GMM Estimation of Spatial Autoregressive Models in a System of Simultaneous Equations with Heteroskedasticity*

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

Paulo Saraiva
Department of Economics,
University of Colorado Boulder
Email: saraiva@colorado.edu

October, 2016

Abstract

This paper proposes a GMM estimation framework for the SAR model in a system of simultaneous equations with heteroskedastic disturbances. Besides linear moment conditions, the proposed GMM estimator also utilizes quadratic moment conditions based on the covariance structure of model disturbances within and across equations. Compared with the QML approach, the GMM estimator is easier to implement and robust under heteroskedasticity of unknown form. We derive the heteroskedasticity-robust standard error for the GMM estimator. Monte Carlo experiments show that the proposed GMM estimator performs well in finite samples.

JEL classification: C31, C36

Key words: simultaneous equations, spatial autoregressive models, quadratic moment conditions, unknown heteroskedasticity

1 Introduction

The spatial autoregressive (SAR) model introduced by Cliff and Ord (1973, 1981) has received considerable attention in various fields of economics as it provides a convenient framework to model the interaction between economic agents. However, with a few exceptions (e.g., Kelejian

*We thank the editor, the associate editor, and two anonymous referees for valuable comments and suggestions. The remaining errors are our own.

and Prucha, 2004; Baltagi and Pirotte, 2011; Yang and Lee, 2017), 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 (2017) study the identification and estimation of the simultaneous-equation SAR model by the full information quasi-maximum likelihood (QML) approach. They give identification conditions for the simultaneous-equation SAR model that are analogous to the rank and order conditions for the classical simultaneous-equation model and derive asymptotic properties of the QML estimator. The QML estimator is asymptotically more efficient than the 3SLS estimator under normality but can be computationally difficult to implement.

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, 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 unknown form. Furthermore, we propose a heteroskedasticity-robust estimator for the asymptotic variance-covariance matrix 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 (2017), 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, we describe the model and give the moment conditions used to construct the GMM estimator. In Section 3, we establish the identification for the model under the GMM framework. We derive the asymptotic properties of the GMM estimator in Section 4. Results of Monte Carlo simulation experiments are reported in Section 5. Section 6 briefly concludes. Proofs are collected in the Appendix.

Throughout the paper, we adopt the following notation. For an $n \times n$ matrix $\mathbf{A} = [a_{ij}]$, let $\text{diag}(\mathbf{A})$ denote an $n \times n$ diagonal matrix with the i-th diagonal element being a_{ii}, i.e., $\text{diag}(\mathbf{A}) = \text{diag}(a_{11}, \cdots, a_{nn})$. Let $\rho(\mathbf{A})$ denotes the spectral radius of the square matrix $\mathbf{A}$. For an $n \times m$ matrix $\mathbf{B} = [b_{ij}]$, the vectorization of $\mathbf{B}$ is denoted by $\text{vec}(\mathbf{B}) = (b_{11}, \cdots, b_{n1}, b_{12}, \cdots, b_{nm})'$.[1] 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 Model and Moment Conditions

2.1 Model

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

$$\mathbf{Y} = \mathbf{Y}\mathbf{\Gamma}_0 + \mathbf{W}\mathbf{Y}\mathbf{\Lambda}_0 + \mathbf{X}\mathbf{B}_0 + \mathbf{U}, \tag{1}$$

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.[2] $\mathbf{W} = [w_{ij}]$ is an $n \times n$ nonstochastic matrix of spatial weights, with w_{ij} representing the proximity between cross sectional units i and j.[3] The diagonal elements of $\mathbf{W}$ are normalized to be zeros. In the literature, $\mathbf{WY}$ is usually referred to as the spatial lag. $\mathbf{\Gamma}_0$, $\mathbf{\Lambda}_0$ 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 $\mathbf{\Gamma}_0$ 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 $\mathbf{\Gamma}_0$, $\mathbf{\Lambda}_0$ 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 $\mathbf{Y}$, $\bar{\mathbf{Y}} = \mathbf{WY}$ and $\mathbf{X}$ that appear in the right hand side of the k-th equation.

[1] 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.

[2] 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.

[3] 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 model has a broad range of applications.

Then, the k-th equation of model (1) is

$$\mathbf{y}_k = \mathbf{Y}_k \boldsymbol{\gamma}_{k,0} + \bar{\mathbf{Y}}_k \boldsymbol{\lambda}_{k,0} + \mathbf{X}_k \boldsymbol{\beta}_{k,0} + \mathbf{u}_k. \tag{2}$$

We maintain the following assumptions regarding the DGP.

Assumption 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 *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 3 $\boldsymbol{\Gamma}_0$ *is nonsingular with a zero diagonal.* $\rho(\boldsymbol{\Lambda}_0(\mathbf{I}_m - \boldsymbol{\Gamma}_0)^{-1}) < 1/\rho(\mathbf{W})$.

Assumption 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 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 3 is from Yang and Lee (2017). Under this assumption, $\mathbf{S} = \mathbf{I}_{mn} - \boldsymbol{\Gamma}_0' \otimes \mathbf{I}_n - \boldsymbol{\Lambda}_0' \otimes \mathbf{W}$ is nonsingular, and hence the simultaneous-equation SAR model (1) has a well defined reduced form

$$\mathbf{y} = \mathbf{S}^{-1}(\mathbf{B}_0' \otimes \mathbf{I}_n)\mathbf{x} + \mathbf{S}^{-1}\mathbf{u}, \tag{3}$$

where $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$, $\mathbf{x} = \mathrm{vec}(\mathbf{X})$, and $\mathbf{u} = \mathrm{vec}(\mathbf{U})$. 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 constraint $|\lambda_{11,0}| < 1/\rho(\mathbf{W})$ for the single-equation SAR model.

2.2 Motivating examples

To illustrate the empirical relevance of the proposed model, we give two motivating examples in this subsection. The first example models the spillover effect of fiscal policies between local jurisdictions. The second example characterizes the conformity effect in social networks.

2.2.1 Fiscal policy interaction

Consider a set of n local jurisdictions. Each jurisdiction chooses expenditures in m publicly provided services to maximize the social welfare. The social welfare function of jurisdiction i is given by

$$V_i = \sum_{k=1}^{m}(\alpha_{ik} + \sum_{l=1}^{m}\rho_{lk}\sum_{j=1}^{n}w_{ij}y_{jl})y_{ik} - \frac{1}{2}\sum_{k=1}^{m}\sum_{l=1}^{m}\phi_{lk}y_{ik}y_{il}, \tag{4}$$

where α_{ik} captures the existing condition of publicly provided service k in jurisdiction i (and other exogenous characteristics of jurisdiction i), y_{ik} represents the spending by jurisdiction i on publicly provided service k, and w_{ij} measures the geographical proximity between jurisdictions i and j (e.g., one could let $w_{ij} = 1$ if jurisdictions i and j share a common border and $w_{ij} = 0$ otherwise).

