_k) = \mathbf{Q}'\mathbf{u}_k(\boldsymbol{\theta}_k) \tag{2.5}$$

and the empirical quadratic moment function based on (2.4) can be written as

$$\mathbf{g}_{2,kl} \equiv \mathbf{g}_{2,kl}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_l) = [\mathbf{\Xi}_1'\mathbf{u}_k(\boldsymbol{\theta}_k), \cdots, \mathbf{\Xi}_p'\mathbf{u}_k(\boldsymbol{\theta}_k)]'\mathbf{u}_l(\boldsymbol{\theta}_l) \tag{2.6}$$

for $k, l = 1, \cdots, m$. Combining both linear and quadratic moment functions by defining

$$\mathbf{g}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{g}_1(\boldsymbol{\theta}) \\ \mathbf{g}_2(\boldsymbol{\theta}) \end{bmatrix},$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \cdots, \boldsymbol{\theta}_m')'$, $\mathbf{g}_1(\boldsymbol{\theta}) = (\mathbf{g}_{1,1}', \cdots, \mathbf{g}_{1,m}')'$, and $\mathbf{g}_2(\boldsymbol{\theta}) = (\mathbf{g}_{2,11}', \cdots, \mathbf{g}_{2,1m}', \mathbf{g}_{2,21}', \cdots, \mathbf{g}_{2,mm}')'$. The identification and estimation of the simultaneous-equation SAR model (2.1) is based on the moment conditions $\mathrm{E}[\mathbf{g}(\boldsymbol{\theta}_0)] = \mathbf{0}$. We maintain the following assumption regarding the moment conditions.

Assumption 2.6 (i) The elements of $\mathbf{Q}$ are uniformly bounded. (ii) The diagonal elements of $\boldsymbol{\Xi}_r$ are zeros, and the row and column sums of $\boldsymbol{\Xi}_r$ are uniformly bounded, for $r = 1, \cdots, p$. (iii) Let $\boldsymbol{\Omega} = \mathrm{Var}[\mathbf{g}(\boldsymbol{\theta}_0)]$. $\lim_{n\to\infty} n^{-1}\boldsymbol{\Omega}$ exists and is nonsingular.

2.3 Identification

Following Yang and Lee (2014), we establish the identification of the simultaneous-equation SAR model in two steps. In the first step, we consider the identification of the reduced form parameters. In the second step, we recover the structural parameters from the reduced form parameters.

2.3.1 Identification of the "pseudo" reduced form parameters

When $\boldsymbol{\Gamma}_0$ is nonsingular, the simultaneous-equation SAR model (2.1) has a "pseudo" reduced form

$$\mathbf{Y} = \mathbf{W}\mathbf{Y}\boldsymbol{\Psi}_0 + \mathbf{X}\boldsymbol{\Pi}_0 + \mathbf{V}, \tag{2.7}$$

where $\boldsymbol{\Psi}_0 = \boldsymbol{\Lambda}_0(\mathbf{I}_m - \boldsymbol{\Gamma}_0)^{-1}$, $\boldsymbol{\Pi}_0 = \mathbf{B}_0(\mathbf{I}_m - \boldsymbol{\Gamma}_0)^{-1}$, and $\mathbf{V} = \mathbf{U}(\mathbf{I}_m - \boldsymbol{\Gamma}_0)^{-1}$. Model (2.7) has the specification of a multivariate SAR model (see, Yang and Lee, 2014; Liu, 2015). First, we consider the identification of the "pseudo" reduced form parameters $\boldsymbol{\Psi}_0 = [\psi_{lk,0}]$ and $\boldsymbol{\Pi}_0 = [\boldsymbol{\pi}_{1,0}, \cdots, \boldsymbol{\pi}_{m,0}]$ under the GMM framework.

The k-th equation in model (2.7) is given by

$$\mathbf{y}_k = \sum_{l=1}^{m} \psi_{lk,0}\mathbf{W}\mathbf{y}_l + \mathbf{X}\boldsymbol{\pi}_{k,0} + \mathbf{v}_k,$$

where

$$\mathbf{W}\mathbf{y}_l = \mathbf{H}_l(\mathbf{B}_0' \otimes \mathbf{I}_n)\mathbf{x} + \mathbf{H}_l\mathbf{v} \tag{2.8}$$

with $\mathbf{H}_l = (\mathbf{i}'_{m,l} \otimes \mathbf{W})[\mathbf{I}_{mn} - (\mathbf{\Psi}'_0 \otimes \mathbf{W})]^{-1}$, $\mathbf{x} = \mathrm{vec}(\mathbf{X})$, and $\mathbf{v} = \mathrm{vec}(\mathbf{V})$. Hence, the residual function for the k-th equation can be written as

$$\mathbf{v}_k(\boldsymbol{\delta}_k) = \mathbf{y}_k - \sum_{l=1}^{m} \psi_{lk} \mathbf{W} \mathbf{y}_l - \mathbf{X}\boldsymbol{\pi}_k = \mathbf{d}_k(\boldsymbol{\delta}_k) + \mathbf{v}_k + \sum_{l=1}^{m} (\psi_{lk,0} - \psi_{lk}) \mathbf{H}_l \mathbf{v}, \tag{2.9}$$

where $\boldsymbol{\delta}_k = (\psi_{1k}, \cdots, \psi_{mk}, \boldsymbol{\pi}'_k)'$ and

$$\mathbf{d}_k(\boldsymbol{\delta}_k) = \sum_{l=1}^{m} (\psi_{lk,0} - \psi_{lk}) \mathrm{E}(\mathbf{W}\mathbf{y}_l) + \mathbf{X}(\boldsymbol{\pi}_{k,0} - \boldsymbol{\pi}_k).$$

The "pseudo" reduced form parameters in model (2.7) can be identified by the moment conditions described in the previous section. Similar to (2.5) and (2.6), the linear moment function can be written as

$$\mathbf{f}_{1,k} \equiv \mathbf{f}_{1,k}(\boldsymbol{\delta}_k) = \mathbf{Q}' \mathbf{v}_k(\boldsymbol{\delta}_k)$$

and the quadratic moment function can be written as

$$\mathbf{f}_{2,kl} \equiv \mathbf{f}_{2,kl}(\boldsymbol{\delta}_k, \boldsymbol{\delta}_l) = [\mathbf{\Xi}'_1 \mathbf{v}_k(\boldsymbol{\delta}_k), \cdots, \mathbf{\Xi}'_p \mathbf{v}_k(\boldsymbol{\delta}_k)]' \mathbf{v}_l(\boldsymbol{\delta}_l)$$

for $k, l = 1, \cdots, m$. Let $\mathbf{f}(\boldsymbol{\delta}) = [\mathbf{f}_1(\boldsymbol{\delta})', \mathbf{f}_2(\boldsymbol{\delta})']'$, where $\boldsymbol{\delta} = (\boldsymbol{\delta}'_1, \cdots, \boldsymbol{\delta}'_m)'$, $\mathbf{f}_1(\boldsymbol{\delta}) = (\mathbf{f}'_{1,1}, \cdots, \mathbf{f}'_{1,m})'$, and $\mathbf{f}_2(\boldsymbol{\delta}) = (\mathbf{f}'_{2,11}, \cdots, \mathbf{f}'_{2,1m}, \mathbf{f}'_{2,21}, \cdots, \mathbf{f}'_{2,mm})'$. For $\boldsymbol{\delta}_0$ to be identified by the moment conditions $\mathrm{E}[\mathbf{f}(\boldsymbol{\delta}_0)] = \mathbf{0}$, the moment equations $\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{f}(\boldsymbol{\delta})] = \mathbf{0}$ need to have a unique solution at $\boldsymbol{\delta} = \boldsymbol{\delta}_0$ (Hansen, 1982).

It follows from (2.9) that

$$\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{f}_{1,k}(\boldsymbol{\delta}_k)] = \lim_{n\to\infty} n^{-1}\mathbf{Q}'\mathbf{d}_k(\boldsymbol{\delta}_k) = \lim_{n\to\infty} n^{-1}\mathbf{Q}'[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \cdots, \mathrm{E}(\mathbf{W}\mathbf{y}_m), \mathbf{X}](\boldsymbol{\delta}_{k,0} - \boldsymbol{\delta}_k)$$

for $k = 1, \cdots, m$. The linear moment equation, $\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{f}_{1,k}(\boldsymbol{\delta}_k)] = \mathbf{0}$, has a unique solution at $\boldsymbol{\delta}_k = \boldsymbol{\delta}_{k,0}$, if $\mathbf{Q}'[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \cdots, \mathrm{E}(\mathbf{W}\mathbf{y}_m), \mathbf{X}]$ has full column rank for n sufficiently large. A necessary condition for this rank condition is that $[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \cdots, \mathrm{E}(\mathbf{W}\mathbf{y}_m), \mathbf{X}]$ has full column rank of $m + K_X$ and $\mathrm{rank}(\mathbf{Q}) \geq m + K_X$ for n sufficiently large.

If, however, $[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \cdots, \mathrm{E}(\mathbf{W}\mathbf{y}_m), \mathbf{X}]$ does not have full column rank,[5] then the model may still be identifiable via the quadratic moment condition. Suppose for some $\bar{m} \in \{0, 1, \cdots, m-1\}$,

[5] From (2.8), $\mathrm{E}(\mathbf{W}\mathbf{y}_k) = \mathbf{H}_k(\mathbf{B}'_0 \otimes \mathbf{I}_n)\mathbf{x}$. For example, if $\mathbf{B}_0 = \mathbf{0}$, then $\mathrm{E}(\mathbf{W}\mathbf{y}_k) = \mathbf{0}$ for $k = 1, \ldots, m$, and thus $[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \ldots, \mathrm{E}(\mathbf{W}\mathbf{y}_m), \mathbf{X}]$ does not have full column rank.

$\mathrm{E}(\mathbf{W}\mathbf{y}_l)$ and the columns of $[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \cdots, \mathrm{E}(\mathbf{W}\mathbf{y}_{\bar{m}}), \mathbf{X}]$ are linearly dependent for some $l \in \{\bar{m}+1, \cdots, m\}$,[6] i.e., $\mathrm{E}(\mathbf{W}\mathbf{y}_l) = \sum_{k=1}^{\bar{m}} c_{1,kl}\mathrm{E}(\mathbf{W}\mathbf{y}_k) + \mathbf{X}\mathbf{c}_{2,l}$ for some vector of constants

$$(c_{1,1l}, \cdots, c_{1,\bar{m}l}, \mathbf{c}_{2,l}') \in \mathbb{R}^{\bar{m}+K_X}.$$

In this case,

$$\mathbf{d}_k(\boldsymbol{\theta}_k) = \sum_{j=1}^{\bar{m}} \mathrm{E}(\mathbf{W}\mathbf{y}_j)[\psi_{jk,0} - \psi_{jk} + \sum_{l=\bar{m}+1}^{m} (\psi_{lk,0} - \psi_{lk})c_{1,jl}] + \mathbf{X}[\boldsymbol{\pi}_{k,0} - \boldsymbol{\pi}_k + \sum_{l=\bar{m}+1}^{m} (\psi_{lk,0} - \psi_{lk})\mathbf{c}_{2,l}],$$

and hence $\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{f}_{1,k}(\boldsymbol{\delta}_k)] = \mathbf{0}$ implies that

$$\begin{aligned} \psi_{jk} &= \psi_{jk,0} + \sum_{l=\bar{m}+1}^{m} (\psi_{lk,0} - \psi_{lk})c_{1,jl} \\ \boldsymbol{\pi}_k &= \boldsymbol{\pi}_{k,0} + \sum_{l=\bar{m}+1}^{m} (\psi_{lk,0} - \psi_{lk})\mathbf{c}_{2,l}, \end{aligned} \tag{2.10}$$

for $j = 1, \cdots, \bar{m}$ and $k = 1, \cdots, m$, provided that $\mathbf{Q}'[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \cdots, \mathrm{E}(\mathbf{W}\mathbf{y}_{\bar{m}}), \mathbf{X}]$ has full column rank for n sufficiently large. Therefore, $(\psi_{1k,0}, \cdots, \psi_{\bar{m}k,0}, \boldsymbol{\pi}_{k,0}')$ can be identified if $\psi_{lk,0}$ (for $l = \bar{m}+1, \cdots, m$) can be identified from the quadratic moment condition.

When $\boldsymbol{\delta}_k$ is characterized by (2.10), we have

$$\begin{aligned} \mathrm{E}[\mathbf{v}_k(\boldsymbol{\delta}_k)'\boldsymbol{\Xi}_r\mathbf{v}_l(\boldsymbol{\delta}_l)] &= \sum_{i=1}^{m}(\psi_{ik,0} - \psi_{ik})\mathrm{tr}[\mathbf{H}_i'\boldsymbol{\Xi}_r\mathrm{E}(\mathbf{v}_l\mathbf{v}')] + \sum_{j=1}^{m}(\psi_{jl,0} - \psi_{jl})\mathrm{tr}[\boldsymbol{\Xi}_r\mathbf{H}_j\mathrm{E}(\mathbf{v}\mathbf{v}_k')] \\ &\quad + \sum_{i=1}^{m}\sum_{j=1}^{m}(\psi_{ik,0} - \psi_{ik})(\psi_{jl,0} - \psi_{jl})\mathrm{tr}[\mathbf{H}_i'\boldsymbol{\Xi}_r\mathbf{H}_j\mathrm{E}(\mathbf{v}\mathbf{v}')], \end{aligned}$$

where $\mathrm{E}(\mathbf{v}\mathbf{v}') = [(\mathbf{I}_m - \boldsymbol{\Gamma}_0')^{-1} \otimes \mathbf{I}_n]\boldsymbol{\Sigma}[(\mathbf{I}_m - \boldsymbol{\Gamma}_0)^{-1} \otimes \mathbf{I}_n]$ and $\mathrm{E}(\mathbf{v}_k\mathbf{v}') = \mathrm{E}(\mathbf{v}\mathbf{v}_k')' = (\mathbf{i}_{m,k}' \otimes \mathbf{I}_n)\mathrm{E}(\mathbf{v}\mathbf{v}')$. Therefore, the quadratic moment equations, $\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{f}_{2,kl}(\boldsymbol{\delta}_k, \boldsymbol{\delta}_l)] = 0$ for $k, l = 1, \cdots, m$, have a unique solution at $\boldsymbol{\Psi}_0$, if the equations

$$\begin{aligned} \lim_{n\to\infty} n^{-1}\{\sum_{i=1}^{m}(\psi_{ik,0} - \psi_{ik})\mathrm{tr}[\mathbf{H}_i'\boldsymbol{\Xi}_r\mathrm{E}(\mathbf{v}_l\mathbf{v}')] + \sum_{j=1}^{m}(\psi_{jl,0} - \psi_{jl})\mathrm{tr}[\boldsymbol{\Xi}_r\mathbf{H}_j\mathrm{E}(\mathbf{v}\mathbf{v}_k')] \\ + \sum_{i=1}^{m}\sum_{j=1}^{m}(\psi_{ik,0} - \psi_{ik})(\psi_{jl,0} - \psi_{jl})\mathrm{tr}[\mathbf{H}_i'\boldsymbol{\Xi}_r\mathbf{H}_j\mathrm{E}(\mathbf{v}\mathbf{v}')]\} &= 0, \end{aligned} \tag{2.11}$$

[6] We adopt the convention that $[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \cdots, \mathrm{E}(\mathbf{W}\mathbf{y}_{\bar{m}}), \mathbf{X}] = \mathbf{X}$ for $\bar{m} = 0$.

for $r = 1, \cdots, p$ and $k, l = 1, \cdots, m$, have a unique solution at $\mathbf{\Psi}_0$.[7] To wrap up, sufficient conditions for the identification of the "pseudo" reduced form parameters are summarized in the following assumption.

Assumption 2.7 At least one of the following conditions holds.

(i) $\lim_{n\to\infty} n^{-1}\mathbf{Q}'[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \cdots, \mathrm{E}(\mathbf{W}\mathbf{y}_m), \mathbf{X}]$ exists and has full column rank.

(ii) $\lim_{n\to\infty} n^{-1}\mathbf{Q}'[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \cdots, \mathrm{E}(\mathbf{W}\mathbf{y}_{\bar{m}}), \mathbf{X}]$ exists and has full column rank for some $0 \leq \bar{m} \leq m-1$. The equations (2.11), for $r = 1, \cdots, p$ and $k, l = 1, \cdots, m$, have a unique solution at $\mathbf{\Psi}_0$.