The first term of the social welfare function (4) captures the social welfare gain from publicly provided services. The marginal social welfare gain of the expenditure y_{ik} is given by $\alpha_{ik} + \sum_{l=1}^{m}\rho_{lk}\sum_{j=1}^{n}w_{ij}y_{jl}$, where $\sum_{j=1}^{n}w_{ij}y_{jl}$ represents the spending by neighboring jurisdictions on publicly provided services and its coefficient ρ_{lk} captures the spillover effect of publicly provided services (see, e.g., Allers and Elhorst, 2011). For example, people may use libraries, parks and other recreation facilities in neighboring jurisdictions. In this case, the spending by a jurisdiction on such facilities might be considered as a substitute for the spending by a neighboring jurisdiction on similar or related facilities and thus generate a negative spillover effect (i.e. $\rho_{lk} < 0$). On the other hand, if the services provided by neighboring jurisdictions are complements (e.g. road networks and business parks, or hospitals and nursing homes), then the spillover effect is expected to be positive (i.e. $\rho_{lk} > 0$). The second term of the social welfare function (4) captures the cost of publicly provided services.[4] The coefficient ϕ_{lk} ($\phi_{lk} = \phi_{kl}$) represents the substitution effect between expenditures by a jurisdiction on publicly provided services k and l.

[4] See Hauptneier et al. (2012) for some discussion on including the cost of publicly provided services instead of imposing a budget constraint in the social welfare function.

Maximizing the social welfare function (4) yields the best response function

$$y_{ik} = \sum_{l=1,l\neq k}^{m} \gamma_{lk} y_{ik} + \sum_{l=1}^{m} \lambda_{lk} \sum_{j=1}^{n} w_{ij} y_{jl} + \alpha_{ik}^{*}$$

where $\gamma_{lk} = -\phi_{lk}/\phi_{kk}$, $\lambda_{lk} = \rho_{lk}/\phi_{kk}$, and $\alpha_{ik}^{*} = \alpha_{ik}/\phi_{kk}$. Suppose $\alpha_{ik}^{*} = \mathbf{x}_{ik}'\boldsymbol{\beta}_k + u_{ik}$, where $\mathbf{x}_{ik}$ is a vector of the observable characteristics of jurisdiction i and u_{ik} captures the unobservable heterogeneity of jurisdiction i. Then the best response function implies the simultaneous-equation SAR model (2).

2.2.2 Social conformity

Patacchini and Zenou (2012) consider a social conformity model where the social norm is given by the average behavior of peers in a certain activity. We generalize their model by defining the social norm based on a portfolio of difference activities. Suppose a set of n individuals interact in a social network. The network topology is captured by the matrix $\mathbf{W} = [w_{ij}]$. Let d_i denote the number of friends of individual i. A possible specification of $\mathbf{W}$ is such that $w_{ij} = 1/d_i$ if individuals i and j are friends and $w_{ij} = 0$ otherwise. Individual i choose effort levels $y_{i1}, \cdots, y_{im}$ simultaneously in m activities to maximize her utility function

$$U_i = \sum_{k=1}^{m} \alpha_{ik} y_{ik} - \frac{1}{2} \sum_{k=1}^{m} \sum_{l=1}^{m} \phi_{lk} y_{ik} y_{il} - \sum_{k=1}^{m} \frac{1}{2} \rho_k (y_{ik} - \sum_{l=1}^{m} \varrho_{lk} \sum_{j=1}^{n} w_{ij} y_{jl})^2. \quad (5)$$

The first term of the utility function (5) captures the payoff from the efforts with the productivity of individual i in activity k given by α_{ik}. The second term is the cost from the efforts with the substitution effect between efforts in different activities captured by ϕ_{lk} ($\phi_{lk} = \phi_{kl}$). The last term reflects the influence of an individual's friends on her own behavior. It is such that each individual wants to minimize the social distance between her own behavior y_{ik} to the social norm of that activity. The social norm for activity k is given by the weighted average behavior of her friends in m activities $\sum_{l=1}^{m} \varrho_{lk} \sum_{j=1}^{n} w_{ij} y_{jl}$ with the weights ϱ_{lk} such that $\sum_{l=1}^{m} \varrho_{lk} = 1$. The coefficient ρ_k captures the taste for conformity.

Maximizing the utility function (5) yields the best response function

$$y_{ik} = \sum_{l=1,l\neq k}^{m} \gamma_{lk} y_{ik} + \sum_{l=1}^{m} \lambda_{lk} \sum_{j=1}^{n} w_{ij} y_{jl} + \alpha_{ik}^{*}$$

where $\gamma_{lk} = -\phi_{lk}/(\phi_{kk} + \rho_k)$, $\lambda_{lk} = \rho_k \varrho_{lk}/(\phi_{kk} + \rho_k)$, and $\alpha^*_{ik} = \alpha_{ik}/(\phi_{kk} + \rho_k)$.[5] Suppose $\alpha^*_{ik} = \mathbf{x}'_{ik}\boldsymbol{\beta}_k + u_{ik}$. Then the best response function leads to the econometric model (2).

2.3 Moment conditions

Following Lee (2007) and Lin and Lee (2010), for the estimation of the simultaneous-equation SAR model (1), we consider both linear moment conditions

$$\mathrm{E}(\mathbf{Q}'\mathbf{u}_k) = \mathbf{0}, \tag{6}$$

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

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

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

$$\mathrm{E}(\mathbf{u}'_k \boldsymbol{\Xi}_r \mathbf{u}_l) = 0, \quad \text{for } r = 1, \cdots, p. \tag{7}$$

As an example, we could use $\mathbf{Q} = [\mathbf{X}, \mathbf{W}\mathbf{X}, \cdots, \mathbf{W}^p\mathbf{X}]$ and $\boldsymbol{\Xi}_1 = \mathbf{W}, \boldsymbol{\Xi}_2 = \mathbf{W}^2 - \mathrm{diag}(\mathbf{W}^2), \cdots, \boldsymbol{\Xi}_p = \mathbf{W}^p - \mathrm{diag}(\mathbf{W}^p)$, where p is some predetermined positive integer, to construct the linear and quadratic moment conditions. The quadratic moment conditions (7) exploit the covariance structure of model disturbances both within and across equations, and hence are more general than the quadratic moment conditions considered in Lee (2007) and Lin and Lee (2010).

Let the residual function for the k-th equation be

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

where $\boldsymbol{\theta}_k = (\boldsymbol{\gamma}'_k, \boldsymbol{\lambda}'_k, \boldsymbol{\beta}'_k)'$. The empirical linear moment functions based on (6) can be written as

$$\mathbf{g}_{1,k} \equiv \mathbf{g}_{1,k}(\boldsymbol{\theta}_k) = \mathbf{Q}'\mathbf{u}_k(\boldsymbol{\theta}_k) \tag{8}$$

[5] Suppose the parameters in the best response function can be identified, then parameters in the utility function (5) can be identified up to a scale factor. This issue is common for simultaneous-equation models (Schmidt, 1976). In this example, if λ_{lk} can be identified, then $\rho_k = \phi_{kk} \sum_{l=1}^m \lambda_{lk}/(1 - \sum_{l=1}^m \lambda_{lk})$, which is identifiable up to a scale factor ϕ_{kk}.

and the empirical quadratic moment functions based on (7) can be written as

$$\mathbf{g}_{2,kl} \equiv \mathbf{g}_{2,kl}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_l) = [\boldsymbol{\Xi}_1' \mathbf{u}_k(\boldsymbol{\theta}_k), \cdots, \boldsymbol{\Xi}_p' \mathbf{u}_k(\boldsymbol{\theta}_k)]' \mathbf{u}_l(\boldsymbol{\theta}_l) \tag{9}$$

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

$$\mathbf{g}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{g}_1(\boldsymbol{\theta}) \\ \mathbf{g}_2(\boldsymbol{\theta}) \end{bmatrix}, \tag{10}$$

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 (1) is then based on the moment conditions $\mathrm{E}[\mathbf{g}(\boldsymbol{\theta}_0)] = \mathbf{0}$. We maintain the following assumption regarding the moment conditions.