2.3.2 Identification of the structural parameters

Provided that the "pseudo" reduced form parameters $\mathbf{\Psi}_0$ and $\mathbf{\Pi}_0$ can be identified from the linear and quadratic moment conditions as discussed above. Then, the identification problem of the structural parameters in $\mathbf{\Theta}_0 = [(\mathbf{I}_m - \mathbf{\Gamma}_0)', -\mathbf{\Lambda}_0', -\mathbf{B}_0']'$ through the linear restrictions $\mathbf{\Psi}_0 = \mathbf{\Lambda}_0(\mathbf{I}_m - \mathbf{\Gamma}_0)^{-1}$ and $\mathbf{\Pi}_0 = \mathbf{B}_0(\mathbf{I}_m - \mathbf{\Gamma}_0)^{-1}$ is essentially the same one as in the classical linear simultaneous equations model (see, e.g., Schmidt, 1970). Let $\boldsymbol{\vartheta}_{k,0}$ denote the k-th column of $\mathbf{\Theta}_0$. Suppose there are R_k restrictions on $\boldsymbol{\vartheta}_{k,0}$ of the form $\mathbf{R}_k\boldsymbol{\vartheta}_{k,0} = \mathbf{0}$ where $\mathbf{R}_k$ is a $R_k \times (2m + K_X)$ matrix of known constants. Following a similar argument in Yang and Lee (2014), the sufficient and necessary **rank** condition for identification with the restrictions $\mathbf{R}_k\boldsymbol{\vartheta}_{k,0} = \mathbf{0}$ is $\mathrm{rank}(\mathbf{R}_k\mathbf{\Theta}_0) = m-1$, and the necessary **order** condition is $R_k \geq m - 1$.

Assumption 2.8 For $k = 1, \cdots, m$, $\mathbf{R}_k\boldsymbol{\vartheta}_{k,0} = \mathbf{0}$ for some $R_k \times (2m + K_X)$ constant matrix $\mathbf{R}_k$ with

$$\mathrm{rank}(\mathbf{R}_k\mathbf{\Theta}_0) = m - 1.$$

[7] A weaker identification condition can be derived based on (2.10) and (2.11) if the constants $c_{1,1l}, \cdots, c_{1,\bar{m}l}, \mathbf{c}_{2,l}$ are known to the researcher.

2.4 GMM Estimation

2.4.1 Consistency and asymptotic normality

Based on the moment conditions $\mathrm{E}[\mathbf{g}(\boldsymbol{\theta}_0)] = \mathbf{0}$, the GMM estimator for the simultaneous-equation SAR model (2.1) is given by

$$\widetilde{\boldsymbol{\theta}}_{gmm} = \arg\min \mathbf{g}(\boldsymbol{\theta})'\mathbf{F}'\mathbf{F}\mathbf{g}(\boldsymbol{\theta}) \tag{2.12}$$

where $\mathbf{F}$ is some conformable matrix such that $\lim_{n\to\infty}\mathbf{F}$ exists with full row rank greater than or equal to $\dim(\boldsymbol{\theta})$.

To characterize the asymptotic distribution of the GMM estimator, first we need to derive $\boldsymbol{\Omega} = \mathrm{Var}[\mathbf{g}(\boldsymbol{\theta}_0)]$ and $\mathbf{D} = -\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta}_0)]$. As $\boldsymbol{\Xi}_r$'s have zero diagonals for $r = 1, \cdots, p$, it follows by Lemmas 2.1 and 2.2 in the Appendix that

$$\boldsymbol{\Omega} = \mathrm{Var}[\mathbf{g}(\boldsymbol{\theta}_0)] = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{22} \end{bmatrix}, \tag{2.13}$$

where

$$\boldsymbol{\Omega}_{11} = \mathrm{Var}[\mathbf{g}_1(\boldsymbol{\theta}_0)] = (\mathbf{I}_m \otimes \mathbf{Q}')\boldsymbol{\Sigma}(\mathbf{I}_m \otimes \mathbf{Q}) = \begin{bmatrix} \mathbf{Q}'\boldsymbol{\Sigma}_{11}\mathbf{Q} & \cdots & \mathbf{Q}'\boldsymbol{\Sigma}_{1m}\mathbf{Q} \\ \vdots & \ddots & \vdots \\ \mathbf{Q}'\boldsymbol{\Sigma}_{1m}\mathbf{Q} & \cdots & \mathbf{Q}'\boldsymbol{\Sigma}_{mm}\mathbf{Q} \end{bmatrix}$$

and $\boldsymbol{\Omega}_{22} = \mathrm{Var}[\mathbf{g}_2(\boldsymbol{\theta}_0)]$ with a typical block matrix in $\boldsymbol{\Omega}_{22}$ given by

$$\mathrm{E}(\mathbf{g}_{2,ij}\mathbf{g}'_{2,kl})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \begin{bmatrix} \mathrm{tr}(\boldsymbol{\Sigma}_{il}\boldsymbol{\Xi}_1\boldsymbol{\Sigma}_{jk}\boldsymbol{\Xi}_1) + \mathrm{tr}(\boldsymbol{\Sigma}_{ik}\boldsymbol{\Xi}_1\boldsymbol{\Sigma}_{jl}\boldsymbol{\Xi}'_1) & \cdots & \mathrm{tr}(\boldsymbol{\Sigma}_{il}\boldsymbol{\Xi}_1\boldsymbol{\Sigma}_{jk}\boldsymbol{\Xi}_1) + \mathrm{tr}(\boldsymbol{\Sigma}_{ik}\boldsymbol{\Xi}_p\boldsymbol{\Sigma}_{jl}\boldsymbol{\Xi}'_p) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\boldsymbol{\Sigma}_{il}\boldsymbol{\Xi}_p\boldsymbol{\Sigma}_{jk}\boldsymbol{\Xi}_1) + \mathrm{tr}(\boldsymbol{\Sigma}_{ik}\boldsymbol{\Xi}_p\boldsymbol{\Sigma}_{jl}\boldsymbol{\Xi}'_1) & \cdots & \mathrm{tr}(\boldsymbol{\Sigma}_{il}\boldsymbol{\Xi}_p\boldsymbol{\Sigma}_{jk}\boldsymbol{\Xi}_p) + \mathrm{tr}(\boldsymbol{\Sigma}_{ik}\boldsymbol{\Xi}_p\boldsymbol{\Sigma}_{jl}\boldsymbol{\Xi}'_p) \end{bmatrix}.$$

The explicit expression for $\mathbf{D}$ depends on the specific restrictions imposed on the model parameters. Let $\mathbf{Z}_k = [\mathbf{Y}_k, \bar{\mathbf{Y}}_k, \mathbf{X}_k]$. Then,

$$\mathbf{D} = -\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta}_0)] = [\mathbf{D}'_1, \mathbf{D}'_2]', \tag{2.14}$$

where

$$\mathbf{D}_1 = -\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_1(\boldsymbol{\theta}_0)] = \begin{bmatrix} \mathbf{Q}'\mathrm{E}(\mathbf{Z}_1) & & \\ & \ddots & \\ & & \mathbf{Q}'\mathrm{E}(\mathbf{Z}_m) \end{bmatrix}$$

and

$$\mathbf{D}_2 = -\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_2(\boldsymbol{\theta}_0)] = \begin{bmatrix} \boldsymbol{\Upsilon}_{1,11} & & \\ \vdots & & \\ \boldsymbol{\Upsilon}_{1,1m} & & \\ & \ddots & \\ & & \boldsymbol{\Upsilon}_{1,m1} \\ & & \vdots \\ & & \boldsymbol{\Upsilon}_{1,mm} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Upsilon}_{2,11} & & \\ & \ddots & \\ & & \boldsymbol{\Upsilon}_{2,1m} \\ & \vdots & \\ \boldsymbol{\Upsilon}_{2,m1} & & \\ & \ddots & \\ & & \boldsymbol{\Upsilon}_{2,mm} \end{bmatrix},$$

with $\boldsymbol{\Upsilon}_{1,kl} = [\mathrm{E}(\mathbf{Z}_k'\boldsymbol{\Xi}_1\mathbf{u}_l), \cdots, \mathrm{E}(\mathbf{Z}_k'\boldsymbol{\Xi}_p\mathbf{u}_l)]'$ and $\boldsymbol{\Upsilon}_{2,kl} = [\mathrm{E}(\mathbf{Z}_l'\boldsymbol{\Xi}_1'\mathbf{u}_k), \cdots, \mathrm{E}(\mathbf{Z}_l'\boldsymbol{\Xi}_p'\mathbf{u}_k)]'$. In the following proposition we establish consistency and asymptotic normality of the GMM estimator $\widetilde{\boldsymbol{\theta}}_{gmm}$ defined in (2.12).

Proposition 2.1 Suppose Assumptions 2.1-2.8 hold, Then,

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

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 $\mathbf{D}$ given in (2.13) and (2.14) respectively.

With $\mathbf{F}'\mathbf{F}$ in (2.12) replaced by $(n^{-1}\boldsymbol{\Omega})^{-1}$, $\mathrm{AsyVar}(\widetilde{\boldsymbol{\theta}}_{gmm})$ reduces to $(\lim_{n\to\infty} n^{-1}\mathbf{D}'\boldsymbol{\Omega}^{-1}\mathbf{D})^{-1}$. Therefore, by the generalized Schwarz inequality, $(n^{-1}\boldsymbol{\Omega})^{-1}$ is the optimal GMM weighting marix. However, since $\boldsymbol{\Omega}$ depends on the unknown matrix $\boldsymbol{\Sigma}$, the GMM estimator with the optimal weighting matrix $(n^{-1}\boldsymbol{\Omega})^{-1}$ is infeasible. The following proposition extends the result in Lin and Lee (2010) to the simultaneous-equation SAR model by suggesting consistent estimators for $n^{-1}\boldsymbol{\Omega}$ and

$n^{-1}\mathbf{D}$ as inspired by White (1980). With consistently estimated $n^{-1}\boldsymbol{\Omega}$ and $n^{-1}\mathbf{D}$, a feasible optimal GMM estimator and a heteroskedasticity-robust estimator of its covariance can be obtained.

Proposition 2.2 Suppose Assumptions 2.1-2.8 hold. Let $\widetilde{\boldsymbol{\theta}}$ be a consistent estimator of $\boldsymbol{\theta}_0$ and $\widetilde{\boldsymbol{\Sigma}}_{kl} = \mathrm{diag}(\widetilde{u}_{1k}\widetilde{u}_{1l}, \cdots, \widetilde{u}_{nk}\widetilde{u}_{nl})$ where $\widetilde{u}_{ik}$ is the i-th element of $\widetilde{\mathbf{u}}_k = \mathbf{u}_k(\widetilde{\boldsymbol{\theta}}_k)$. Let $n^{-1}\widetilde{\mathbf{D}}$ and $n^{-1}\widetilde{\boldsymbol{\Omega}}$ be estimators of $n^{-1}\boldsymbol{\Omega}$ and $n^{-1}\mathbf{D}$, with $\boldsymbol{\theta}_0$ and $\boldsymbol{\Sigma}_{kl}$ in $\boldsymbol{\Omega}$ and $\mathbf{D}$ replaced by $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\Sigma}}_{kl}$ respectively. Then $n^{-1}\widetilde{\mathbf{D}} - n^{-1}\mathbf{D} = \mathrm{o_p}(1)$ and $n^{-1}\widetilde{\boldsymbol{\Omega}} - n^{-1}\boldsymbol{\Omega} = \mathrm{o_p}(1)$.

Finally Proposition 2.3 establishes asymptotic normality of the feasible optimal GMM estimator.

Proposition 2.3 Suppose Assumptions 2.1-2.8 hold. The optimal GMM estimator is given by

$$\widehat{\boldsymbol{\theta}}_{gmm} = \arg\min \mathbf{g}(\boldsymbol{\theta})'\widetilde{\boldsymbol{\Omega}}\mathbf{g}(\boldsymbol{\theta}), \tag{2.15}$$

where $n^{-1}\widetilde{\boldsymbol{\Omega}}$ is a consistent estimator of $n^{-1}\boldsymbol{\Omega}$ such that $n^{-1}\widetilde{\boldsymbol{\Omega}} - n^{-1}\boldsymbol{\Omega} = \mathrm{o_p}(1)$. Then $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{d}} \mathrm{N}(\mathbf{0}, (\lim_{n\to\infty} n^{-1}\mathbf{D}'\boldsymbol{\Omega}\mathbf{D})^{-1})$.

Note that, the 3SLS estimator can be treated as a special case of the optimal GMM estimator using only linear moment conditions, i.e.,

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

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{Z}_m \end{bmatrix}$$

and $\widetilde{\mathbf{P}} = (\mathbf{I}_m \otimes \mathbf{Q})[(\mathbf{I}_m \otimes \mathbf{Q}')\widetilde{\boldsymbol{\Sigma}}(\mathbf{I}_m \otimes \mathbf{Q})]^{-1}(\mathbf{I}_m \otimes \mathbf{Q}')$. Similar to Proposition 2.3, we can show that

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

Since $\mathbf{D}'\boldsymbol{\Omega}^{-1}\mathbf{D} - \mathbf{D}_1'\boldsymbol{\Omega}_{11}^{-1}\mathbf{D}_1 = \mathbf{D}_2'\boldsymbol{\Omega}_{22}^{-1}\mathbf{D}_2$, which is positive semi-definite, the proposed GMM estimator is asymptotically more efficient than the 3SLS estimator.

2.4.2 Best moment conditions under homoskedasticity

The above optimal GMM estimator is only "optimal" given the moment conditions. The asymptotic efficiency of the optimal GMM estimator can be improved by using the "best" moment conditions. As discussed in Lin and Lee (2010), under heteroskedasticity of an unknown form, the best moment conditions may not be available. However, under homoskedasticity, it is possible to find the best $\mathbf{Q}$ and $\mathbf{\Xi}_r$'s that satisfy Assumption 2.6. In general, the best $\mathbf{Q}$ and $\mathbf{\Xi}_r$'s depend on the specification of the simultaneous-equation model. For expositional purpose, we consider a two-equation SAR model given by

$$\begin{aligned}\mathbf{y}_1 &= \gamma_{21,0}\mathbf{y}_2 + \lambda_{11,0}\mathbf{W}\mathbf{y}_1 + \lambda_{21,0}\mathbf{W}\mathbf{y}_2 + \mathbf{X}_1\boldsymbol{\beta}_{1,0} + \mathbf{u}_1 \\ \mathbf{y}_2 &= \gamma_{12,0}\mathbf{y}_1 + \lambda_{12,0}\mathbf{W}\mathbf{y}_1 + \lambda_{22,0}\mathbf{W}\mathbf{y}_2 + \mathbf{X}_2\boldsymbol{\beta}_{2,0} + \mathbf{u}_2\end{aligned} \tag{2.16}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are respectively $n \times K_1$ and $n \times K_2$ submatrices of $\mathbf{X}$. Suppose $\mathbf{u}_1$ and $\mathbf{u}_2$ are $n \times 1$ vectors of i.i.d. random variable with zero mean and $\mathrm{E}(\mathbf{u}_1\mathbf{u}_1') = \sigma_{11}\mathbf{I}_n$, $\mathrm{E}(\mathbf{u}_2\mathbf{u}_2') = \sigma_{22}\mathbf{I}_n$ and $\mathrm{E}(\mathbf{u}_1\mathbf{u}_2') = \sigma_{12}\mathbf{I}_n$.

With the residual functions

$$\begin{aligned}\mathbf{u}_1(\boldsymbol{\theta}_1) &= \mathbf{y}_1 - \gamma_{21}\mathbf{y}_2 - \lambda_{11}\mathbf{W}\mathbf{y}_1 - \lambda_{21}\mathbf{W}\mathbf{y}_2 - \mathbf{X}_1\boldsymbol{\beta}_1 \\ \mathbf{u}_2(\boldsymbol{\theta}_2) &= \mathbf{y}_2 - \gamma_{12}\mathbf{y}_1 - \lambda_{12}\mathbf{W}\mathbf{y}_1 - \lambda_{22}\mathbf{W}\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2,\end{aligned}$$

the moment functions are given by $\mathbf{g}(\boldsymbol{\theta}) = [\mathbf{g}_1(\boldsymbol{\theta})', \mathbf{g}_2(\boldsymbol{\theta})']'$, where

$$\mathbf{g}_1(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{Q}'\mathbf{u}_1(\boldsymbol{\theta}_1) \\ \mathbf{Q}'\mathbf{u}_2(\boldsymbol{\theta}_2) \end{bmatrix}$$

and

$$\mathbf{g}_2(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{g}_{2,11}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_1) \\ \mathbf{g}_{2,12}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ \mathbf{g}_{2,21}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_1) \\ \mathbf{g}_{2,22}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_2) \end{bmatrix},$$

with $\mathbf{g}_{2,kl}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_l) = [\boldsymbol{\Xi}_1'\mathbf{u}_k(\boldsymbol{\theta}_k), \cdots, \boldsymbol{\Xi}_p'\mathbf{u}_k(\boldsymbol{\theta}_k)]'\mathbf{u}_l(\boldsymbol{\theta}_l)$. Then, the asymptotic covariance matrix for the optimal GMM estimator defined in (2.15) is $\text{AsyVar}(\widehat{\boldsymbol{\theta}}_{gmm}) = (\lim_{n\to\infty} n^{-1}\mathbf{D}'\boldsymbol{\Omega}\mathbf{D})^{-1}$. Under homoskedasticity,

$$\boldsymbol{\Omega} = \text{Var}[\mathbf{g}(\boldsymbol{\theta}_0)] = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{22} \end{bmatrix},$$

where

$$\boldsymbol{\Omega}_{11} = \text{Var}[\mathbf{g}_1(\boldsymbol{\theta}_0)] = \begin{bmatrix} \sigma_{11}\mathbf{Q}'\mathbf{Q} & \sigma_{12}\mathbf{Q}'\mathbf{Q} \\ \sigma_{12}\mathbf{Q}'\mathbf{Q} & \sigma_{22}\mathbf{Q}'\mathbf{Q} \end{bmatrix}$$

and

$$\begin{aligned} \boldsymbol{\Omega}_{22} &= \text{Var}[\mathbf{g}_2(\boldsymbol{\theta}_0)] \\ &= \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{22}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & \sigma_{22}\sigma_{12} \\ \sigma_{12}^2 & \sigma_{22}\sigma_{12} & \sigma_{22}\sigma_{12} & \sigma_{22}^2 \end{bmatrix} \otimes \boldsymbol{\Delta}_1 + \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & \sigma_{22}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{22}\sigma_{12} \\ \sigma_{12}^2 & \sigma_{22}\sigma_{12} & \sigma_{22}\sigma_{12} & \sigma_{22}^2 \end{bmatrix} \otimes \boldsymbol{\Delta}_2, \end{aligned}$$

with

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

Furthermore,

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

where

$$\mathbf{D}_1 = -\text{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}_1(\boldsymbol{\theta}_0)] = \begin{bmatrix} \mathbf{Q}'\text{E}(\mathbf{Z}_1) & \\ & \mathbf{Q}'\text{E}(\mathbf{Z}_m) \end{bmatrix}$$

and

$$\mathbf{D}_2 = -\text{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}_2(\boldsymbol{\theta}_0)] = \begin{bmatrix} \boldsymbol{\Upsilon}_{1,11} & \\ \boldsymbol{\Upsilon}_{1,12} & \\ & \boldsymbol{\Upsilon}_{1,21} \\ & \boldsymbol{\Upsilon}_{1,22} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Upsilon}_{2,11} & \\ & \boldsymbol{\Upsilon}_{2,12} \\ \boldsymbol{\Upsilon}_{2,21} & \\ & \boldsymbol{\Upsilon}_{2,22} \end{bmatrix},$$

with $\boldsymbol{\Upsilon}_{1,kl} = [\mathrm{E}(\mathbf{Z}_k'\boldsymbol{\Xi}_1\mathbf{u}_l), \cdots, \mathrm{E}(\mathbf{Z}_k'\boldsymbol{\Xi}_p\mathbf{u}_l)]'$ and $\boldsymbol{\Upsilon}_{2,kl} = [\mathrm{E}(\mathbf{Z}_l'\boldsymbol{\Xi}_1'\mathbf{u}_k), \cdots, \mathrm{E}(\mathbf{Z}_l'\boldsymbol{\Xi}_p'\mathbf{u}_k)]'$.

Let

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

and $\boldsymbol{\Omega}^* = \mathrm{Var}[\mathbf{g}^*(\boldsymbol{\theta}_0)]$, where $\mathbf{g}_1^*(\boldsymbol{\theta}) = (\mathbf{I}_2 \otimes \mathbf{Q}^*)'\mathbf{u}(\boldsymbol{\theta})$ and

$$\mathbf{g}_2^*(\boldsymbol{\theta}) = [\mathbf{g}_{2,11}^*(\boldsymbol{\theta}_1, \boldsymbol{\theta}_1)', \mathbf{g}_{2,12}^*(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)', \mathbf{g}_{2,21}^*(\boldsymbol{\theta}_2, \boldsymbol{\theta}_1)', \mathbf{g}_{2,22}^*(\boldsymbol{\theta}_2, \boldsymbol{\theta}_2)']'$$

with

$$\mathbf{g}_{2,kl}^*(\boldsymbol{\theta}_k, \boldsymbol{\theta}_l) = [\mathbf{u}_k(\boldsymbol{\theta}_k)'\boldsymbol{\Xi}_1^*\mathbf{u}_l(\boldsymbol{\theta}_l), \ldots, \mathbf{u}_k(\boldsymbol{\theta}_k)'\boldsymbol{\Xi}_{p^*}^*\mathbf{u}_l(\boldsymbol{\theta}_l)]'.$$

The following equation gives the best GMM (BGMM) estimator

$$\tilde{\boldsymbol{\theta}}_{bgmm} = \arg\min \mathbf{g}^*(\boldsymbol{\theta})'\boldsymbol{\Omega}^{*-1}\mathbf{g}^*(\boldsymbol{\theta}) \tag{2.17}$$

with the optimal $\mathbf{Q}^*$ and $\{\boldsymbol{\Xi}_j^* : j = 1, \ldots, p^*\}$.

Setting $p^* = 8$, the equations bellow define the optimal choice of IV matrix and weights for the quadratic moment functions,

$$\mathbf{Q}^* = [\mathbf{G}_{11}\mathbf{X}_1, \mathbf{G}_{12}\mathbf{X}_2, \mathbf{G}_{21}\mathbf{X}_1, \mathbf{G}_{22}\mathbf{X}_2, \mathbf{S}_{11}^{-1}\mathbf{X}_1, \mathbf{S}_{12}^{-1}\mathbf{X}_2, \mathbf{S}_{12}^{-1}\mathbf{X}_1, \mathbf{S}_{22}^{-1}\mathbf{X}_2, \mathbf{X}] \tag{2.18}$$

$$\boldsymbol{\Xi}_1^* = (\mathbf{i}_{2,1}' \otimes \mathbf{W})\mathbf{S}^{-1}(\mathbf{i}_{2,1} \otimes \mathbf{I}_n) - \mathrm{diag}[(\mathbf{i}_{2,1}' \otimes \mathbf{W})\mathbf{S}^{-1}(\mathbf{i}_{2,1} \otimes \mathbf{I}_n)] \tag{2.19}$$

$$\boldsymbol{\Xi}_2^* = (\mathbf{i}_{2,2}' \otimes \mathbf{W})\mathbf{S}^{-1}(\mathbf{i}_{2,2} \otimes \mathbf{I}_n) - \mathrm{diag}[(\mathbf{i}_{2,2}' \otimes \mathbf{W})\mathbf{S}^{-1}(\mathbf{i}_{2,2} \otimes \mathbf{I}_n)] \tag{2.20}$$

$$\boldsymbol{\Xi}_3^* = (\mathbf{i}_{2,1}' \otimes \mathbf{W})\mathbf{S}^{-1}(\mathbf{i}_{2,2} \otimes \mathbf{I}_n) - \mathrm{diag}[(\mathbf{i}_{2,1}' \otimes \mathbf{W})\mathbf{S}^{-1}(\mathbf{i}_{2,2} \otimes \mathbf{I}_n)] \tag{2.21}$$

$$\boldsymbol{\Xi}_4^* = (\mathbf{i}_{2,2}' \otimes \mathbf{W})\mathbf{S}^{-1}(\mathbf{i}_{2,1} \otimes \mathbf{I}_n) - \mathrm{diag}[(\mathbf{i}_{2,2}' \otimes \mathbf{W})\mathbf{S}^{-1}(\mathbf{i}_{2,1} \otimes \mathbf{I}_n)] \tag{2.22}$$

$$\boldsymbol{\Xi}_5^* = (\mathbf{i}_{2,1}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{i}_{2,1} \otimes \mathbf{I}_n) - \mathrm{diag}[(\mathbf{i}_{2,1}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{i}_{2,1} \otimes \mathbf{I}_n)] \tag{2.23}$$

$$\boldsymbol{\Xi}_6^* = (\mathbf{i}_{2,2}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{i}_{2,2} \otimes \mathbf{I}_n) - \mathrm{diag}[(\mathbf{i}_{2,2}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{i}_{2,2} \otimes \mathbf{I}_n)] \tag{2.24}$$

$$\boldsymbol{\Xi}_7^* = (\mathbf{i}_{2,1}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{i}_{2,2} \otimes \mathbf{I}_n) - \mathrm{diag}[(\mathbf{i}_{2,1}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{i}_{2,2} \otimes \mathbf{I}_n)] \tag{2.25}$$

$$\boldsymbol{\Xi}_8^* = (\mathbf{i}_{2,2}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{i}_{2,1} \otimes \mathbf{I}_n) - \mathrm{diag}[(\mathbf{i}_{2,2}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{i}_{2,1} \otimes \mathbf{I}_n)] \tag{2.26}$$

where $\mathbf{S} = [(\boldsymbol{\Gamma}_0' \otimes \mathbf{I}_n) - (\boldsymbol{\Lambda}_0' \otimes \mathbf{W})]$; $\mathbf{S}_{rs}^{-1} = (\mathbf{i}_{2,r}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{i}_{2,s} \otimes \mathbf{I}_n)$ and $\mathbf{G}_{rs} = \mathbf{W}\mathbf{S}_{rs}^{-1}$, for $r, s = 1, 2$.

The following proposition establish efficiency of the BGMM estimator constructed with $\mathbf{Q}^*$ and $\{\boldsymbol{\Xi}_j^*\}_{j=1}^8$.

Proposition 2.4 Suppose Assumptions 2.2-2.7 hold. Suppose $\mathbf{u}$ has zero mean and variance $\mathbf{\Sigma}\otimes\mathbf{I}_n$, where $\mathbf{\Sigma}$ is a 2×2 symmetric positive definite matrix. The BGMM estimator in equation (2.17) constructed with $\mathbf{Q}^*$ and $\{\mathbf{\Xi}_j^*\}_{j=1}^8$ defined in equations (2.18)-(2.26) is efficient in the class of GMM estimators defined in equation (2.12).

2.5 Monte Carlo

In this section we perform a small Monte Carlo simulation study. We consider the following model

$$\begin{aligned}\mathbf{y}_1 &= \lambda_{11,0}\mathbf{W}\mathbf{y}_1 + \lambda_{21,0}\mathbf{W}\mathbf{y}_2 + \gamma_{21,0}\mathbf{y}_2 + \mathbf{x}_1\beta_{1,0} + \mathbf{u}_1 \\ \mathbf{y}_2 &= \lambda_{12,0}\mathbf{W}\mathbf{y}_1 + \lambda_{22,0}\mathbf{W}\mathbf{y}_2 + \gamma_{12,0}\mathbf{y}_1 + \mathbf{x}_2\beta_{2,0} + \mathbf{u}_2\end{aligned}$$

with $\mathbf{x}_1, \mathbf{x}_2 \sim \mathrm{N}(1,1)$; $\beta_{1,0} = \beta_{2,0} = 0.6$; $\lambda_{11,0} = \lambda_{22,0} = \lambda_{12,0} = \lambda_{21,0} = 0.1$; and $\gamma_{21,0} = \gamma_{12,0} = 0.2$. We consider both the conditional heteroskedastic case with $\mathrm{Var}[u_{s,i}|x_{s,i}] = \sigma_{ss,i} = x_{s,i}^2$ and $\mathrm{Cov}[u_{1,i}, u_{2,i}|x_{1,i}, x_{2,i}] = \sigma_{12,i} = \rho\sqrt{\sigma_{11,i}\sigma_{22,i}}$; and the homoskedastic case with $\sigma_{ss,i} = 2$ and $\sigma_{12,i} = \rho$, where $\rho = 0.1$, 0.5 and 0.9. In both heteroskedastic and homoskedastic cases, disturbances are conditionally normally distributed centered at 0. Note that the parameter choices are such that the homoskedastic and heteroskedastic cases are comparable. Let $\mathbf{W}_0$ denote the spatial weights matrix for the study of crimes across 49 districts in Columbus, Ohio, in Anselin (1988). For $n = 98$, we set $\mathbf{W} = \mathbf{I}_2 \otimes \mathbf{W}_0$, and for $n = 490$, we set $\mathbf{W} = \mathbf{I}_{10} \otimes \mathbf{W}_0$. We conduct 1000 replications in the simulation experiment for each of the different specifications. Eight estimators are considered

a) two stages least squares (2SLS) with linear moment condition $\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\theta}_1)$;

b) three stages least squares (3SLS-ht) taking into account the possibility of heteroskedasticity in construction of $\mathbf{\Omega}$ based on moment conditions $(\mathbf{I}_2 \otimes \mathbf{Q}')\mathbf{u}(\boldsymbol{\theta})$;

c) three stages least squares (3SLS-hm) ignoring the possibility of heteroskedasticity in construction of $\mathbf{\Omega}$ based on moment conditions $(\mathbf{I}_2 \otimes \mathbf{Q}')\mathbf{u}(\boldsymbol{\theta})$;

d) "single equation" generalized method of moments (GMM1-ht) estimation taking into account the possibility of heteroskedasticity in construction of $\mathbf{\Omega}$ based on linear moment condition $\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\theta}_1)$ and quadratic moment condition $\mathbf{u}_1(\boldsymbol{\theta}_1)'\mathbf{\Xi}\mathbf{u}_1(\boldsymbol{\theta}_1)$;

e) "single equation" generalized method of moments (GMM1-hm) estimation ignoring the possibility of heteroskedasticity in construction of $\mathbf{\Omega}$ based on $\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\theta}_1)$ and $\mathbf{u}_1(\boldsymbol{\theta}_1)'\mathbf{\Xi}\mathbf{u}_1(\boldsymbol{\theta}_1)$ moment conditions;

f) "simultaneous equation" generalized method of moments (GMM2-ht) estimation taking into account the possibility of heteroskedasticity in construction of $\mathbf{\Omega}$ based on the linear moment conditions $(\mathbf{I}_2 \otimes \mathbf{Q}')\mathbf{u}(\boldsymbol{\theta})$ and quadratic moment conditions $\mathbf{u}_1(\boldsymbol{\theta}_1)'\mathbf{\Xi}\mathbf{u}_1(\boldsymbol{\theta}_1)$, $\mathbf{u}_1(\boldsymbol{\theta}_1)'\mathbf{\Xi}\mathbf{u}_2(\boldsymbol{\theta}_2)$, $\mathbf{u}_2(\boldsymbol{\theta}_2)'\mathbf{\Xi}\mathbf{u}_1(\boldsymbol{\theta}_1)$ and $\mathbf{u}_2(\boldsymbol{\theta}_2)'\mathbf{\Xi}\mathbf{u}_2(\boldsymbol{\theta}_2)$;

g) "simultaneous equation" generalized method of moments (GMM2-hm) estimation ignoring the possibility of heteroskedasticity in construction of $\mathbf{\Omega}$ based on linear moment conditions $(\mathbf{I}_2 \otimes \mathbf{Q}')\mathbf{u}(\boldsymbol{\theta})$, and quadratic moment conditions $\mathbf{u}_1(\boldsymbol{\theta}_1)'\mathbf{\Xi}\mathbf{u}_1(\boldsymbol{\theta}_1)$, $\mathbf{u}_1(\boldsymbol{\theta}_1)'\mathbf{\Xi}\mathbf{u}_2(\boldsymbol{\theta}_2)$, $\mathbf{u}_2(\boldsymbol{\theta}_2)'\mathbf{\Xi}\mathbf{u}_1(\boldsymbol{\theta}_1)$ and $\mathbf{u}_2(\boldsymbol{\theta}_2)'\mathbf{\Xi}\mathbf{u}_2(\boldsymbol{\theta}_2)$ and;

h) maximum likelihood estimator (MLE).