Assumption 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) $\lim_{n\to\infty} n^{-1}\boldsymbol{\Omega}$ *exists and is nonsingular, where* $\boldsymbol{\Omega} = \mathrm{Var}[\mathbf{g}(\boldsymbol{\theta}_0)]$.

3 Identification

Following Yang and Lee (2017), we establish the identification of the simultaneous-equation SAR model in two steps. In the first step, we consider the identification of the "pseudo" reduced form parameters in equation (11). In the second step, we recover the structural parameters from the "pseudo" reduced form parameters.

3.1 Identification of the "pseudo" reduced form parameters

When $\boldsymbol{\Gamma}_0$ is nonsingular, the simultaneous-equation SAR model (1) has a "pseudo" reduced form

$$\mathbf{Y} = \mathbf{W}\mathbf{Y}\boldsymbol{\Psi}_0 + \mathbf{X}\boldsymbol{\Pi}_0 + \mathbf{V}, \tag{11}$$

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 $\mathbf{V} = \mathbf{U}(\mathbf{I}_m - \boldsymbol{\Gamma}_0)^{-1}$. Equation (11) has the specification of a multivariate SAR model (see, Yang and Lee, 2017; 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 (11) 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

$$\mathbf{W}\mathbf{y}_l = \mathbf{H}_l(\boldsymbol{\Pi}_0' \otimes \mathbf{I}_n)\mathbf{x} + \mathbf{H}_l\mathbf{v} \tag{12}$$

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

$$\mathbf{v}_k(\boldsymbol{\delta}_k) = \mathbf{y}_k - \sum_{l=1}^{m} \psi_{lk}\mathbf{W}\mathbf{y}_l - \mathbf{X}\boldsymbol{\pi}_k = \mathbf{d}_k(\boldsymbol{\delta}_k) + \mathbf{v}_k + \sum_{l=1}^{m}(\psi_{lk,0} - \psi_{lk})\mathbf{H}_l\mathbf{v}, \tag{13}$$

where $\boldsymbol{\delta}_k = (\psi_{1k}, \cdots, \psi_{mk}, \boldsymbol{\pi}_k')'$ and $\mathbf{d}_k(\boldsymbol{\delta}_k) = [\mathrm{E}(\mathbf{W}\mathbf{y}_1), \cdots, \mathrm{E}(\mathbf{W}\mathbf{y}_m), \mathbf{X}](\boldsymbol{\delta}_{k,0} - \boldsymbol{\delta}_k)$.

The "pseudo" reduced form parameters in model (11) can be identified by the moment conditions described in the previous section. Similar to (8) and (9), the linear moment functions 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 functions can be written as

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

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 (13) 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 equations, $\lim_{n\to\infty} n^{-1}\mathrm{E}[\mathbf{f}_{1,k}(\boldsymbol{\delta}_k)] = \mathbf{0}$, have 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 $\text{rank}(\mathbf{Q}) \geq m+K_X$ for n sufficiently large.

If, however, $[\text{E}(\mathbf{W}\mathbf{y}_1), \cdots, \text{E}(\mathbf{W}\mathbf{y}_m), \mathbf{X}]$ does not have full column rank, then the model may still be identifiable via the quadratic moment conditions. Suppose for some $\bar{m} \in \{0, 1, \cdots, m-1\}$, $\text{E}(\mathbf{W}\mathbf{y}_l)$ and the columns of $[\text{E}(\mathbf{W}\mathbf{y}_1), \cdots, \text{E}(\mathbf{W}\mathbf{y}_{\bar{m}}), \mathbf{X}]$ are linearly dependent,[6] i.e., $\text{E}(\mathbf{W}\mathbf{y}_l) = \sum_{k=1}^{\bar{m}} c_{1,kl}\text{E}(\mathbf{W}\mathbf{y}_k) + \mathbf{X}\mathbf{c}_{2,l}$ for some vector of constants $(c_{1,1l}, \cdots, c_{1,\bar{m}l}, \mathbf{c}'_{2,l}) \in \mathbb{R}^{\bar{m}+K_X}$, for $l = \bar{m}+1, \cdots, m$. In this case,

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

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

$$\begin{aligned} \psi_{jk} &= \psi_{jk,0} + \sum_{l=\bar{m}+1}^{m} (\psi_{lk,0} - \psi_{lk})c_{1,jl} \\ \boldsymbol{\pi}_k &= \boldsymbol{\pi}_{k,0} + \sum_{l=\bar{m}+1}^{m} (\psi_{lk,0} - \psi_{lk})\mathbf{c}_{2,l}, \end{aligned} \tag{14}$$

for $j = 1, \cdots, \bar{m}$ and $k = 1, \cdots, m$, provided that $\mathbf{Q}'[\text{E}(\mathbf{W}\mathbf{y}_1), \cdots, \text{E}(\mathbf{W}\mathbf{y}_{\bar{m}}), \mathbf{X}]$ 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 conditions.

With $\boldsymbol{\delta}_k$ given by (14), we have

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

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

$$\begin{aligned} \lim_{n\to\infty} n^{-1}\{\sum_{i=1}^{m}(\psi_{ik,0} - \psi_{ik})\text{tr}[\mathbf{H}'_i\boldsymbol{\Xi}_r\text{E}(\mathbf{v}_l\mathbf{v}')] + \sum_{j=1}^{m}(\psi_{jl,0} - \psi_{jl})\text{tr}[\boldsymbol{\Xi}_r\mathbf{H}_j\text{E}(\mathbf{v}\mathbf{v}'_k)] \\ + \sum_{i=1}^{m}\sum_{j=1}^{m}(\psi_{ik,0} - \psi_{ik})(\psi_{jl,0} - \psi_{jl})\text{tr}[\mathbf{H}'_i\boldsymbol{\Xi}_r\mathbf{H}_j\text{E}(\mathbf{v}\mathbf{v}')]\} &= 0, \end{aligned} \tag{15}$$

[6] We adopt the convention that $[\text{E}(\mathbf{W}\mathbf{y}_1), \cdots, \text{E}(\mathbf{W}\mathbf{y}_{\bar{m}}), \mathbf{X}] = \mathbf{X}$ 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 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 (15), for* $r = 1, \cdots, p$ *and* $k, l = 1, \cdots, m$, *have a unique solution at* $\mathbf{\Psi}_0$.