We use $\mathbf{Q} = [\mathbf{WX}, \mathbf{W}^2\mathbf{X}, \mathbf{X}]$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2]$, and $\mathbf{\Xi} = \mathbf{W} - \text{diag}(\mathbf{W})$. The $\mathbf{\Omega}$ matrix was constructed using 2SLS residuals. Results are reported in tables 2.1-2.4 in the appendix. We summarize the main observations bellow.

(i) Construction of $\mathbf{\Omega}$ using the sandwich method introduced in Proposition 2.2 does not significantly reduce efficiency of estimators even under homoskedasticity. Under heteroskedasticity, the results do not indicate efficiency gain in using the sandwich method.

(ii) Efficiency follows the following order, from most efficient to least efficient, MLE, GMM2, GMM1, 3SLS and 2SLS.

(iii) For the sample size of $n = 490$ under heteroskedasticity reported in Table 2.2, mean of GMM2 estimates are in general closer to true parameter values than MLE, especially in the estimation of $\lambda_{12,0}$. In particular, the percentage bias[8] of the MLE of $\lambda_{12,0}$ is given by roughly 26%, 23% and 19% for ρ equal to 0.1, 0.5 and 0.9 respectively. Under homoskedasticity, we observe from Table 2.4, those numbers drop to approximately 4%, 4% and 3%. The percentage bias of all GMM2-ht estimate for $n = 490$ fall below the 5%. These results suggests that GMM2-ht is robust under heteroskedasticity, whereas MLE tend to be biased.

(iv) On an early 2011 13-inch display MacBook Pro with 2.3 Gz Intel Core i5 processor and 8GB of memory, the simulation took 49 hours to run. The GMM estimators and MLE where calculated in MATLAB using `fminuc` optimization procedure with user provided gradient. Excluding MLE from the simulation, running time drops to about 12 hours. This implies that about 75% of the time to run the simulation was dedicated to running MLE. Furthermore, for $n = 98$, GMM2-ht averaged 0.248(0.171) seconds per estimation while MLE averaged 1.311(0.372) seconds and for $n = 490$, GMM2-ht averaged 1.328(0.173) seconds per estimation whereas MLE averaged 11.219(1.608) seconds, where the values in parenthesis correspond to the standard deviation in running time. This is mainly due to the fact that the GMM estimators considered in this simulation have objective functions that are polynomial, whereas the objective function of the MLE is considerably more cumbersome. Due to the computational ease of optimizing polynomial functions, it takes considerably less time to compute the GMM estimators considered in this paper.

2.6 Conclusion

In this paper, we propose a general GMM framework for the estimation of System of Simultaneous Equations SAR models with unknown heteroskedasticity. 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 regressors. We establish the consistency and asymptotic normality of the proposed the

[8] The percentage bias is given by bias divided by the true parameter value.

GMM estimator and discuss the optimal choice of moment conditions. We also provide a method for consistently estimate the variance covariance matrix of the GMM estimator under unknown heteroskedasticity. The approach taken for variance covariance estimation follows closely White (1980), which has been referred as the sandwich method by applied researchers.

Our Monte Carlo study show that the proposed estimator perform well in finite samples. In particular, the estimator in Proposition 2.3 constructed with $\tilde{\mathbf{\Omega}}$ in Proposition 2.2 is robust under heteroskedasticity with no apparent loss in efficiency under homoskedasticity. The simulation suggests that MLE is biased in the presence of heteroskedasticity, whereas our GMM estimator's unbiasedness is not affected by the presence of heteroskedasticity. Furthermore, the computational cost imposed by the proposed estimator is drastically smaller than MLE's. We believe that the gain in precision in the case of unknown heteroskedastic error, the computational ease and consistency of variance estimation of our estimator more than offsets the efficiency gains of MLE.

Appendix 2.1 Lemmas

In the following, we list some lemmas useful for proving the main results in this paper.

Lemma 2.1 Let $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ be $n \times n$ nonstochastic matrices with zero diagonals. Let $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{\epsilon}_3, \boldsymbol{\epsilon}_4$ be $n \times 1$ vectors of independent random variables with zero mean. Let $\mathbf{\Sigma}_{kl} = \mathrm{E}(\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_l')$ for $k, l = 1, 2, 3, 4$. Then,

$$\mathrm{E}(\boldsymbol{\epsilon}_1' \mathbf{A} \boldsymbol{\epsilon}_2 \boldsymbol{\epsilon}_3' \mathbf{B} \boldsymbol{\epsilon}_4) = \mathrm{tr}(\mathbf{\Sigma}_{13} \mathbf{A} \mathbf{\Sigma}_{24} \mathbf{B}') + \mathrm{tr}(\mathbf{\Sigma}_{14} \mathbf{A} \mathbf{\Sigma}_{23} \mathbf{B}).$$

Proof: As $a_{ii} = b_{ii} = 0$ for all i,

$$\begin{aligned}
& \mathrm{E}(\boldsymbol{\epsilon}_1'\mathbf{A}\boldsymbol{\epsilon}_2\boldsymbol{\epsilon}_3'\mathbf{B}\boldsymbol{\epsilon}_4) \\
= \; & \mathrm{E}(\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n a_{ij}b_{kl}\epsilon_{1,i}\epsilon_{2,j}\epsilon_{3,k}\epsilon_{4,l}) \\
= \; & \sum_{i=1}^n a_{ii}b_{ii}\mathrm{E}(\epsilon_{1,i}\epsilon_{2,i}\epsilon_{3,i}\epsilon_{4,i}) + \sum_{i=1}^n\sum_{j\neq i}^n a_{ii}b_{jj}\mathrm{E}(\epsilon_{1,i}\epsilon_{2,i})\mathrm{E}(\epsilon_{3,j}\epsilon_{4,j}) \\
& + \sum_{i=1}^n\sum_{j\neq i}^n a_{ij}b_{ij}\mathrm{E}(\epsilon_{1,i}\epsilon_{3,i})\mathrm{E}(\epsilon_{2,j}\epsilon_{4,j}) + \sum_{i=1}^n\sum_{j\neq i}^n a_{ij}b_{ji}\mathrm{E}(\epsilon_{1,i}\epsilon_{4,i})\mathrm{E}(\epsilon_{2,j}\epsilon_{3,j}) \\
= \; & \mathrm{tr}(\boldsymbol{\Sigma}_{13}\mathbf{A}\boldsymbol{\Sigma}_{24}\mathbf{B}') + \mathrm{tr}(\boldsymbol{\Sigma}_{14}\mathbf{A}\boldsymbol{\Sigma}_{23}\mathbf{B}).
\end{aligned}$$

■

Lemma 2.2 Let $\mathbf{A} = [a_{ij}]$ be an $n \times n$ nonstochastic matrix with a zero diagonal and $\mathbf{c} = (c_1, \cdots, c_n)$ be an $n \times 1$ nonstochastic vector. Let $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{\epsilon}_3$ be $n \times 1$ vectors of independent random variables with zero mean. Then,

$$\mathrm{E}(\boldsymbol{\epsilon}_1'\mathbf{A}\boldsymbol{\epsilon}_2\boldsymbol{\epsilon}_3'\mathbf{c}) = 0.$$

Proof: As $a_{ii} = 0$ for all i,

$$\mathrm{E}(\boldsymbol{\epsilon}_1'\mathbf{A}\boldsymbol{\epsilon}_2\boldsymbol{\epsilon}_3'\mathbf{c}) = \mathrm{E}(\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n a_{ij}c_k\epsilon_{1,i}\epsilon_{2,j}\epsilon_{3,k}) = \sum_{i=1}^n a_{ii}c_i\mathrm{E}(\epsilon_{1,i}\epsilon_{2,i}\epsilon_{3,i}) = 0.$$

■

Lemma 2.3 Let $\mathbf{A}$ be an $mn \times mn$ nonstochastic matrix with row and column sums uniformly bounded in absolute value. Suppose $\mathbf{u}$ satisfies Assumption 2.1. Then (i) $n^{-1}\mathbf{u}'\mathbf{A}\mathbf{u} = \mathrm{O_p}(1)$ and (ii) $n^{-1}[\mathbf{u}'\mathbf{A}\mathbf{u} - \mathrm{E}(\mathbf{u}'\mathbf{A}\mathbf{u})] = \mathrm{o_p}(1)$.

Proof: A trivial extension of Lemma A.3 in Lin and Lee (2010). ■

Lemma 2.4 Let $\mathbf{A}$ be an $mn \times mn$ nonstochastic matrix with row and column sums uniformly bounded in absolute value. Let $\mathbf{c}$ be an $mn \times 1$ nonstochastic vector with uniformly bounded elements. Suppose $\mathbf{u}$ satisfies Assumption 2.1. Then $n^{-1/2}\mathbf{c}'\mathbf{A}\mathbf{u} = \mathrm{O_p}(1)$ and $n^{-1}\mathbf{c}'\mathbf{A}\mathbf{u} = \mathrm{o_p}(1)$. Furthermore, if $\lim_{n\to\infty} n^{-1}\mathbf{c}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\mathbf{c}$ exists and is positive definite, then

$$n^{-1/2}\mathbf{c}'\mathbf{A}\mathbf{u} \xrightarrow{\mathrm{d}} \mathrm{N}(\mathbf{0}, \lim_{n\to\infty} n^{-1}\mathbf{c}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\mathbf{c}).$$

Proof: A trivial extension of Lemma A.4 in Lin and Lee (2010). ■

Lemma 2.5 Let $\mathbf{A}_{kl}$ be an $n \times n$ nonstochastic matrix with row and column sums uniformly bounded in absolute value and $\mathbf{c}_k$ an $n \times 1$ nonstochastic vector with uniformly bounded elements for $k, l = 1, \cdots, m$. Suppose $\mathbf{u}$ satisfies Assumption 2.1. Let $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon)$, where $\epsilon = \sum_{k=1}^m \mathbf{c}_k'\mathbf{u}_k + \sum_{k=1}^m \sum_{l=1}^m [\mathbf{u}_k'\mathbf{A}_{kl}\mathbf{u}_l - \mathrm{tr}(\mathbf{A}_{kl}\boldsymbol{\Sigma}_{kl})]$. If $n^{-1}\sigma_\epsilon^2$ is bounded away from zero, then $\sigma_\epsilon^{-1}\epsilon \xrightarrow{\mathrm{d}} \mathrm{N}(0, 1)$.

Proof: A trivial extension of Lemma 3 in Yang and Lee (2014). ■

Lemma 2.6 Let $\mathbf{c}_1$ and $\mathbf{c}_2$ be $mn \times 1$ nonstochastic vectors with uniformly bounded elements. Let $\mathbf{S} = \mathbf{I}_{mn} - (\boldsymbol{\Gamma}_0' \otimes \mathbf{I}_n) - (\boldsymbol{\Lambda}_0' \otimes \mathbf{W})$ and $\widetilde{\mathbf{S}} = \mathbf{I}_{mn} - (\widetilde{\boldsymbol{\Gamma}}' \otimes \mathbf{I}_n) - (\widetilde{\boldsymbol{\Lambda}}' \otimes \mathbf{W})$, where $\widetilde{\boldsymbol{\Gamma}}$ and $\widetilde{\boldsymbol{\Lambda}}$ are consistent estimators of $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Lambda}_0$ respectively. Then, $n^{-1}\mathbf{c}_1'(\widetilde{\mathbf{S}}^{-1} - \mathbf{S}^{-1})\mathbf{c}_2 = \mathrm{o_p}(1)$.

Proof: A trivial extension of Lemma A.9 in Lee (2007). ■

Lemma 2.7 Let $\mathbf{f}(\boldsymbol{\theta}) = [\mathbf{g}_1(\boldsymbol{\theta})', \mathbf{f}_2(\boldsymbol{\theta})']'$ with $\mathrm{E}[\mathbf{f}(\boldsymbol{\theta}_0)] = \mathbf{0}$. Define $\mathbf{D}_i = -\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{f}_i(\boldsymbol{\theta})]$ and $\boldsymbol{\Omega}_{ij} = \mathrm{E}[\mathbf{f}_i(\boldsymbol{\theta})\mathbf{f}_j(\boldsymbol{\theta})']$ for $i, j = 1, 2$. The following statements are equivalent (i) $\mathbf{f}_2$ is redundant given $\mathbf{f}_1$; (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 $\mathbf{D}_1 = \boldsymbol{\Omega}_{11}\mathbf{A}$.

Proof: See Breusch et al. (1999). ■

Lemma 2.8 Consider the set of moment conditions $\mathbf{f}(\boldsymbol{\theta}) = [\mathbf{f}_1(\boldsymbol{\theta})', \mathbf{f}_2(\boldsymbol{\theta})', \mathbf{f}_3(\boldsymbol{\theta})']'$ with $\mathrm{E}[\mathbf{f}(\boldsymbol{\theta}_0)] = \mathbf{0}$. Then $(\mathbf{f}_2', \mathbf{f}_3')'$ is redundant given $\mathbf{f}_1$ if and only if $\mathbf{f}_2$ is redundant given $\mathbf{f}_1$ and $\mathbf{f}_3$ is redundant given $\mathbf{f}_1$.

Proof: See Breusch et al. (1999). ■

Appendix 2.2 Proofs

Proof of Proposition 2.1: For consistency, we first need to show that $n^{-1}\mathbf{F}\mathbf{g}(\boldsymbol{\theta}) - n^{-1}\mathrm{E}[\mathbf{F}\mathbf{g}(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta}$. Suppose the i-th row of $\mathbf{F}$ can be written as

$$\mathbf{F}_i = [\mathbf{f}_{i,1}, \cdots, \mathbf{f}_{i,m}, f_{i,11,1}, \cdots, f_{i,11,p}, \cdots, f_{i,mm,1}, \cdots, f_{i,mm,p}].$$

Then,

$$\mathbf{F}_i\mathbf{g}(\boldsymbol{\theta}) = \sum_{k=1}^{m} \mathbf{f}_{i,k}\mathbf{Q}'\mathbf{u}_k(\boldsymbol{\theta}_k) + \sum_{k=1}^{m}\sum_{l=1}^{m}\sum_{r=1}^{p} f_{i,kl,r}\mathbf{u}_k(\boldsymbol{\theta}_k)'\boldsymbol{\Xi}_r\mathbf{u}_l(\boldsymbol{\theta}_l).$$