Example 1 *To better understand the identification conditions in Assumption 7, consider the "pseudo" reduced form equations (11) with* $m = 2$ *and* $\mathbf{X} = [\underline{\mathbf{x}}, \mathbf{W}\underline{\mathbf{x}}]$, *where* $\underline{\mathbf{x}}$ *is* $n \times 1$ *vector of individual-specific exogenous characteristics. In this case, equations (11) can be written as*

$$\begin{aligned} \mathbf{y}_1 &= \psi_{11,0}\mathbf{W}\mathbf{y}_1 + \psi_{21,0}\mathbf{W}\mathbf{y}_2 + \pi_{11,0}\underline{\mathbf{x}} + \pi_{21,0}\mathbf{W}\underline{\mathbf{x}} + \mathbf{u}_1 \\ \mathbf{y}_2 &= \psi_{12,0}\mathbf{W}\mathbf{y}_1 + \psi_{22,0}\mathbf{W}\mathbf{y}_2 + \pi_{12,0}\underline{\mathbf{x}} + \pi_{22,0}\mathbf{W}\underline{\mathbf{x}} + \mathbf{u}_2. \end{aligned} \tag{16}$$

The "pseudo" reduced form equations (16) have the specification of a multivariate SAR model (Yang and Lee, 2017). In (16), $\mathbf{W}\underline{\mathbf{x}}$ *is a spatial lag of the exogenous variable and its coefficient captures the "contextual effect" (Manski, 1993). How to identify the endogenous peer effect captured by* $\psi_{kk,0}$ *from the contextual effect has been a major interest in the literature of social interaction models. In the multivariate SAR model (16), the identification problem is even more challenging because of the presence of the "cross-equation peer effect" captured by* $\psi_{lk,0}$ $(l \neq k)$.

Identification of model (16) can be achieved via linear moment conditions if Assumption 7 (i) *holds, or via linear and quadratic moment conditions if Assumption 7* (ii) *holds. A necessary condition for Assumption 7* (i) *is that* $[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \mathrm{E}(\mathbf{W}\mathbf{y}_2), \underline{\mathbf{x}}, \mathbf{W}\underline{\mathbf{x}}]$ *has full column rank for* n *sufficiently large. It follows by a similar argument as in Cohen-Cole et al. (2016) that* $[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \mathrm{E}(\mathbf{W}\mathbf{y}_2), \underline{\mathbf{x}}, \mathbf{W}\underline{\mathbf{x}}]$ *has full column rank if and only if the matrices* $\mathbf{I}_n, \mathbf{W}, \mathbf{W}^2, \mathbf{W}^3$ *are linearly independent*[8] *and the*

[7] A weaker identification condition can be derived based on (14) and (15) if the constants $c_{1,1l}, \cdots, c_{1,\bar{m}l}, \mathbf{c}_{2,l}$ are known to the researcher.

[8] For example, the weights matrix $\mathbf{W}$ for a star network is given by

$$\mathbf{W} = \begin{bmatrix} 0 & (n-1)^{-1} & \cdots & (n-1)^{-1} \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}.$$

As $\mathbf{W}^3 = \mathbf{W}$, the linear independence of $\mathbf{I}_n, \mathbf{W}, \mathbf{W}^2, \mathbf{W}^3$ does not hold.

parameter matrix

$$\begin{bmatrix} \pi_{11,0} & \pi_{12,0} & 1 \\ \psi_{21,0}\pi_{12,0} - \psi_{22,0}\pi_{11,0} + \pi_{21,0} & \psi_{12,0}\pi_{11,0} - \psi_{11,0}\pi_{12,0} + \pi_{22,0} & -(\psi_{11,0} + \psi_{22,0}) \\ \psi_{21,0}\pi_{22,0} - \psi_{22,0}\pi_{21,0} & \psi_{12,0}\pi_{21,0} - \psi_{11,0}\pi_{22,0} & \psi_{11,0}\psi_{22,0} - \psi_{12,0}\psi_{21,0} \end{bmatrix} \quad (17)$$

has full column rank.

Suppose $\pi_{11,0} = \pi_{21,0} = \pi_{12,0} = \pi_{22,0} = 0$ *in the DGP. Then the parameter matrix (17) does not have full column rank. Therefore,* $[\mathrm{E}(\mathbf{W}\mathbf{y}_1), \mathrm{E}(\mathbf{W}\mathbf{y}_2), \underline{\mathbf{x}}, \mathbf{W}\underline{\mathbf{x}}]$ *does not have full column rank and Assumption 7* (i) *fails to hold. In this case, model (16) can be identified if Assumption 7* (ii) *holds. From (12),* $\mathrm{E}(\mathbf{W}\mathbf{y}_k) = \mathbf{H}_k(\mathbf{\Pi}_0' \otimes \mathbf{I}_n)\mathbf{x} = \mathbf{0}$ *as* $\mathbf{\Pi}_0 = \mathbf{0}$*, which implies that* $\mathbf{d}_k(\boldsymbol{\delta}_k) = [\mathrm{E}(\mathbf{W}\mathbf{y}_1), \mathrm{E}(\mathbf{W}\mathbf{y}_m), \underline{\mathbf{x}}, \mathbf{W}\underline{\mathbf{x}}](\boldsymbol{\delta}_{k,0} - \boldsymbol{\delta}_k) = (\pi_{1k,0} - \pi_{1k})\underline{\mathbf{x}} + (\pi_{2k,0} - \pi_{2k})\mathbf{W}\underline{\mathbf{x}}$*. Hence, from the linear moment equations* $\lim_{n\to\infty} n^{-1}\mathbf{Q}'\mathbf{d}_k(\boldsymbol{\theta}_k) = \mathbf{0}$*, only* $\pi_{1k,0}$ *and* $\pi_{2k,0}$ *can be identified if* $\lim_{n\to\infty} n^{-1}\mathbf{Q}'[\underline{\mathbf{x}}, \mathbf{W}\underline{\mathbf{x}}]$ *exists and has full column rank. Identification of* $\psi_{lk,0}$ *has to be achieved via the quadratic moment equations (15).*

It follows from (15) that, for $k = l = 1$*,*

$$\begin{aligned} &(\psi_{11,0} - \psi_{11}) \lim_{n\to\infty} n^{-1}\mathrm{tr}[\mathbf{H}_1'(\mathbf{\Xi}_r + \mathbf{\Xi}_r')\mathrm{E}(\mathbf{v}_1\mathbf{v}')] + (\psi_{21,0} - \psi_{21}) \lim_{n\to\infty} n^{-1}\mathrm{tr}[\mathbf{H}_2'(\mathbf{\Xi}_r + \mathbf{\Xi}_r')\mathrm{E}(\mathbf{v}_1\mathbf{v}')] \\ &+(\psi_{11,0} - \psi_{11})^2 \lim_{n\to\infty} n^{-1}\mathrm{tr}[\mathbf{H}_1'\mathbf{\Xi}_r\mathbf{H}_1\mathrm{E}(\mathbf{v}\mathbf{v}')] + (\psi_{21,0} - \psi_{21})^2 \lim_{n\to\infty} n^{-1}\mathrm{tr}[\mathbf{H}_2'\mathbf{\Xi}_r\mathbf{H}_2\mathrm{E}(\mathbf{v}\mathbf{v}')] \\ &+(\psi_{11,0} - \psi_{11})(\psi_{21,0} - \psi_{21}) \lim_{n\to\infty} n^{-1}\mathrm{tr}[\mathbf{H}_1'(\mathbf{\Xi}_r + \mathbf{\Xi}_r')\mathbf{H}_2\mathrm{E}(\mathbf{v}\mathbf{v}')] = 0, \end{aligned} \quad (18)$$

for $r = 1, \cdots, p$*. If the matrix*

$$\lim_{n\to\infty} n^{-1} \begin{bmatrix} \mathrm{tr}[\mathbf{H}_1'(\mathbf{\Xi}_1 + \mathbf{\Xi}_1')\mathrm{E}(\mathbf{v}_1\mathbf{v}')] & \cdots & \mathrm{tr}[\mathbf{H}_1'(\mathbf{\Xi}_p + \mathbf{\Xi}_p')\mathrm{E}(\mathbf{v}_1\mathbf{v}')] \\ \mathrm{tr}[\mathbf{H}_2'(\mathbf{\Xi}_1 + \mathbf{\Xi}_1')\mathrm{E}(\mathbf{v}_1\mathbf{v}')] & \cdots & \mathrm{tr}[\mathbf{H}_2'(\mathbf{\Xi}_p + \mathbf{\Xi}_p')\mathrm{E}(\mathbf{v}_1\mathbf{v}')] \\ \mathrm{tr}[\mathbf{H}_1'\mathbf{\Xi}_1\mathbf{H}_1\mathrm{E}(\mathbf{v}\mathbf{v}')] & \cdots & \mathrm{tr}[\mathbf{H}_1'\mathbf{\Xi}_p\mathbf{H}_1\mathrm{E}(\mathbf{v}\mathbf{v}')] \\ \mathrm{tr}[\mathbf{H}_2'\mathbf{\Xi}_1\mathbf{H}_2\mathrm{E}(\mathbf{v}\mathbf{v}')] & \cdots & \mathrm{tr}[\mathbf{H}_2'\mathbf{\Xi}_p\mathbf{H}_2\mathrm{E}(\mathbf{v}\mathbf{v}')] \\ \mathrm{tr}[\mathbf{H}_1'(\mathbf{\Xi}_1 + \mathbf{\Xi}_1')\mathbf{H}_2\mathrm{E}(\mathbf{v}\mathbf{v}')] & \cdots & \mathrm{tr}[\mathbf{H}_1'(\mathbf{\Xi}_p + \mathbf{\Xi}_p')\mathbf{H}_2\mathrm{E}(\mathbf{v}\mathbf{v}')] \end{bmatrix}$$

has full row rank, (18) has a unique solution at $(\psi_{11,0}, \psi_{21,0})$ *and hence* $(\psi_{11,0}, \psi_{21,0})$ *can be identified. Similarly,* $(\psi_{12,0}, \psi_{22,0})$ *can be identified from (15) with* $k = l = 2$*.*

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-equation model (see, e.g., Schmidt, 1976). 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 as in Yang and Lee (2017), the sufficient and necessary *rank* condition for identification is $\text{rank}(\mathbf{R}_k\mathbf{\Theta}_0) = m-1$, and the necessary *order* condition is $R_k \geq m-1$, for $k = 1, \cdots, m$.

Assumption 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*

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

Example 2 *To better understand the rank condition in Assumption 8, consider the model*

$$\begin{aligned} \mathbf{y}_1 &= \gamma_{21,0}\mathbf{y}_2 + \lambda_{11,0}\mathbf{W}\mathbf{y}_1 + \mathbf{X}\boldsymbol{\beta}_{1,0} + \mathbf{u}_1 \\ \mathbf{y}_2 &= \gamma_{12,0}\mathbf{y}_1 + \lambda_{22,0}\mathbf{W}\mathbf{y}_2 + \mathbf{X}\boldsymbol{\beta}_{2,0} + \mathbf{u}_2 \end{aligned} \tag{19}$$

with "pseudo" reduced-form equations

$$\begin{aligned} \mathbf{y}_1 &= \psi_{11,0}\mathbf{W}\mathbf{y}_1 + \psi_{21,0}\mathbf{W}\mathbf{y}_2 + \mathbf{X}\boldsymbol{\pi}_{1,0} + \mathbf{v}_1 \\ \mathbf{y}_2 &= \psi_{12,0}\mathbf{W}\mathbf{y}_1 + \psi_{22,0}\mathbf{W}\mathbf{y}_2 + \mathbf{X}\boldsymbol{\pi}_{2,0} + \mathbf{v}_2, \end{aligned} \tag{20}$$

where

$$\begin{bmatrix} \psi_{11,0} & \psi_{12,0} \\ \psi_{21,0} & \psi_{22,0} \end{bmatrix} = (1-\gamma_{12,0}\gamma_{21,0})^{-1} \begin{bmatrix} \lambda_{11,0} & \gamma_{12,0}\lambda_{11,0} \\ \gamma_{21,0}\lambda_{22,0} & \lambda_{22,0} \end{bmatrix} \tag{21}$$

and

$$[\boldsymbol{\pi}_{1,0}, \boldsymbol{\pi}_{2,0}] = (1-\gamma_{12,0}\gamma_{21,0})^{-1}[\boldsymbol{\beta}_{1,0} + \gamma_{21,0}\boldsymbol{\beta}_{2,0}, \boldsymbol{\beta}_{2,0} + \gamma_{12,0}\boldsymbol{\beta}_{1,0}]. \tag{22}$$

Suppose Assumption 7 holds for (20) and thus the "pseudo" reduced-form parameters can be identi-

fied. Then, the structural parameters

$$\boldsymbol{\Theta}_0 = [(\mathbf{I}_m - \boldsymbol{\Gamma}_0)', -\boldsymbol{\Lambda}_0', -\mathbf{B}_0']' = \begin{bmatrix} 1 & -\gamma_{21,0} & -\lambda_{11,0} & 0 & -\boldsymbol{\beta}_{1,0}' \\ -\gamma_{12,0} & 1 & 0 & -\lambda_{22,0} & -\boldsymbol{\beta}_{2,0}' \end{bmatrix}'$$

can be identified from (21) and (22) if the rank condition holds. The exclusion restriction for the first equation of model (19) can be represented by $\mathbf{R}_1 = [0, 0, 0, -1, \mathbf{0}_{1\times K_X}]$*. Then* $\mathbf{R}_1\boldsymbol{\Theta}_0 = [0, \lambda_{22,0}]$*, which has rank* 1 *if* $\lambda_{22,0} \neq 0$*. Similarly, the exclusion restriction for the second equation can be represented by* $\mathbf{R}_2 = [0, 0, -1, 0, \mathbf{0}_{1\times K_X}]$*. Then* $\mathbf{R}_2\boldsymbol{\Theta}_0 = [\lambda_{11,0}, 0]$ *which has rank* 1 *if* $\lambda_{11,0} \neq 0$*. Indeed, if* $\lambda_{11,0} = \lambda_{22,0} = 0$ *in the DGP, then (19) becomes a classical linear simultaneous equations model, which cannot be identified without imposing exclusion restrictions on* $\boldsymbol{\beta}_{k,0}$*.*

4 GMM Estimation

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 (1) is given by

$$\widetilde{\boldsymbol{\theta}}_{gmm} = \arg\min \mathbf{g}(\boldsymbol{\theta})'\mathbf{F}'\mathbf{F}\mathbf{g}(\boldsymbol{\theta}) \tag{23}$$

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 $\dim(\boldsymbol{\theta})$. Usually, $\mathbf{F}'\mathbf{F}$ is referred to as the GMM weighting matrix.