Let $\bar{\boldsymbol{\gamma}}_{k,0} = (\bar{\gamma}_{1k,0}, \cdots, \bar{\gamma}_{mk,0})'$ and $\bar{\boldsymbol{\lambda}}_{k,0} = (\bar{\lambda}_{1k,0}, \cdots, \bar{\lambda}_{mk,0})'$ denote, respectively, the k-th column of $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Lambda}_0$, including the restricted parameters. From the reduced form (2.2),

$$\begin{aligned} \mathbf{u}_k(\boldsymbol{\theta}_k) &= \mathbf{y}_k - \mathbf{Y}_k\boldsymbol{\gamma}_k - \bar{\mathbf{Y}}_k\boldsymbol{\lambda}_k - \mathbf{X}_k\boldsymbol{\beta}_k \qquad (2.27)\\ &= \mathbf{Y}_k(\boldsymbol{\gamma}_{k,0} - \boldsymbol{\gamma}_k) + \bar{\mathbf{Y}}_k(\boldsymbol{\lambda}_{k,0} - \boldsymbol{\lambda}_k) + \mathbf{X}_k(\boldsymbol{\beta}_{k,0} - \boldsymbol{\beta}_k) + \mathbf{u}_k \\ &= \mathbf{Y}(\bar{\boldsymbol{\gamma}}_{k,0} - \bar{\boldsymbol{\gamma}}_k) + \mathbf{W}\mathbf{Y}(\bar{\boldsymbol{\lambda}}_{k,0} - \bar{\boldsymbol{\lambda}}_k) + \mathbf{X}_k(\boldsymbol{\beta}_{k,0} - \boldsymbol{\beta}_k) + \mathbf{u}_k \\ &= \mathbf{d}_k(\boldsymbol{\theta}_k) + \mathbf{u}_k + \sum_{l=1}^{m}[(\bar{\gamma}_{lk,0} - \bar{\gamma}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{I}_n) + (\bar{\lambda}_{lk,0} - \bar{\lambda}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{W})]\mathbf{S}^{-1}\mathbf{u} \end{aligned}$$

where

$$\mathbf{d}_k(\boldsymbol{\theta}_k) = \sum_{l=1}^{m}[(\bar{\gamma}_{lk,0} - \bar{\gamma}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{I}_n) + (\bar{\lambda}_{lk,0} - \bar{\lambda}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{W})]\mathbf{S}^{-1}(\mathbf{B}'_0 \otimes \mathbf{I}_n)\mathbf{x} + \mathbf{X}_k(\boldsymbol{\beta}_{k,0} - \boldsymbol{\beta}_k)$$

and $\mathbf{S} = (\boldsymbol{\Gamma}'_0 \otimes \mathbf{I}_n) - (\boldsymbol{\Lambda}'_0 \otimes \mathbf{W})$. This implies that

$$\mathrm{E}[\mathbf{Q}'\mathbf{u}_k(\boldsymbol{\theta}_k)] = \mathbf{Q}'\mathbf{d}_k(\boldsymbol{\theta}_k)$$

and

$$
\begin{aligned}
&\mathrm{E}[\mathbf{u}_k(\boldsymbol{\theta}_k)'\boldsymbol{\Xi}_r\mathbf{u}_l(\boldsymbol{\theta}_l)] = \mathbf{d}_k(\boldsymbol{\theta}_k)'\boldsymbol{\Xi}_r\mathbf{d}_l(\boldsymbol{\theta}_l) \\
&+\sum_{j=1}^{m}(\bar{\gamma}_{jl,0}-\bar{\gamma}_{jl})\mathrm{tr}[\boldsymbol{\Xi}_r(\mathbf{i}'_{m,j}\otimes\mathbf{I}_n)\mathbf{S}^{-1}\mathrm{E}(\mathbf{u}\mathbf{u}'_k)] + \sum_{j=1}^{m}(\bar{\lambda}_{jl,0}-\bar{\lambda}_{jl})\mathrm{tr}[\boldsymbol{\Xi}_r(\mathbf{i}'_{m,j}\otimes\mathbf{W})\mathbf{S}^{-1}\mathrm{E}(\mathbf{u}\mathbf{u}'_k)] \\
&+\sum_{i=1}^{m}(\bar{\gamma}_{ik,0}-\bar{\gamma}_{ik})\mathrm{tr}[\boldsymbol{\Xi}'_r(\mathbf{i}'_{m,i}\otimes\mathbf{I}_n)\mathbf{S}^{-1}\mathrm{E}(\mathbf{u}\mathbf{u}'_l)] + \sum_{i=1}^{m}(\bar{\lambda}_{ik,0}-\bar{\lambda}_{ik})\mathrm{tr}[\boldsymbol{\Xi}'_r(\mathbf{i}'_{m,i}\otimes\mathbf{W})\mathbf{S}^{-1}\mathrm{E}(\mathbf{u}\mathbf{u}'_l)] \\
&+\sum_{i=1}^{m}\sum_{j=1}^{m}(\bar{\gamma}_{ik,0}-\bar{\gamma}_{ik})(\bar{\gamma}_{jl,0}-\bar{\gamma}_{jl})\mathrm{tr}[\mathbf{S}'^{-1}(\mathbf{i}'_{m,i}\otimes\mathbf{I}_n)'\boldsymbol{\Xi}_r(\mathbf{i}'_{m,j}\otimes\mathbf{I}_n)\mathbf{S}^{-1}\boldsymbol{\Sigma}] \\
&+\sum_{i=1}^{m}\sum_{j=1}^{m}(\bar{\lambda}_{ik,0}-\bar{\lambda}_{ik})(\bar{\gamma}_{jl,0}-\bar{\gamma}_{jl})\mathrm{tr}[\mathbf{S}'^{-1}(\mathbf{i}'_{m,i}\otimes\mathbf{W})'\boldsymbol{\Xi}_r(\mathbf{i}'_{m,j}\otimes\mathbf{I}_n)\mathbf{S}^{-1}\boldsymbol{\Sigma}] \\
&+\sum_{i=1}^{m}\sum_{j=1}^{m}(\bar{\gamma}_{ik,0}-\bar{\gamma}_{ik})(\bar{\lambda}_{jl,0}-\bar{\lambda}_{jl})\mathrm{tr}[\mathbf{S}'^{-1}(\mathbf{i}'_{m,i}\otimes\mathbf{I}_n)'\boldsymbol{\Xi}_r(\mathbf{i}'_{m,j}\otimes\mathbf{W})\mathbf{S}^{-1}\boldsymbol{\Sigma}] \\
&+\sum_{i=1}^{m}\sum_{j=1}^{m}(\bar{\lambda}_{ik,0}-\bar{\lambda}_{ik})(\bar{\lambda}_{jl,0}-\bar{\lambda}_{jl})\mathrm{tr}[\mathbf{S}'^{-1}(\mathbf{i}'_{m,i}\otimes\mathbf{W})'\boldsymbol{\Xi}_r(\mathbf{i}'_{m,j}\otimes\mathbf{W})\mathbf{S}^{-1}\boldsymbol{\Sigma}].
\end{aligned}
$$

As $\mathbf{F}_i\mathbf{g}(\boldsymbol{\theta})$ is a quadratic function of $\boldsymbol{\theta}$ and the parameter space of $\boldsymbol{\theta}$ is bounded, it follows by Lemmas 2.3 and 2.4 that $n^{-1}\mathbf{F}_i\mathbf{g}(\boldsymbol{\theta}) - n^{-1}\mathrm{E}[\mathbf{F}_i\mathbf{g}(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta}$. Furthermore, $n^{-1}\mathrm{E}[\mathbf{F}\mathbf{g}(\boldsymbol{\theta})]$ is uniformly equicontinuous in $\boldsymbol{\theta}$. The identification condition and the uniform equicontinuity of $n^{-1}\mathrm{E}[\mathbf{F}\mathbf{g}(\boldsymbol{\theta})]$ imply that the identification uniqueness condition for $n^{-2}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})']\mathbf{F}'\mathbf{F}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})]$ holds. Therefore, $\widetilde{\boldsymbol{\theta}}_{gmm}$ is a consistent estimator of $\boldsymbol{\theta}_0$ (White, 1994).

For the asymptotic normality, we use the mean value theorem to write

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

where $\overline{\boldsymbol{\theta}}$ is as convex combination of $\widetilde{\boldsymbol{\theta}}_{gmm}$ and $\boldsymbol{\theta}_0$. By Lemma 2.5 together with the Cramer Wald device, $\frac{1}{\sqrt{n}}\mathbf{F}\mathbf{g}(\boldsymbol{\theta}_0)$ converges in distribution to $\mathrm{N}(\mathbf{0}, \lim_{n\to\infty} n^{-1}\mathbf{F}\boldsymbol{\Omega}\mathbf{F}')$. Furthermore, consistency of $\widetilde{\boldsymbol{\theta}}_{gmm}$ implies that $\overline{\boldsymbol{\theta}}$ also converges in probability to $\boldsymbol{\theta}_0$. Therefore it suffices to show that $n^{-1}\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta}) - \lim_{n\to\infty} n^{-1}\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta}$. We divide the remainder of the proof into two parts focusing respectively on the partial derivatives of $\mathbf{g}_1(\boldsymbol{\theta})$ and $\mathbf{g}_2(\boldsymbol{\theta})$.

First, note that

$$\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_1(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{Q}'\mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{Q}'\mathbf{Z}_m \end{bmatrix}.$$

From the reduced form (2.2), we have

$$\mathbf{y}_l = (\mathbf{i}'_{m,l} \otimes \mathbf{I}_n)\mathbf{S}^{-1}[(\mathbf{B}'_0 \otimes \mathbf{I}_n)\mathbf{x} + \mathbf{u}].$$

By Lemma 2.4, we have $n^{-1}\mathbf{Q}'\mathbf{y}_l - n^{-1}\mathrm{E}(\mathbf{Q}'\mathbf{y}_l) = \mathrm{o_p}(1)$ and $n^{-1}\mathbf{Q}'\mathbf{W}\mathbf{y}_l = n^{-1}\mathrm{E}(\mathbf{Q}'\mathbf{W}\mathbf{y}_l) + \mathrm{o_p}(1)$, which implies that that $n^{-1}\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_1(\boldsymbol{\theta}) - n^{-1}\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_1(\boldsymbol{\theta})] = \mathrm{o_p}(1)$.

Next, note that

$$\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_2(\boldsymbol{\theta}) = -\begin{bmatrix} \bar{\boldsymbol{\Upsilon}}_{1,11}(\boldsymbol{\theta}_1) & & \\ \vdots & & \\ \bar{\boldsymbol{\Upsilon}}_{1,1m}(\boldsymbol{\theta}_m) & & \\ & \ddots & \\ & & \bar{\boldsymbol{\Upsilon}}_{1,m1}(\boldsymbol{\theta}_1) \\ & & \vdots \\ & & \bar{\boldsymbol{\Upsilon}}_{1,mm}(\boldsymbol{\theta}_m) \end{bmatrix} - \begin{bmatrix} \bar{\boldsymbol{\Upsilon}}_{2,11}(\boldsymbol{\theta}_1) & & \\ & \ddots & \\ & & \bar{\boldsymbol{\Upsilon}}_{2,1m}(\boldsymbol{\theta}_1) \\ & \vdots & \\ \bar{\boldsymbol{\Upsilon}}_{2,m1}(\boldsymbol{\theta}_m) & & \\ & \ddots & \\ & & \bar{\boldsymbol{\Upsilon}}_{2,mm}(\boldsymbol{\theta}_m) \end{bmatrix},$$

where $\bar{\boldsymbol{\Upsilon}}_{1,kl}(\boldsymbol{\theta}_l) = [\mathbf{Z}'_k\boldsymbol{\Xi}_1\mathbf{u}_l(\boldsymbol{\theta}_l), \cdots, \mathbf{Z}'_k\boldsymbol{\Xi}_p\mathbf{u}_l(\boldsymbol{\theta}_l)]'$ and $\bar{\boldsymbol{\Upsilon}}_{2,kl}(\boldsymbol{\theta}_k) = [\mathbf{Z}'_l\boldsymbol{\Xi}'_1\mathbf{u}_k(\boldsymbol{\theta}_k), \cdots, \mathbf{Z}'_l\boldsymbol{\Xi}'_p\mathbf{u}_k(\boldsymbol{\theta}_k)]'$. As $\mathbf{u}_k(\boldsymbol{\theta}_k)$ can be expanded as (2.27), it follows by Lemmas 2.3 and 2.4 that $n^{-1}\mathbf{Z}'_k\boldsymbol{\Xi}_r\mathbf{u}_l(\boldsymbol{\theta}_l) - n^{-1}\mathrm{E}[\mathbf{Z}'_k\boldsymbol{\Xi}_r\mathbf{u}_l(\boldsymbol{\theta}_l)] = \mathrm{o_p}(1)$ and $n^{-1}\mathbf{Z}'_l\boldsymbol{\Xi}'_r\mathbf{u}_k(\boldsymbol{\theta}_k) - n^{-1}\mathrm{E}[\mathbf{Z}'_l\boldsymbol{\Xi}'_r\mathbf{u}_k(\boldsymbol{\theta}_k)] = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta}$. Therefore, $n^{-1}\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_2(\boldsymbol{\theta}) - n^{-1}\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_2(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ and the desired result follows. ■

Proof of Proposition 2.2: We divide the proof into two parts. First we prove the consistency of $n^{-1}\widetilde{\boldsymbol{\Omega}}$. Then we prove the consistency of $n^{-1}\widetilde{\mathbf{D}}$.

(1) To show $n^{-1}\widetilde{\boldsymbol{\Omega}} - n^{-1}\boldsymbol{\Omega} = \mathrm{o_p}(1)$, we need to show that (1a) $n^{-1}\mathbf{Q}'\widetilde{\boldsymbol{\Sigma}}_{kl}\mathbf{Q} - n^{-1}\mathbf{Q}'\boldsymbol{\Sigma}_{kl}\mathbf{Q} = \mathrm{o_p}(1)$ and (1b) $n^{-1}\mathrm{tr}(\widetilde{\boldsymbol{\Sigma}}_{kl}\mathbf{A}_1\widetilde{\boldsymbol{\Sigma}}_{st}\mathbf{A}_2) - n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}_{kl}\mathbf{A}_1\boldsymbol{\Sigma}_{st}\mathbf{A}_2) = \mathrm{o_p}(1)$, for $k, l, s, t = 1, \cdots, m$, for $n \times n$ zero-diagonal matrices $\mathbf{A}_1 = [a_{1,ij}]$ and $\mathbf{A}_2 = [a_{2,ij}]$ that are uniformly bounded in row and column

sums. The consistency of $n^{-1}\mathbf{Q}'\widetilde{\boldsymbol{\Sigma}}_{kl}\mathbf{Q}$ in (1a) can be shown by a similar argument as in White (1980). Thus, we focus on the consistency of $n^{-1}\mathrm{tr}(\widetilde{\boldsymbol{\Sigma}}_{kl}\mathbf{A}_1\widetilde{\boldsymbol{\Sigma}}_{st}\mathbf{A}_2) = n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}\widetilde{u}_{ik}\widetilde{u}_{il}\widetilde{u}_{js}\widetilde{u}_{jt}$. It follows by a similar argument as in Lin and Lee (2010) that $n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}u_{ik}u_{il}u_{js}u_{jt} - n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}\sigma_{i,kl}\sigma_{j,st} = \mathrm{o_p}(1)$. Therefore, to show (1b) holds, we only need to show that $n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}\widetilde{u}_{ik}\widetilde{u}_{il}\widetilde{u}_{js}\widetilde{u}_{jt} - n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}u_{ik}u_{il}u_{js}u_{jt} = \mathrm{o_p}(1)$.

Note that

$$\begin{aligned}
& n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}\widetilde{u}_{ik}\widetilde{u}_{il}\widetilde{u}_{js}\widetilde{u}_{jt} - n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}u_{ik}u_{il}u_{js}u_{jt} \\
= \; & n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}(\widetilde{u}_{ik}\widetilde{u}_{il} - u_{ik}u_{il})u_{js}u_{jt} + n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}u_{ik}u_{il}(\widetilde{u}_{js}\widetilde{u}_{jt} - u_{js}u_{jt}) \\
& + n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}(\widetilde{u}_{ik}\widetilde{u}_{il} - u_{ik}u_{il})(\widetilde{u}_{js}\widetilde{u}_{jt} - u_{js}u_{jt}).
\end{aligned}$$

From (2.27), we have

$$\widetilde{\mathbf{u}}_k = \mathbf{u}_k(\widetilde{\boldsymbol{\theta}}_k) = \mathbf{y}_k - \mathbf{Y}_k\widetilde{\boldsymbol{\gamma}}_k - \bar{\mathbf{Y}}_k\widetilde{\boldsymbol{\lambda}}_k - \mathbf{X}_k\widetilde{\boldsymbol{\beta}}_k = \mathbf{d}_k(\widetilde{\boldsymbol{\theta}}_k) + \mathbf{u}_k + \mathbf{e}_k(\widetilde{\boldsymbol{\theta}}_k)$$

where

$$\begin{aligned}
\mathbf{d}_k(\widetilde{\boldsymbol{\theta}}_k) &= \sum_{l=1}^{m}[(\bar{\gamma}_{lk,0} - \widetilde{\bar{\gamma}}_{lk})(\mathbf{i}'_{m,l}\otimes\mathbf{I}_n) + (\bar{\lambda}_{lk,0} - \widetilde{\bar{\lambda}}_{lk})(\mathbf{i}'_{m,l}\otimes\mathbf{W})]\mathbf{S}^{-1}(\mathbf{B}'_0\otimes\mathbf{I}_n)\mathbf{x} + \mathbf{X}_k(\boldsymbol{\beta}_{k,0} - \widetilde{\boldsymbol{\beta}}_k) \\
\mathbf{e}_k(\widetilde{\boldsymbol{\theta}}_k) &= \sum_{l=1}^{m}[(\bar{\gamma}_{lk,0} - \widetilde{\bar{\gamma}}_{lk})(\mathbf{i}'_{m,l}\otimes\mathbf{I}_n) + (\bar{\lambda}_{lk,0} - \widetilde{\bar{\lambda}}_{lk})(\mathbf{i}'_{m,l}\otimes\mathbf{W})]\mathbf{S}^{-1}\mathbf{u}
\end{aligned}$$

and $\mathbf{S} = \mathbf{I}_{mn} - (\boldsymbol{\Gamma}'_0\otimes\mathbf{I}_n) - (\boldsymbol{\Lambda}'_0\otimes\mathbf{W})$. Let d_{ik} and e_{ik} denote the i-th element of $\mathbf{d}_k(\widetilde{\boldsymbol{\theta}}_k)$ and $\mathbf{e}_k(\widetilde{\boldsymbol{\theta}}_k)$ respectively. Then,

$$\widetilde{u}_{ik}\widetilde{u}_{il} = u_{ik}u_{il} + d_{ik}d_{il} + e_{ik}e_{il} + (u_{ik}d_{il} + d_{ik}u_{il}) + (u_{ik}e_{il} + e_{ik}u_{il}) + (d_{ik}e_{il} + e_{ik}d_{il}).$$