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 A.1 and A.2 in the Appendix that

$$\boldsymbol{\Omega} = \mathrm{Var}[\mathbf{g}(\boldsymbol{\theta}_0)] = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{22} \end{bmatrix}, \tag{24}$$

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

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

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,

$$\mathbf{D} = -\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta}_0)] = [\mathbf{D}'_1, \mathbf{D}'_2]', \tag{25}$$

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

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

with $\mathbf{\Upsilon}_{1,kl} = [\mathrm{E}(\mathbf{Z}'_k\mathbf{\Xi}_1\mathbf{u}_l), \cdots, \mathrm{E}(\mathbf{Z}'_k\mathbf{\Xi}_p\mathbf{u}_l)]'$ and $\mathbf{\Upsilon}_{2,kl} = [\mathrm{E}(\mathbf{Z}'_l\mathbf{\Xi}'_1\mathbf{u}_k), \cdots, \mathrm{E}(\mathbf{Z}'_l\mathbf{\Xi}'_p\mathbf{u}_k)]'$. In the following proposition we establish consistency and asymptotic normality of the GMM estimator $\widetilde{\boldsymbol{\theta}}_{gmm}$ defined in (23).

Proposition 1 *Suppose Assumptions 1-8 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

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

with $\boldsymbol{\Omega}$ *and* $\mathbf{D}$ *defined in (24) and (25) respectively.*

With $\mathbf{F}'\mathbf{F}$ in (23) replaced by $(n^{-1}\boldsymbol{\Omega})^{-1}$, $\text{AsyVar}(\widetilde{\boldsymbol{\theta}}_{gmm})$ 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 matrix. 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 $n^{-1}\mathbf{D}$ under heteroskedasticity of unknown form. With consistently estimated $n^{-1}\boldsymbol{\Omega}$ and $n^{-1}\mathbf{D}$, the feasible optimal GMM estimator and its heteroskedasticity-robust standard error can be obtained.

Proposition 2 *Suppose Assumptions 1-8 hold. Let* $\widetilde{\boldsymbol{\theta}}$ *be a consistent estimator of* $\boldsymbol{\theta}_0$ *and* $\widetilde{\boldsymbol{\Sigma}}_{kl} = \text{diag}(\widetilde{u}_{1k}\widetilde{u}_{1l}, \cdots, \widetilde{u}_{nk}\widetilde{u}_{nl})$ *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,* $n^{-1}\widetilde{\mathbf{D}} - n^{-1}\mathbf{D} = \text{o}_\text{p}(1)$ *and* $n^{-1}\widetilde{\boldsymbol{\Omega}} - n^{-1}\boldsymbol{\Omega} = \text{o}_\text{p}(1)$.

Finally Proposition 3 establishes asymptotic normality of the feasible optimal GMM estimator.

Proposition 3 *Suppose Assumptions 1-8 hold. The optimal GMM estimator is given by*

$$\widehat{\boldsymbol{\theta}}_{gmm} = \arg\min \mathbf{g}(\boldsymbol{\theta})'\widetilde{\boldsymbol{\Omega}}\mathbf{g}(\boldsymbol{\theta}), \tag{26}$$

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

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 $\widetilde{\mathbf{P}} = (\mathbf{I}_m \otimes \mathbf{Q})[(\mathbf{I}_m \otimes \mathbf{Q}')\widetilde{\mathbf{\Sigma}}(\mathbf{I}_m \otimes \mathbf{Q})]^{-1}(\mathbf{I}_m \otimes \mathbf{Q}')$. Similar to Proposition 3, we can show that

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

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

4.2 Best moment conditions under homoskedasticity

The above optimal GMM estimator is only "optimal" given the chosen moment conditions. The asymptotic efficiency of the optimal GMM estimator can be improved by choosing the "best" moment conditions. As discussed in Lin and Lee (2010), under heteroskedasticity of unknown form, the best moment conditions may not be available. However, under homoskedasticity, it is possible to find the best linear and quadratic moment conditions with $\mathbf{Q}$ and $\mathbf{\Xi}_r$'s satisfying Assumption 6. In general, the best moment conditions depend on the model specification. For expositional purpose, we consider a two-equation SAR model given by

$$\begin{aligned} \mathbf{y}_1 &= \gamma_{21,0}\mathbf{y}_2 + \lambda_{11,0}\mathbf{W}\mathbf{y}_1 + \lambda_{21,0}\mathbf{W}\mathbf{y}_2 + \mathbf{X}_1\boldsymbol{\beta}_{1,0} + \mathbf{u}_1 \\ \mathbf{y}_2 &= \gamma_{12,0}\mathbf{y}_1 + \lambda_{12,0}\mathbf{W}\mathbf{y}_1 + \lambda_{22,0}\mathbf{W}\mathbf{y}_2 + \mathbf{X}_2\boldsymbol{\beta}_{2,0} + \mathbf{u}_2 \end{aligned} \tag{27}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are respectively $n \times K_1$ and $n \times K_2$ sub-matrices of $\mathbf{X}$. Suppose $\mathbf{u}_1$ and $\mathbf{u}_2$ are $n \times 1$ vectors of i.i.d. random variables with zero mean and $\text{E}(\mathbf{u}_1\mathbf{u}_1') = \sigma_{11}\mathbf{I}_n$, $\text{E}(\mathbf{u}_2\mathbf{u}_2') = \sigma_{22}\mathbf{I}_n$ and $\text{E}(\mathbf{u}_1\mathbf{u}_2') = \sigma_{12}\mathbf{I}_n$. The reduced form of (27) is

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \mathbf{S}^{-1} \begin{bmatrix} \mathbf{X}_1\boldsymbol{\beta}_{1,0} \\ \mathbf{X}_2\boldsymbol{\beta}_{2,0} \end{bmatrix} + \mathbf{S}^{-1} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \tag{28}$$

where

$$\mathbf{S} = \mathbf{I}_{2n} - \mathbf{\Gamma}_0' \otimes \mathbf{I}_n - \mathbf{\Lambda}_0' \otimes \mathbf{W} = \begin{bmatrix} 1 & -\gamma_{12,0} \\ -\gamma_{21,0} & 1 \end{bmatrix}' \otimes \mathbf{I}_n - \begin{bmatrix} \lambda_{11,0} & \lambda_{12,0} \\ \lambda_{21,0} & \lambda_{22,0} \end{bmatrix}' \otimes \mathbf{W}.$$

Let $(\mathbf{S}^{-1})_{kl}$ denote the (k,l)-th block matrix of $\mathbf{S}^{-1}$, i.e., $(\mathbf{S}^{-1})_{kl} = (\mathbf{i}'_{2,k} \otimes \mathbf{I}_n)\mathbf{S}^{-1}(\mathbf{i}_{2,l} \otimes \mathbf{I}_n)$, for $k,l = 1,2$. Then, from (28),

$$\begin{aligned} \mathbf{y}_1 &= (\mathbf{S}^{-1})_{11}\mathbf{X}_1\boldsymbol{\beta}_{1,0} + (\mathbf{S}^{-1})_{12}\mathbf{X}_2\boldsymbol{\beta}_{2,0} + (\mathbf{S}^{-1})_{11}\mathbf{u}_1 + (\mathbf{S}^{-1})_{12}\mathbf{u}_2 \\ \mathbf{y}_2 &= (\mathbf{S}^{-1})_{21}\mathbf{X}_1\boldsymbol{\beta}_{1,0} + (\mathbf{S}^{-1})_{22}\mathbf{X}_2\boldsymbol{\beta}_{2,0} + (\mathbf{S}^{-1})_{21}\mathbf{u}_1 + (\mathbf{S}^{-1})_{22}\mathbf{u}_2. \end{aligned} \tag{29}$$

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 variance-covariance matrix for the optimal GMM estimator defined in (26) is $\text{AsyVar}(\widehat{\boldsymbol{\theta}}_{gmm}) = (\lim_{n\to\infty} n^{-1}\mathbf{D}'\boldsymbol{\Omega}\mathbf{D})^{-1}$. Under homoskedasticity, $\boldsymbol{\Omega}$ defined in (24) can be simplified so that

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

and

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

with

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

Furthermore, $\mathbf{D}$ defined in (25) can be simplified so that

$$\mathbf{D}_1 = -\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}_1(\boldsymbol{\theta}_0)] = \begin{bmatrix} \mathbf{Q}'\mathrm{E}(\mathbf{Z}_1) & \\ & \mathbf{Q}'\mathrm{E}(\mathbf{Z}_2) \end{bmatrix} \tag{30}$$

and

$$\mathbf{D}_2 = -\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}_2\left(\boldsymbol{\theta}_0\right)] = \begin{bmatrix} \boldsymbol{\Upsilon}_{1,11} & \\ \boldsymbol{\Upsilon}_{1,12} & \\ & \boldsymbol{\Upsilon}_{1,21} \\ & \boldsymbol{\Upsilon}_{1,22} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Upsilon}_{2,11} & \\ & \boldsymbol{\Upsilon}_{2,12} \\ \boldsymbol{\Upsilon}_{2,21} & \\ & \boldsymbol{\Upsilon}_{2,22} \end{bmatrix}, \tag{31}$$

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 $\mathbf{G}_{kl} = \mathbf{W}(\mathbf{S}^{-1})_{kl}$ for $k, l = 1, 2$. From the reduced form (29),

$$\mathrm{E}(\mathbf{Z}_k) = [(\mathbf{S}^{-1})_{3-k,1}\mathbf{X}_1\boldsymbol{\beta}_{1,0}+(\mathbf{S}^{-1})_{3-k,2}\mathbf{X}_2\boldsymbol{\beta}_{2,0}, \mathbf{G}_{11}\mathbf{X}_1\boldsymbol{\beta}_{1,0}+\mathbf{G}_{12}\mathbf{X}_2\boldsymbol{\beta}_{2,0}, \mathbf{G}_{21}\mathbf{X}_1\boldsymbol{\beta}_{1,0}+\mathbf{G}_{22}\mathbf{X}_2\boldsymbol{\beta}_{2,0}, \mathbf{X}_k]$$

and

$$
\mathrm{E}(\mathbf{Z}_k'\mathbf{A}\mathbf{u}_l) = \begin{bmatrix} \sigma_{1l}\mathrm{tr}[\mathbf{A}'(\mathbf{S}^{-1})_{3-k,1}] + \sigma_{l2}\mathrm{tr}[\mathbf{A}'(\mathbf{S}^{-1})_{3-k,2}] \\ \sigma_{1l}\mathrm{tr}(\mathbf{A}'\mathbf{G}_{11}) + \sigma_{l2}\mathrm{tr}(\mathbf{A}'\mathbf{G}_{12}) \\ \sigma_{1l}\mathrm{tr}(\mathbf{A}'\mathbf{G}_{21}) + \sigma_{l2}\mathrm{tr}(\mathbf{A}'\mathbf{G}_{22}) \\ \mathbf{0} \end{bmatrix}
$$

for $k, l = 1, 2$, where $\mathbf{A}$ can be replaced by either Ξ_r or Ξ_r' for $r = 1, \cdots, p$.

As we can see now, both $\boldsymbol{\Omega}$ and $\mathbf{D}$ depend on $\mathbf{Q}$ and Ξ_r's. The best $\mathbf{Q}$ and Ξ_r's are those that minimize the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\theta}}_{gmm}$ given by $(\lim_{n\to\infty} n^{-1}\mathbf{D}'\boldsymbol{\Omega}\mathbf{D})^{-1}$. Under homoskedasticity, we can use the criterion for the redundancy of moment conditions (Breusch et al., 1999) to show that the best $\mathbf{Q}$ and Ξ_r's satisfying Assumption 6 are

$$
\begin{aligned}
\mathbf{Q}^* &= [\mathbf{X}, (\mathbf{S}^{-1})_{11}\mathbf{X}_1, (\mathbf{S}^{-1})_{12}\mathbf{X}_2, (\mathbf{S}^{-1})_{21}\mathbf{X}_1, (\mathbf{S}^{-1})_{22}\mathbf{X}_2, \mathbf{G}_{11}\mathbf{X}_1, \mathbf{G}_{12}\mathbf{X}_2, \mathbf{G}_{21}\mathbf{X}_1, \mathbf{G}_{22}\mathbf{X}_2] \\
\Xi_1^* &= (\mathbf{S}^{-1})_{11} - \mathrm{diag}[(\mathbf{S}^{-1})_{11}] \qquad \Xi_5^* = \mathbf{G}_{11} - \mathrm{diag}(\mathbf{G}_{11}) \\
\Xi_2^* &= (\mathbf{S}^{-1})_{12} - \mathrm{diag}[(\mathbf{S}^{-1})_{12}] \qquad \Xi_6^* = \mathbf{G}_{12} - \mathrm{diag}(\mathbf{G}_{12}) \\
\Xi_3^* &= (\mathbf{S}^{-1})_{21} - \mathrm{diag}[(\mathbf{S}^{-1})_{21}] \qquad \Xi_7^* = \mathbf{G}_{21} - \mathrm{diag}(\mathbf{G}_{21}) \\
\Xi_4^* &= (\mathbf{S}^{-1})_{22} - \mathrm{diag}[(\mathbf{S}^{-1})_{22}] \qquad \Xi_8^* = \mathbf{G}_{22} - \mathrm{diag}(\mathbf{G}_{22}).
\end{aligned} \tag{32}
$$

Proposition 4 *For model (27) with homoskedastic disturbances, the best* $\mathbf{Q}$ *and* Ξ_r*'s that satisfy Assumption 6 are given in (32).*

The best moment conditions with $\mathbf{Q}^*$ and Ξ_r^*'s given in (32) are not feasible as $\mathbf{Q}^*$ and Ξ_r^*'s involve unknown parameters $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Lambda}_0$. With some consistent preliminary estimators for $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Lambda}_0$, the feasible best moment conditions can be obtained in a similar manner as in Lee (2007). Note that, $(\mathbf{S}^{-1})_{kl} = \sum_{j=0}^{\infty} c_j\mathbf{W}^j$ for some constants $c_0, c_1, \cdots$, for $k, l = 1, 2$. Thus, we can use the leading order terms of the series expansion, i.e., $\mathbf{I}_n, \mathbf{W}, \cdots, \mathbf{W}^p$ to approximate the unknown $(\mathbf{S}^{-1})_{kl}$ and $\mathbf{G}_{kl} = \mathbf{W}(\mathbf{S}^{-1})_{kl}$ in $\mathbf{Q}^*$ and Ξ_r^*'s. Since the approximation error goes to zero very fast as p increases, in practice, we could use $\mathbf{Q} = [\mathbf{X}, \mathbf{W}\mathbf{X}, \cdots, \mathbf{W}^p\mathbf{X}]$ and $\Xi_1 = \mathbf{W}, \Xi_2 = \mathbf{W}^2 - \mathrm{diag}(\mathbf{W}^2), \cdots, \Xi_p = \mathbf{W}^p - \mathrm{diag}(\mathbf{W}^p)$ for some small p to construct the moment conditions.