To show $n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}(\widetilde{u}_{ik}\widetilde{u}_{il} - u_{ik}u_{il})u_{js}u_{jt} = \mathrm{o_p}(1)$, we focus on terms that are of higher orders in u_{il}. One of such terms is

$$\begin{aligned}
n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}e_{ik}u_{il}u_{js}u_{jt} &= \sum_{r=1}^{m}(\bar{\gamma}_{rk,0} - \widetilde{\bar{\gamma}}_{rk})n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}u_{il}u_{js}u_{jt}(\mathbf{i}'_{m,r}\otimes\mathbf{i}'_{n,i})\mathbf{S}^{-1}\mathbf{u} \\
&\quad + \sum_{r=1}^{m}(\bar{\lambda}_{rk,0} - \widetilde{\bar{\lambda}}_{rk})n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}u_{il}u_{js}u_{jt}(\mathbf{i}'_{m,r}\otimes\mathbf{w}_i)\mathbf{S}^{-1}\mathbf{u},
\end{aligned}$$

where $\mathbf{w}_i$ denotes the i-th row of $\mathbf{W}$. By Assumption 2.1, we can show $\mathrm{E}|u_{hk}u_{il}u_{js}u_{jt}| \leq c$ for some constant c using Cauchy's inequality, which implies $\mathrm{E}|n^{-1}\sum_{i,j=1}^{n}\sum_{r=1}^{m} a_{1,ij}a_{2,ji}u_{il}u_{js}u_{jt}(\mathbf{i}'_{m,r} \otimes \mathbf{i}'_{n,i})\mathbf{S}^{-1}\mathbf{u}| = \mathrm{O}(1)$ and $\mathrm{E}|n^{-1}\sum_{i,j=1}^{n}\sum_{r=1}^{m} a_{1,ij}a_{2,ji}u_{il}u_{js}u_{jt}(\mathbf{i}'_{m,r} \otimes \mathbf{w}_i)\mathbf{S}^{-1}\mathbf{u}| = \mathrm{O}(1)$ because $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{W}$ and $\mathbf{S}^{-1}$ are uniformly bounded in row and column sums. Hence,

$$n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}e_{ik}u_{il}u_{js}u_{jt} = \mathrm{o_p}(1).$$

Similarly, we can show other terms in $n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}(\widetilde{u}_{ik}\widetilde{u}_{il} - u_{ik}u_{il})u_{js}u_{jt}$ are of order $\mathrm{o_p}(1)$. With a similar argument as above or as in Lin and Lee (2010),

$$n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}u_{ik}u_{il}(\widetilde{u}_{js}\widetilde{u}_{jt} - u_{js}u_{jt}) = \mathrm{o_p}(1)$$

and $n^{-1}\sum_{i,j=1}^{n} a_{1,ij}a_{2,ji}(\widetilde{u}_{ik}\widetilde{u}_{il} - u_{ik}u_{il})(\widetilde{u}_{js}\widetilde{u}_{jt} - u_{js}u_{jt}) = \mathrm{o_p}(1)$. Therefore, the consistency of $n^{-1}\mathrm{tr}(\widetilde{\mathbf{\Sigma}}_{kl}\mathbf{A}_1\widetilde{\mathbf{\Sigma}}_{st}\mathbf{A}_2)$ in (1b) follows.

(2) Some typical entries in $\mathbf{D}$ that involve unknown parameters are $\mathbf{Q}'\mathrm{E}(\mathbf{y}_k) = \mathbf{Q}'(\mathbf{i}'_{m,k} \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{B}'_0 \otimes \mathbf{I}_n)\mathbf{x}$, $\mathbf{Q}'\mathrm{E}(\mathbf{W}\mathbf{y}_k) = \mathbf{Q}'(\mathbf{i}'_{m,k} \otimes \mathbf{W}_n)\mathbf{S}^{-1}(\mathbf{B}'_0 \otimes \mathbf{I}_n)\mathbf{x}$, $\mathrm{E}(\mathbf{u}'_l\mathbf{A}\mathbf{y}_k) = \mathrm{tr}[(\mathbf{i}_{m,l} \otimes \mathbf{A})(\mathbf{i}'_{m,k} \otimes \mathbf{I}_n)\mathbf{S}^{-1}\mathbf{\Sigma}]$ and $\mathrm{E}(\mathbf{u}'_l\mathbf{A}\mathbf{W}\mathbf{y}_k) = \mathrm{tr}[(\mathbf{i}_{m,l} \otimes \mathbf{A})(\mathbf{i}'_{m,k} \otimes \mathbf{W})\mathbf{S}^{-1}\mathbf{\Sigma}]$, where $\mathbf{A} = [a_{ij}]$ is an $n \times n$ zero-diagonal matrix uniformly bounded in row and column sums. To show $n^{-1}\widetilde{\mathbf{D}} - n^{-1}\mathbf{D} = \mathrm{o_p}(1)$, we need to show that (2a) $n^{-1}\mathbf{Q}'(\mathbf{i}'_{m,k} \otimes \mathbf{I}_n)\widetilde{\mathbf{S}}^{-1}(\widetilde{\mathbf{B}}'_0 \otimes \mathbf{I}_n)\mathbf{x} - n^{-1}\mathbf{Q}'(\mathbf{i}'_{m,k} \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{B}'_0 \otimes \mathbf{I}_n)\mathbf{x} = \mathrm{o_p}(1)$ and $n^{-1}\mathbf{Q}'(\mathbf{i}'_{m,k} \otimes \mathbf{W})\widetilde{\mathbf{S}}^{-1}(\widetilde{\mathbf{B}}'_0 \otimes \mathbf{I}_n)\mathbf{x} - n^{-1}\mathbf{Q}'(\mathbf{i}'_{m,k} \otimes \mathbf{W})\mathbf{S}^{-1}(\mathbf{B}'_0 \otimes \mathbf{I}_n)\mathbf{x} = \mathrm{o_p}(1)$; (2b) $n^{-1}\mathrm{tr}[(\mathbf{i}_{m,l} \otimes \mathbf{A})(\mathbf{i}'_{m,k} \otimes \mathbf{I}_n)\widetilde{\mathbf{S}}^{-1}\widetilde{\mathbf{\Sigma}}] - n^{-1}\mathrm{tr}[(\mathbf{i}_{m,l} \otimes \mathbf{A})(\mathbf{i}'_{m,k} \otimes \mathbf{I}_n)\mathbf{S}^{-1}\mathbf{\Sigma}] = \mathrm{o_p}(1)$ and $n^{-1}\mathrm{tr}[(\mathbf{i}_{m,l} \otimes \mathbf{A})(\mathbf{i}'_{m,k} \otimes \mathbf{W})\widetilde{\mathbf{S}}^{-1}\widetilde{\mathbf{\Sigma}}] - n^{-1}\mathrm{tr}[(\mathbf{i}_{m,l} \otimes \mathbf{A})(\mathbf{i}'_{m,k} \otimes \mathbf{W})\mathbf{S}^{-1}\mathbf{\Sigma}] = \mathrm{o_p}(1)$. where $\widetilde{\mathbf{S}} = \mathbf{I}_{mn} - (\widetilde{\mathbf{\Gamma}}' \otimes \mathbf{I}_n) - (\widetilde{\mathbf{\Lambda}}' \otimes \mathbf{W})$ and

$$\widetilde{\mathbf{\Sigma}} = \begin{bmatrix} \widetilde{\mathbf{\Sigma}}_{11} & \cdots & \widetilde{\mathbf{\Sigma}}_{1m} \\ \vdots & \ddots & \vdots \\ \widetilde{\mathbf{\Sigma}}_{m1} & \cdots & \widetilde{\mathbf{\Sigma}}_{mm} \end{bmatrix}.$$

As (2a) follows by Lemma 2.6 and (2b) follows by a similar argument as in Lin and Lee (2010), we conclude $n^{-1}\widetilde{\mathbf{D}} - n^{-1}\mathbf{D} = \mathrm{o_p}(1)$. ■

Proof of Proposition 2.3: For consistency, note that

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

From the proof of Proposition 2.1, $n^{-1}\mathbf{g}(\boldsymbol{\theta})'\boldsymbol{\Omega}^{-1}\mathbf{g}(\boldsymbol{\theta}) - n^{-1}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})'\boldsymbol{\Omega}^{-1}\mathbf{g}(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta}$. Hence, it suffices to show that $n^{-1}\mathbf{g}(\boldsymbol{\theta})'(\widetilde{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\mathbf{g}(\boldsymbol{\theta}) = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta}$. Let $\|\cdot\|$ denote the euclidian norm for vectors and matrices. Then,

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

From the proof of Proposition 2.1, $n^{-1}\mathbf{g}(\boldsymbol{\theta}) - n^{-1}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta}$. As

$$n^{-1}\mathrm{E}[\mathbf{Q}'\mathbf{u}_k(\boldsymbol{\theta}_k)] = n^{-1}\mathbf{Q}'\mathbf{d}_k(\boldsymbol{\theta}_k) = \mathrm{O}(1)$$

and

$$n^{-1}\mathrm{E}[\mathbf{u}_k(\boldsymbol{\theta}_k)'\boldsymbol{\Xi}_r\mathbf{u}_l(\boldsymbol{\theta}_l)] = n^{-1}\mathbf{d}_k(\boldsymbol{\theta}_k)'\boldsymbol{\Xi}_r\mathbf{d}_l(\boldsymbol{\theta}_l) + n^{-1}\mathrm{tr}[\mathbf{G}_k(\boldsymbol{\theta}_k)'\boldsymbol{\Xi}_r\mathbf{G}_l(\boldsymbol{\theta}_l)\boldsymbol{\Sigma}] = \mathrm{O}(1)$$

uniformly in $\boldsymbol{\theta}$, where

$$\mathbf{d}_k(\boldsymbol{\theta}_k) = \sum_{l=1}^{m}[(\bar{\gamma}_{lk,0} - \bar{\gamma}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{I}_n) + (\bar{\lambda}_{lk,0} - \bar{\lambda}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{W})]\mathbf{S}^{-1}(\mathbf{B}'_0 \otimes \mathbf{I}_n)\mathbf{x} + \mathbf{X}_k(\boldsymbol{\beta}_{k,0} - \boldsymbol{\beta}_k)$$

and $\mathbf{G}_k(\boldsymbol{\theta}_k) = \sum_{l=1}^{m}[(\bar{\gamma}_{lk,0} - \bar{\bar{\gamma}}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{I}_n) + (\bar{\lambda}_{lk,0} - \bar{\bar{\lambda}}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{W})]\mathbf{S}^{-1}$, it follows that $n^{-1}\|\mathbf{g}(\boldsymbol{\theta})\| = \mathrm{O_p}(1)$ uniformly in $\boldsymbol{\theta}$. Therefore, $n^{-1}\mathbf{g}(\boldsymbol{\theta})'(\widetilde{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\mathbf{g}(\boldsymbol{\theta}) = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta}$.

For the asymptotic distribution, by the mean value theorem, for some convex combination of $\widehat{\boldsymbol{\theta}}_{gmm}$ and $\boldsymbol{\theta}_0$ denoted by $\overline{\boldsymbol{\theta}}$,

$$\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0) &= -\left[\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm})'\left(\frac{1}{n}\widetilde{\boldsymbol{\Omega}}\right)^{-1}\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\overline{\boldsymbol{\theta}})\right]^{-1}\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{gmm})'\left(\frac{1}{n}\widetilde{\boldsymbol{\Omega}}\right)^{-1}\frac{1}{\sqrt{n}}\mathbf{g}(\boldsymbol{\theta}_0) \\
&= \left[\frac{1}{n}\mathbf{D}\left(\frac{1}{n}\boldsymbol{\Omega}\right)^{-1}\frac{1}{n}\mathbf{D}\right]^{-1}\frac{1}{n}\mathbf{D}\left(\frac{1}{n}\boldsymbol{\Omega}\right)^{-1}\frac{1}{\sqrt{n}}\mathbf{g}(\boldsymbol{\theta}_0) + \mathrm{o_p}(1) \\
&\xrightarrow{\mathrm{d}} \mathrm{N}\left(\mathbf{0}, \left[\lim_{n\to\infty} n^{-1}\mathbf{D}'\boldsymbol{\Omega}\mathbf{D}\right]^{-1}\right)
\end{aligned}$$

where the asymptotic distribution statement is implied by Lemma 2.5. ■

Proof of Proposition 2.4: Similar to Liu et al. (2010) and Liu and Saraiva (2015), we use Breusch et al. (1999), to show that any additional linear and/or quadratic moments are redundant. In order to obtain the desired result, we need to closer inspect the matrix $\mathbf{D}$, for Lemma 2.6

establishes that if $\mathbf{D} = \mathrm{E}[\mathbf{g}(\boldsymbol{\theta}_0)\mathbf{g}^*(\boldsymbol{\theta}_0)']\mathbf{A}$, any moment condition of the form in $\mathbf{g}$ will be redundant given $\mathbf{g}^*$. Furthermore, if $\mathbf{g}$ will be redundant given $\mathbf{g}^*$, then by Lemma 2.8 any sub-set of moment conditions of $\mathbf{g}$ will also be redundant given $\mathbf{g}^*$.

Defining $\boldsymbol{\theta}_k = (\lambda_{1k}, \lambda_{2k}, \phi_k, \boldsymbol{\beta}_k')'$ and

$$\frac{\partial}{\partial \boldsymbol{\theta}_k'}\mathbf{u}_k(\boldsymbol{\theta}_k) = \mathbf{Z}_k = [\mathbf{W}\mathbf{y}_1, \mathbf{W}\mathbf{y}_2, \mathbf{y}_s, \mathbf{X}_k]$$

leads to the following sub-matrix associated with the linear moment conditions,

$$\begin{aligned}\mathbf{D}_1 &= -\mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_1(\boldsymbol{\theta}_0)\right] \\ &= \begin{bmatrix} \mathbf{Q}'\mathrm{E}(\mathbf{Z}_1) & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}'\mathrm{E}(\mathbf{Z}_2) \end{bmatrix}.\end{aligned}$$

Combining $\mathbf{y}_k = (\mathbf{i}_{2,k}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}[\mathrm{vec}(\mathbf{X}\mathbf{B}_0) + \mathbf{u}]$ and the definition of $\mathbf{Z}_k$, we obtain

$$\mathrm{E}(\mathbf{Z}_k) = \left[\mathbf{G}_{11}\mathbf{X}_1\boldsymbol{\beta}_{1,0} + \mathbf{G}_{12}\mathbf{X}_2\boldsymbol{\beta}_{2,0},\quad \mathbf{G}_{21}\mathbf{X}_1\boldsymbol{\beta}_{1,0} + \mathbf{G}_{22}\mathbf{X}_2\boldsymbol{\beta}_{2,0},\quad \mathbf{S}_{s1}^{-1}\mathbf{X}_1\boldsymbol{\beta}_{1,0} + \mathbf{S}_{s2}^{-1}\mathbf{X}_2\boldsymbol{\beta}_{2,0},\quad X_k\right]$$

for $k = 1, 2$ and $s \neq k$.

For the quadratic moment conditions, note that $\bar{\boldsymbol{\Upsilon}}_{1,kl} = [\mathrm{E}(\mathbf{Z}_k'\boldsymbol{\Xi}_1\mathbf{u}_l), \cdots, \mathrm{E}(\mathbf{Z}_k'\boldsymbol{\Xi}_p\mathbf{u}_l)]'$ and $\bar{\boldsymbol{\Upsilon}}_{2,kl} = [\mathrm{E}(\mathbf{Z}_l'\boldsymbol{\Xi}_1'\mathbf{u}_k), \cdots, \mathrm{E}(\mathbf{Z}_l'\boldsymbol{\Xi}_p'\mathbf{u}_k)]'$. Using the definition of $\mathbf{Z}_k$, we have that

$$\begin{aligned}\mathrm{E}(\mathbf{Z}_k'\boldsymbol{\Xi}_j\mathbf{u}_l) &= \begin{bmatrix} \mathrm{E}([(\mathbf{i}_{2,1}' \otimes \mathbf{W})\mathbf{S}^{-1}\mathbf{u}]'\boldsymbol{\Xi}_j(\mathbf{i}_{2,l}' \otimes \mathbf{I}_n)\mathbf{u}) \\ \mathrm{E}([(\mathbf{i}_{2,2}' \otimes \mathbf{W})\mathbf{S}^{-1}\mathbf{u}]'\boldsymbol{\Xi}_j(\mathbf{i}_{2,l}' \otimes \mathbf{I}_n)\mathbf{u}) \\ \mathrm{E}([(\mathbf{i}_{2,s}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}\mathbf{u}]'\boldsymbol{\Xi}_j(\mathbf{i}_{2,l}' \otimes \mathbf{I}_n)\mathbf{u}) \\ \mathbf{0}_{K_k \times 1} \end{bmatrix}' \\ &= \begin{bmatrix} \mathrm{tr}(\boldsymbol{\Xi}_j'(\mathbf{i}_{2,1}' \otimes \mathbf{W})\mathbf{S}^{-1}(\boldsymbol{\Sigma}\mathbf{i}_{2,l} \otimes \mathbf{I}_n)) \\ \mathrm{tr}(\boldsymbol{\Xi}_j'(\mathbf{i}_{2,2}' \otimes \mathbf{W})\mathbf{S}^{-1}(\boldsymbol{\Sigma}\mathbf{i}_{2,l} \otimes \mathbf{I}_n)) \\ \mathrm{tr}(\boldsymbol{\Xi}_j'(\mathbf{i}_{2,s}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\boldsymbol{\Sigma}\mathbf{i}_{2,l} \otimes \mathbf{I}_n)) \\ \mathbf{0}_{K_k \times 1} \end{bmatrix}'.\end{aligned}$$

Similarly

$$\mathrm{E}(\mathbf{Z}_k'\boldsymbol{\Xi}_j'\mathbf{u}_l) = \begin{bmatrix} \mathrm{tr}(\boldsymbol{\Xi}_j(\mathbf{i}_{2,1}' \otimes \mathbf{W})\mathbf{S}^{-1}(\boldsymbol{\Sigma}\mathbf{i}_{2,l} \otimes \mathbf{I}_n)) \\ \mathrm{tr}(\boldsymbol{\Xi}_j(\mathbf{i}_{2,2}' \otimes \mathbf{W})\mathbf{S}^{-1}(\boldsymbol{\Sigma}\mathbf{i}_{2,l} \otimes \mathbf{I}_n)) \\ \mathrm{tr}(\boldsymbol{\Xi}_j(\mathbf{i}_{2,s}' \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\boldsymbol{\Sigma}\mathbf{i}_{2,l} \otimes \mathbf{I}_n)) \\ \mathbf{0}_{K_k \times 1} \end{bmatrix}'$$

for $k, l = 1, 2$ and $s \neq k$. To ease visualization we look at each sub-matrix of

$$\mathbf{D}_2 = [\mathbf{D}_{2,11}', \mathbf{D}_{2,12}', \mathbf{D}_{2,21}', \mathbf{D}_{2,22}']'$$

separately. Substituting $\bar{\boldsymbol{\Upsilon}}_{1,kl}$ and $\bar{\boldsymbol{\Upsilon}}_{2,kl}$ into $\mathbf{D}_2$, we have

$$\begin{aligned} \mathbf{D}_{2,11} &= -\mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_{2,11}(\boldsymbol{\theta}_0)\right] = \left[\bar{\boldsymbol{\Upsilon}}_{1,11} + \bar{\boldsymbol{\Upsilon}}_{2,11}, \quad \mathbf{0}_{p \times K_2+3}\right] \\ &= \begin{bmatrix} \mathrm{tr}[\boldsymbol{\Xi}_1^{(s)}\mathbf{W}\mathbf{H}_1(\boldsymbol{\Sigma}\mathbf{i}_{2,1} \otimes \mathbf{I}_n)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p^{(s)}\mathbf{W}\mathbf{H}_1(\boldsymbol{\Sigma}\mathbf{i}_{2,1} \otimes \mathbf{I}_n)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1^{(s)}\mathbf{W}\mathbf{H}_2(\boldsymbol{\Sigma}\mathbf{i}_{2,1} \otimes \mathbf{I}_n)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p^{(s)}\mathbf{W}\mathbf{H}_2(\boldsymbol{\Sigma}\mathbf{i}_{2,1} \otimes \mathbf{I}_n)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1^{(s)}\mathbf{H}_2(\boldsymbol{\Sigma}\mathbf{i}_{2,1} \otimes \mathbf{I}_n)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p^{(s)}\mathbf{H}_2(\boldsymbol{\Sigma}\mathbf{i}_{2,1} \otimes \mathbf{I}_n)] \\ \mathbf{0}_{(K_1+K_2+3)\times 1} & \cdots & \mathbf{0}_{(K_1+K_2+3)\times 1} \end{bmatrix}' \\ &= \begin{bmatrix} \mathrm{tr}[\boldsymbol{\Xi}_1^{(s)}(\sigma_{11}\boldsymbol{\Xi}_1^* + \sigma_{12}\boldsymbol{\Xi}_3^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p^{(s)}(\sigma_{11}\boldsymbol{\Xi}_1^* + \sigma_{12}\boldsymbol{\Xi}_3^*)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1^{(s)}(\sigma_{11}\boldsymbol{\Xi}_4^* + \sigma_{12}\boldsymbol{\Xi}_2^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p^{(s)}(\sigma_{11}\boldsymbol{\Xi}_4^* + \sigma_{12}\boldsymbol{\Xi}_2^*)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1^{(s)}(\sigma_{11}\boldsymbol{\Xi}_8^* + \sigma_{12}\boldsymbol{\Xi}_6^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p^{(s)}(\sigma_{11}\boldsymbol{\Xi}_8^* + \sigma_{12}\boldsymbol{\Xi}_6^*)] \\ \mathbf{0}_{(K_1+K_2+3)\times 1} & \cdots & \mathbf{0}_{(K_1+K_2+3)\times 1} \end{bmatrix}'; \end{aligned}$$

$$\mathbf{D}_{2,12} = -\mathrm{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}_{2,12}(\boldsymbol{\theta}_0)\right] = \left[\bar{\boldsymbol{\Upsilon}}_{1,12}, \quad \bar{\boldsymbol{\Upsilon}}_{2,12}\right]$$

$$= \begin{bmatrix} \mathrm{tr}[\boldsymbol{\Xi}_1'(\sigma_{12}\boldsymbol{\Xi}_1^* + \sigma_{22}\boldsymbol{\Xi}_3^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p'(\sigma_{12}\boldsymbol{\Xi}_1^* + \sigma_{22}\boldsymbol{\Xi}_3^*)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1'(\sigma_{12}\boldsymbol{\Xi}_4^* + \sigma_{22}\boldsymbol{\Xi}_2^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p'(\sigma_{12}\boldsymbol{\Xi}_4^* + \sigma_{22}\boldsymbol{\Xi}_2^*)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1'(\sigma_{12}\boldsymbol{\Xi}_8^* + \sigma_{22}\boldsymbol{\Xi}_6^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p'(\sigma_{12}\boldsymbol{\Xi}_8^* + \sigma_{22}\boldsymbol{\Xi}_6^*)] \\ \mathbf{0}_{K_1\times 1} & \cdots & \mathbf{0}_{K_1\times 1} \\ \mathrm{tr}[\boldsymbol{\Xi}_1(\sigma_{11}\boldsymbol{\Xi}_1^* + \sigma_{12}\boldsymbol{\Xi}_3^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p(\sigma_{11}\boldsymbol{\Xi}_1^* + \sigma_{12}\boldsymbol{\Xi}_3^*)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1(\sigma_{11}\boldsymbol{\Xi}_4^* + \sigma_{12}\boldsymbol{\Xi}_2^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p(\sigma_{12}\boldsymbol{\Xi}_4^* + \sigma_{22}\boldsymbol{\Xi}_2^*)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1(\sigma_{11}\boldsymbol{\Xi}_5^* + \sigma_{12}\boldsymbol{\Xi}_7^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p(\sigma_{12}\boldsymbol{\Xi}_5^* + \sigma_{22}\boldsymbol{\Xi}_7^*)] \\ \mathbf{0}_{K_2\times 1} & \cdots & \mathbf{0}_{K_2\times 1} \end{bmatrix}';$$

$$\mathbf{D}_{2,21} = -\mathrm{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}_{2,21}(\boldsymbol{\theta}_0)\right] = \left[\bar{\boldsymbol{\Upsilon}}_{2,21}, \quad \bar{\boldsymbol{\Upsilon}}_{1,21}\right]$$

$$= \begin{bmatrix} \mathrm{tr}[\boldsymbol{\Xi}_1(\sigma_{12}\boldsymbol{\Xi}_1^* + \sigma_{22}\boldsymbol{\Xi}_3^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p(\sigma_{12}\boldsymbol{\Xi}_1^* + \sigma_{22}\boldsymbol{\Xi}_3^*)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1(\sigma_{12}\boldsymbol{\Xi}_4^* + \sigma_{22}\boldsymbol{\Xi}_2^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p(\sigma_{12}\boldsymbol{\Xi}_4^* + \sigma_{22}\boldsymbol{\Xi}_2^*)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1(\sigma_{12}\boldsymbol{\Xi}_8^* + \sigma_{22}\boldsymbol{\Xi}_6^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p(\sigma_{12}\boldsymbol{\Xi}_8^* + \sigma_{22}\boldsymbol{\Xi}_6^*)] \\ \mathbf{0}_{K_1\times 1} & \cdots & \mathbf{0}_{K_1\times 1} \\ \mathrm{tr}[\boldsymbol{\Xi}_1'(\sigma_{11}\boldsymbol{\Xi}_1^* + \sigma_{12}\boldsymbol{\Xi}_3^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p'(\sigma_{11}\boldsymbol{\Xi}_1^* + \sigma_{12}\boldsymbol{\Xi}_3^*)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1'(\sigma_{11}\boldsymbol{\Xi}_4^* + \sigma_{12}\boldsymbol{\Xi}_2^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p'(\sigma_{12}\boldsymbol{\Xi}_4^* + \sigma_{22}\boldsymbol{\Xi}_2^*)] \\ \mathrm{tr}[\boldsymbol{\Xi}_1'(\sigma_{11}\boldsymbol{\Xi}_5^* + \sigma_{12}\boldsymbol{\Xi}_7^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p'(\sigma_{12}\boldsymbol{\Xi}_5^* + \sigma_{22}\boldsymbol{\Xi}_7^*)] \\ \mathbf{0}_{K_2\times 1} & \cdots & \mathbf{0}_{K_2\times 1} \end{bmatrix}';$$

and

$$
\begin{aligned}
\mathbf{D}_{2,22} &= -\mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_{2,22}(\boldsymbol{\theta}_0)\right] = \left[\mathbf{0}_{p\times K_1+3}, \quad \bar{\boldsymbol{\Upsilon}}_{1,22} + \bar{\boldsymbol{\Upsilon}}_{2,22}\right] \\
&= \begin{bmatrix}
\mathbf{0}_{(K_1+3)\times 1} & \cdots & \mathbf{0}_{(K_1+3)\times 1} \\
\mathrm{tr}[\boldsymbol{\Xi}_1^{(s)}(\sigma_{12}\boldsymbol{\Xi}_1^* + \sigma_{22}\boldsymbol{\Xi}_3^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p^{(s)}(\sigma_{12}\boldsymbol{\Xi}_1^* + \sigma_{22}\boldsymbol{\Xi}_3^*)] \\
\mathrm{tr}[\boldsymbol{\Xi}_1^{(s)}(\sigma_{12}\boldsymbol{\Xi}_4^* + \sigma_{22}\boldsymbol{\Xi}_2^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p^{(s)}(\sigma_{12}\boldsymbol{\Xi}_4^* + \sigma_{22}\boldsymbol{\Xi}_2^*)] \\
\mathrm{tr}[\boldsymbol{\Xi}_1^{(s)}(\sigma_{12}\boldsymbol{\Xi}_5^* + \sigma_{22}\boldsymbol{\Xi}_7^*)] & \cdots & \mathrm{tr}[\boldsymbol{\Xi}_p^{(s)}(\sigma_{12}\boldsymbol{\Xi}_5^* + \sigma_{22}\boldsymbol{\Xi}_7^*)] \\
\mathbf{0}_{K_2\times 1} & \cdots & \mathbf{0}_{K_2\times 1}
\end{bmatrix}';
\end{aligned}
$$

with $\mathbf{H}_k = (\mathbf{i}'_{2,k} \otimes \mathbf{I}_n)\mathbf{S}^{-1}$.

Let

$$\mathbf{A}_1 = \frac{1}{|\mathbf{\Sigma}|}\begin{bmatrix}
\sigma_{22}\boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\sigma_{12}\boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\sigma_{22}\boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\sigma_{12}\boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \sigma_{22}\boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\sigma_{12}\boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \sigma_{22}\boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\sigma_{12}\boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \sigma_{22}\boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \sigma_{22}\boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\sigma_{12}\boldsymbol{\beta}_{1,0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\sigma_{12}\boldsymbol{\beta}_{2,0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{22}\mathbf{C}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\sigma_{12}\mathbf{C}_2 \\
-\sigma_{12}\boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{11}\boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
-\sigma_{12}\boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{11}\boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & -\sigma_{12}\boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{11}\boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & -\sigma_{12}\boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{11}\boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & -\sigma_{12}\boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & -\sigma_{12}\boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{11}\boldsymbol{\beta}_{1,0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{11}\boldsymbol{\beta}_{2,0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & -\sigma_{12}\mathbf{C}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{11}\mathbf{C}_2
\end{bmatrix}$$

and

$$\begin{aligned}
\mathbf{A}_2 &= \begin{bmatrix} (\boldsymbol{\pi}_{11} \otimes \mathbf{i}_{8,1}) & (\boldsymbol{\pi}_{12} \otimes \mathbf{i}_{8,2}) & (\boldsymbol{\pi}_{12} \otimes \mathbf{i}_{8,6}) & \mathbf{0} & (\boldsymbol{\pi}_{21} \otimes \mathbf{i}_{8,1}) & (\boldsymbol{\pi}_{22} \otimes \mathbf{i}_{8,2}) & (\boldsymbol{\pi}_{21} \otimes \mathbf{i}_{8,5}) & \mathbf{0} \end{bmatrix} \\
&+ \begin{bmatrix} (\boldsymbol{\pi}_{12} \otimes \mathbf{i}_{8,3}) & (\boldsymbol{\pi}_{11} \otimes \mathbf{i}_{8,4}) & (\boldsymbol{\pi}_{11} \otimes \mathbf{i}_{8,8}) & \mathbf{0} & (\boldsymbol{\pi}_{22} \otimes \mathbf{i}_{8,3}) & (\boldsymbol{\pi}_{21} \otimes \mathbf{i}_{8,4}) & (\boldsymbol{\pi}_{22} \otimes \mathbf{i}_{8,7}) & \mathbf{0} \end{bmatrix}
\end{aligned}$$

where

$$\boldsymbol{\pi}_{11} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & \sigma_{22}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{22}\sigma_{12} \\ \sigma_{12}^2 & \sigma_{22}\sigma_{12} & \sigma_{22}\sigma_{12} & \sigma_{22}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ 0 \\ 0 \end{bmatrix} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{11}^2} \begin{bmatrix} \sigma_{22} \\ 0 \\ -\sigma_{12} \\ 0 \end{bmatrix}$$

$$\boldsymbol{\pi}_{12} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & \sigma_{22}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{22}\sigma_{12} \\ \sigma_{12}^2 & \sigma_{22}\sigma_{12} & \sigma_{22}\sigma_{12} & \sigma_{22}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{12} \\ \sigma_{22} \\ 0 \\ 0 \end{bmatrix} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{11}^2} \begin{bmatrix} 0 \\ \sigma_{22} \\ 0 \\ -\sigma_{12} \end{bmatrix}.$$

$$\boldsymbol{\pi}_{21} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & \sigma_{22}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{22}\sigma_{12} \\ \sigma_{12}^2 & \sigma_{22}\sigma_{12} & \sigma_{22}\sigma_{12} & \sigma_{22}^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ \sigma_{11} \\ \sigma_{12} \end{bmatrix} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{11}^2} \begin{bmatrix} -\sigma_{12} \\ 0 \\ \sigma_{11} \\ 0 \end{bmatrix}$$

and

$$\boldsymbol{\pi}_{22} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & \sigma_{22}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{22}\sigma_{12} \\ \sigma_{12}^2 & \sigma_{22}\sigma_{12} & \sigma_{22}\sigma_{12} & \sigma_{22}^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{11}^2} \begin{bmatrix} 0 \\ -\sigma_{12} \\ 0 \\ \sigma_{11} \end{bmatrix}$$

The following inverse is essential to verifying the definitions of $\boldsymbol{\pi}_{11}$, $\boldsymbol{\pi}_{12}$, $\boldsymbol{\pi}_{21}$ and $\boldsymbol{\pi}_{22}$.

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & \sigma_{22}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{22}\sigma_{12} \\ \sigma_{12}^2 & \sigma_{22}\sigma_{12} & \sigma_{22}\sigma_{12} & \sigma_{22}^2 \end{bmatrix}^{-1} =$$

$$\frac{1}{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^2} \begin{bmatrix} \sigma_{22}^2 & -\sigma_{22}\sigma_{12} & -\sigma_{22}\sigma_{12} & \sigma_{12}^2 \\ -\sigma_{22}\sigma_{12} & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & -\sigma_{11}\sigma_{12} \\ -\sigma_{22}\sigma_{12} & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & -\sigma_{11}\sigma_{12} \\ \sigma_{12}^2 & -\sigma_{11}\sigma_{12} & -\sigma_{11}\sigma_{12} & \sigma_{11}^2 \end{bmatrix}$$

Defining $\mathbf{A} = [\mathbf{A}_1', \mathbf{A}_2']'$, note that since $\mathbf{D} = \mathrm{E}\left[\mathbf{g}(\boldsymbol{\theta}_0)\mathbf{g}^*(\boldsymbol{\theta}_0)'\right]\mathbf{A}$, by Lemma 2.6 $\mathbf{g}$ is redundant given $\mathbf{g}^*$. ■

Appendix 2.3 Tables

Table 2.1: 2SLS, 3SLS, GMM and MLE Estimation ($n = 98$, under Heteroskedasticity)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.1$	$\lambda_{21,0} = 0.1$	$\beta_{1,0} = 0.6$
$\rho = 0.1$				
2SLS	.228(.293)[.294]	.114(.523)[.523]	.082(.538)[.538]	.580(.211)[.212]
3SLS-ht	.237(.356)[.358]	.121(.569)[.569]	.080(.587)[.587]	.560(.227)[.230]
3SLS-hm	.234(.356)[.358]	.119(.564)[.564]	.090(.575)[.575]	.563(.222)[.225]
GMM1-ht	.232(.234)[.236]	.052(.194)[.200]	.125(.312)[.313]	.580(.197)[.198]
GMM1-hm	.229(.238)[.240]	.045(.199)[.207]	.134(.330)[.332]	.584(.196)[.196]
GMM2-ht	.274(.322)[.331]	.081(.232)[.233]	.089(.236)[.236]	.548(.216)[.222]
GMM2-hm	.268(.310)[.318]	.075(.209)[.210]	.093(.257)[.257]	.561(.210)[.213]
MLE	.174(.233)[.234]	.086(.079)[.080]	.145(.105)[.114]	.596(.189)[.189]
$\rho = 0.5$				
2SLS	.242(.322)[.325]	.097(.535)[.535]	.088(.607)[.608]	.572(.206)[.208]
3SLS-ht	.237(.435)[.436]	.109(.590)[.590]	.093(.704)[.705]	.558(.219)[.223]
3SLS-hm	.226(.409)[.410]	.108(.586)[.586]	.104(.683)[.683]	.561(.216)[.219]
GMM1-ht	.258(.227)[.234]	.051(.211)[.216]	.115(.325)[.325]	.571(.193)[.196]
GMM1-hm	.255(.227)[.233]	.045(.220)[.226]	.125(.336)[.337]	.575(.191)[.193]
GMM2-ht	.277(.327)[.336]	.080(.259)[.260]	.088(.278)[.278]	.546(.210)[.217]
GMM2-hm	.262(.333)[.339]	.079(.268)[.269]	.090(.295)[.295]	.562(.205)[.208]
MLE	.168(.242)[.244]	.087(.079)[.080]	.138(.101)[.108]	.597(.190)[.190]
$\rho = 0.9$				
2SLS	.251(.293)[.298]	.106(.528)[.528]	.080(.533)[.533]	.564(.210)[.213]
3SLS-ht	.232(.352)[.353]	.126(.544)[.545]	.085(.564)[.564]	.556(.219)[.224]
3SLS-hm	.215(.357)[.357]	.123(.559)[.559]	.103(.578)[.578]	.562(.216)[.219]
GMM1-ht	.278(.206)[.221]	.057(.223)[.227]	.100(.305)[.305]	.561(.192)[.196]
GMM1-hm	.282(.217)[.231]	.050(.232)[.238]	.109(.318)[.318]	.563(.190)[.194]
GMM2-ht	.272(.303)[.312]	.090(.270)[.270]	.087(.307)[.308]	.543(.212)[.219]
GMM2-hm	.249(.300)[.304]	.085(.254)[.255]	.092(.309)[.309]	.564(.205)[.208]
MLE	.139(.671)[.673]	.098(.177)[.177]	.133(.125)[.129]	.601(.254)[.254]

Mean(SD)[RMSE]

Table 2.2: 2SLS, 3SLS, GMM and MLE Estimation ($n = 490$, under Heteroskedasticity)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.1$	$\lambda_{21,0} = 0.1$	$\beta_{1,0} = 0.6$
$\rho = 0.1$				
2SLS	.203(.103)[.103]	.105(.187)[.187]	.099(.195)[.195]	.599(.092)[.092]
3SLS-ht	.203(.106)[.106]	.102(.189)[.189]	.100(.195)[.195]	.599(.092)[.092]
3SLS-hm	.202(.106)[.106]	.105(.188)[.188]	.099(.196)[.196]	.598(.092)[.092]
GMM1-ht	.202(.098)[.098]	.085(.085)[.086]	.120(.135)[.137]	.599(.091)[.091]
GMM1-hm	.206(.104)[.104]	.072(.098)[.102]	.139(.152)[.157]	.593(.091)[.091]
GMM2-ht	.209(.093)[.093]	.098(.072)[.072]	.099(.097)[.097]	.597(.090)[.090]
GMM2-hm	.207(.093)[.094]	.098(.072)[.072]	.100(.097)[.097]	.599(.089)[.089]
MLE	.188(.082)[.083]	.095(.022)[.022]	.126(.033)[.042]	.600(.086)[.086]
$\rho = 0.5$				
2SLS	.204(.102)[.102]	.105(.188)[.188]	.097(.195)[.195]	.598(.092)[.092]
3SLS-ht	.200(.108)[.108]	.105(.189)[.189]	.100(.196)[.196]	.599(.092)[.092]
3SLS-hm	.198(.107)[.107]	.107(.188)[.188]	.099(.196)[.196]	.599(.092)[.092]
GMM1-ht	.208(.093)[.093]	.086(.096)[.097]	.115(.140)[.141]	.598(.089)[.089]
GMM1-hm	.209(.097)[.098]	.076(.104)[.107]	.132(.153)[.157]	.593(.088)[.088]
GMM2-ht	.208(.091)[.092]	.098(.083)[.083]	.100(.108)[.108]	.596(.089)[.089]
GMM2-hm	.204(.091)[.091]	.099(.083)[.083]	.101(.108)[.108]	.600(.088)[.088]
MLE	.187(.081)[.082]	.097(.020)[.021]	.123(.033)[.040]	.600(.086)[.086]
$\rho = 0.9$				
2SLS	.204(.102)[.103]	.107(.189)[.190]	.095(.196)[.196]	.597(.092)[.092]
3SLS-ht	.198(.110)[.110]	.107(.191)[.191]	.099(.197)[.197]	.599(.092)[.092]
3SLS-hm	.194(.109)[.109]	.110(.190)[.191]	.100(.198)[.198]	.599(.092)[.092]
GMM1-ht	.213(.091)[.092]	.088(.113)[.113]	.109(.150)[.151]	.596(.088)[.088]
GMM1-hm	.213(.092)[.093]	.088(.118)[.118]	.110(.158)[.159]	.595(.086)[.086]
GMM2-ht	.207(.091)[.091]	.099(.099)[.099]	.101(.122)[.122]	.596(.088)[.088]
GMM2-hm	.200(.090)[.090]	.100(.100)[.100]	.101(.122)[.122]	.601(.087)[.087]
MLE	.187(.080)[.081]	.098(.018)[.018]	.119(.029)[.035]	.601(.086)[.086]

Mean(SD)[RMSE]

Table 2.3: 2SLS, 3SLS, GMM and MLE Estimation ($n = 98$, under Homoskedasticity)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.1$	$\lambda_{21,0} = 0.1$	$\beta_{1,0} = 0.6$
$\rho = 0.1$				
2SLS	.215(.264)[.264]	.137(.495)[.496]	.064(.516)[.518]	.575(.158)[.161]
3SLS-ht	.216(.308)[.308]	.137(.548)[.549]	.075(.571)[.571]	.560(.175)[.179]
3SLS-hm	.216(.315)[.315]	.138(.539)[.541]	.077(.565)[.565]	.558(.172)[.177]
GMM1-ht	.226(.225)[.226]	.059(.200)[.204]	.125(.320)[.321]	.575(.149)[.151]
GMM1-hm	.222(.214)[.215]	.055(.199)[.204]	.133(.319)[.321]	.577(.146)[.148]
GMM2-ht	.253(.274)[.279]	.087(.218)[.218]	.094(.255)[.255]	.558(.163)[.169]
GMM2-hm	.254(.273)[.278]	.082(.193)[.193]	.100(.250)[.250]	.556(.159)[.165]
MLE	.174(.221)[.222]	.084(.073)[.075]	.146(.108)[.118]	.593(.140)[.140]
$\rho = 0.5$				
2SLS	.221(.267)[.267]	.122(.482)[.482]	.073(.498)[.498]	.573(.157)[.159]
3SLS-ht	.203(.315)[.315]	.133(.533)[.534]	.086(.555)[.555]	.564(.171)[.175]
3SLS-hm	.201(.323)[.323]	.131(.521)[.522]	.093(.547)[.547]	.563(.169)[.173]
GMM1-ht	.247(.217)[.222]	.059(.226)[.230]	.114(.337)[.338]	.570(.145)[.148]
GMM1-hm	.244(.212)[.217]	.058(.219)[.223]	.120(.333)[.334]	.572(.143)[.146]
GMM2-ht	.244(.268)[.272]	.090(.248)[.248]	.094(.286)[.286]	.564(.156)[.161]
GMM2-hm	.244(.266)[.269]	.088(.243)[.243]	.098(.277)[.277]	.563(.152)[.156]
MLE	.165(.222)[.225]	.088(.074)[.075]	.142(.101)[.110]	.599(.143)[.143]
$\rho = 0.9$				
2SLS	.227(.275)[.276]	.105(.470)[.470]	.084(.494)[.494]	.572(.152)[.155]
3SLS-ht	.191(.330)[.330]	.122(.509)[.509]	.100(.546)[.546]	.572(.165)[.168]
3SLS-hm	.187(.331)[.331]	.121(.491)[.491]	.107(.528)[.528]	.572(.161)[.164]
GMM1-ht	.269(.210)[.221]	.060(.249)[.253]	.101(.339)[.339]	.567(.138)[.142]
GMM1-hm	.265(.205)[.215]	.059(.246)[.249]	.107(.338)[.338]	.569(.135)[.139]
GMM2-ht	.235(.267)[.269]	.089(.250)[.250]	.095(.311)[.312]	.573(.148)[.151]
GMM2-hm	.232(.257)[.259]	.089(.242)[.243]	.100(.301)[.301]	.573(.143)[.145]
MLE	.157(.229)[.233]	.089(.078)[.079]	.140(.100)[.108]	.604(.145)[.145]

Mean(SD)[RMSE]

Table 2.4: 2SLS, 3SLS, GMM and MLE Estimation ($n = 490$, under Homoskedasticity)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.1$	$\lambda_{21,0} = 0.1$	$\beta_{1,0} = 0.6$
$\rho = 0.1$				
2SLS	.202(.103)[.103]	.104(.186)[.186]	.097(.190)[.190]	.598(.067)[.067]
3SLS-ht	.202(.107)[.107]	.104(.188)[.188]	.098(.193)[.193]	.598(.068)[.068]
3SLS-hm	.202(.106)[.106]	.104(.187)[.187]	.098(.191)[.191]	.597(.067)[.067]
GMM1-ht	.206(.103)[.104]	.072(.101)[.105]	.138(.155)[.159]	.592(.071)[.071]
GMM1-hm	.204(.103)[.103]	.075(.098)[.101]	.136(.152)[.156]	.592(.070)[.071]
GMM2-ht	.208(.096)[.096]	.098(.074)[.074]	.100(.101)[.101]	.598(.065)[.065]
GMM2-hm	.207(.095)[.095]	.098(.074)[.074]	.101(.101)[.101]	.597(.064)[.064]
MLE	.187(.090)[.091]	.096(.023)[.023]	.125(.036)[.044]	.599(.060)[.060]
$\rho = 0.5$				
2SLS	.203(.102)[.103]	.103(.186)[.186]	.097(.190)[.190]	.598(.067)[.067]
3SLS-ht	.199(.107)[.107]	.104(.188)[.188]	.099(.193)[.193]	.599(.068)[.068]
3SLS-hm	.198(.106)[.106]	.105(.187)[.187]	.098(.191)[.191]	.598(.067)[.067]
GMM1-ht	.209(.098)[.099]	.081(.105)[.107]	.124(.156)[.158]	.595(.066)[.067]
GMM1-hm	.209(.098)[.098]	.077(.105)[.107]	.130(.155)[.158]	.592(.066)[.066]
GMM2-ht	.204(.094)[.094]	.099(.085)[.085]	.101(.110)[.110]	.600(.063)[.063]
GMM2-hm	.204(.093)[.093]	.099(.084)[.084]	.101(.109)[.109]	.599(.063)[.063]
MLE	.187(.089)[.090]	.097(.020)[.021]	.123(.032)[.039]	.600(.060)[.060]
$\rho = 0.9$				
2SLS	.204(.102)[.102]	.103(.186)[.186]	.096(.190)[.190]	.598(.067)[.067]
3SLS-ht	.195(.107)[.107]	.105(.189)[.189]	.099(.193)[.193]	.600(.068)[.068]
3SLS-hm	.195(.107)[.107]	.106(.187)[.187]	.099(.191)[.191]	.600(.067)[.067]
GMM1-ht	.212(.094)[.095]	.090(.119)[.119]	.109(.158)[.159]	.596(.064)[.064]
GMM1-hm	.212(.094)[.094]	.090(.119)[.119]	.109(.159)[.159]	.596(.063)[.063]
GMM2-ht	.201(.092)[.092]	.099(.099)[.099]	.102(.120)[.120]	.601(.062)[.062]
GMM2-hm	.201(.091)[.091]	.099(.098)[.098]	.102(.119)[.119]	.601(.062)[.062]
MLE	.187(.089)[.090]	.098(.019)[.019]	.119(.030)[.036]	.602(.061)[.061]

Mean(SD)[RMSE]

Bibliography

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

Baltagi, B. H. and Pirotte, A. (2011). Seemingly unrelated regressions with spatial error components, Empirical Economics **40**: 5–49.

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. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model, International Economic Review **40**: 509–533.

Kelejian, H. H. and Prucha, I. R. (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-oending 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. (2015). GMM identification and estimation of peer effects in a system of simultaneous equations. Working Paper, University of Colorado Boulder.

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.

Liu, X., Lee, L. F. and Bollinger, C. (2010). An efficient GMM estimator of spatial autoregressive models, Journal of Econometrics **159**: 303–319.

Liu, X. and Saraiva, P. (2015). GMM estimation of SAR models with endogenous regressors. Working Paper, University of Colorado Boulder.

Newey, W. 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, Amsterdam, 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, NY.

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. III, Elsevier, North-Holland, Amsterdam, pp. 249–277.

Schmidt, P. (1970). Econometrics, Marcel Dekker, New York, NY.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica **48**: 817–838.

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

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