5 Monte Carlo

To study the finite sample performance of the proposed GMM estimator, we conduct a small Monte Carlo simulation experiment. The model considered in the experiment is given by

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

In the DGP, we set $\gamma_{21,0} = \gamma_{12,0} = 0.2$, $\lambda_{11,0} = \lambda_{22,0} = 0.4$, and $\lambda_{21,0} = \lambda_{12,0} = 0.2$. For the spatial weights matrix $\mathbf{W} = [w_{ij}]$, we take guidance from the specification in Kelejian and Prucha (2010). More specifically, we divide the sample into two halves, with each half formulating a circle. In the first circle (containing $n/2$ cross sectional units), every unit is connected with the unit ahead of it and the unit behind. In the second circle, every unit is connected with the 9 units ahead of it and the 9 units behind. There are no connections across those two circles. Let d_i be the number of connections of the i-th unit. We set $w_{ij} = 1/d_i$ if units i and j are connected and $w_{ij} = 0$ otherwise. We generate $\mathbf{x}_1$ and $\mathbf{x}_2$ from independent standard normal distributions. We generate $\mathbf{u}_1 = (u_{11}, \cdots, u_{n1})'$ and $\mathbf{u}_2 = (u_{12}, \cdots, u_{n2})'$ such that $u_{i1} = \sqrt{\varsigma_i}\epsilon_{i1}$ and $u_{i2} = \sqrt{\varsigma_i}\epsilon_{i2}$, where ϵ_{i1} and ϵ_{i2} are respectively the i-th elements of $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ with

$$\begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{bmatrix} \sim \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{I}_n & \sigma_{12}\mathbf{I}_n \\ \sigma_{12}\mathbf{I}_n & \mathbf{I}_n \end{bmatrix}\right).$$

For the heteroskedastic case, we set $\varsigma_i = \sqrt{d_i/2}$. The average variances of $\mathbf{u}_1$ and $\mathbf{u}_2$ are 2. For the homoskedastic case, we set $\varsigma_i = 2$.

We conduct 1000 replications in the simulation experiment for different specifications with $n \in \{250, 500\}$, $\sigma_{12} \in \{0.3, 0.5, 0.7\}$, and $(\beta_{1,0}, \beta_{2,0}) \in \{(1,1), (0.5, 0.5)\}$. Let

$$\begin{aligned}\mathbf{u}_1(\boldsymbol{\theta}_1) &= \mathbf{y}_1 - \gamma_{21}\mathbf{y}_2 - \lambda_{11}\mathbf{W}\mathbf{y}_1 - \lambda_{21}\mathbf{W}\mathbf{y}_2 - \beta_1\mathbf{x}_1 \\ \mathbf{u}_2(\boldsymbol{\theta}_2) &= \mathbf{y}_2 - \gamma_{12}\mathbf{y}_1 - \lambda_{12}\mathbf{W}\mathbf{y}_1 - \lambda_{22}\mathbf{W}\mathbf{y}_2 - \beta_2\mathbf{x}_2,\end{aligned}$$

where $\boldsymbol{\theta}_1 = (\gamma_{21}, \lambda_{11}, \lambda_{21}, \beta_1)$ and $\boldsymbol{\theta}_2 = (\gamma_{12}, \lambda_{12}, \lambda_{22}, \beta_2)$. Let $\mathbf{Q} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{W}\mathbf{x}_1, \mathbf{W}\mathbf{x}_2, \mathbf{W}^2\mathbf{x}_1, \mathbf{W}^2\mathbf{x}_2]$ and $\boldsymbol{\Xi} = \mathbf{W}$. We consider the following estimators:

(a) The 2SLS estimator of $\boldsymbol{\theta}_1$ based on the linear moment function $\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\theta}_1)$.

(b) The 3SLS estimator of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ based on the linear moment function $(\mathbf{I}_2 \otimes \mathbf{Q})'\mathbf{u}(\boldsymbol{\theta})$, where $\mathbf{u}(\boldsymbol{\theta}) = [\mathbf{u}_1(\boldsymbol{\theta}_1)', \mathbf{u}_2(\boldsymbol{\theta}_2)']'$.

(c) The single-equation GMM (GMM-1) estimator of $\boldsymbol{\theta}_1$ based on the linear moment function $\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\theta}_1)$ and quadratic moment function $\mathbf{u}_1(\boldsymbol{\theta}_1)'\boldsymbol{\Xi}\mathbf{u}_1(\boldsymbol{\theta}_1)$.

(d) The system GMM (GMM-2) estimator of $\boldsymbol{\theta}$ based on the linear moment function $(\mathbf{I}_2 \otimes \mathbf{Q})'\mathbf{u}(\boldsymbol{\theta})$ and the quadratic moment functions $\mathbf{u}_1(\boldsymbol{\theta}_1)'\boldsymbol{\Xi}\mathbf{u}_1(\boldsymbol{\theta}_1)$, $\mathbf{u}_1(\boldsymbol{\theta}_1)'\boldsymbol{\Xi}\mathbf{u}_2(\boldsymbol{\theta}_2)$, and $\mathbf{u}_2(\boldsymbol{\theta}_2)'\boldsymbol{\Xi}\mathbf{u}_2(\boldsymbol{\theta}_2)$.

(e) The QML estimator of $\boldsymbol{\theta}$ described in Yang and Lee (2017).

Among the above estimators, the 2SLS and GMM-1 are equation-by-equation "limited information" estimators, while 3SLS, GMM-2 and QML are "full information" estimators. To obtain the heteroskedasticity-robust optimal weighting matrix for the 3SLS and GMM estimators, we consider preliminary estimators of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ given by

$$\begin{aligned}\widetilde{\boldsymbol{\theta}}_1 &= \arg\min \mathbf{u}_1(\boldsymbol{\theta}_1)'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{u}_1(\boldsymbol{\theta}_1) + [\mathbf{u}_1(\boldsymbol{\theta}_1)'\boldsymbol{\Xi}\mathbf{u}_1(\boldsymbol{\theta}_1)]^2 \\ \widetilde{\boldsymbol{\theta}}_2 &= \arg\min \mathbf{u}_2(\boldsymbol{\theta}_2)'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{u}_2(\boldsymbol{\theta}_2) + [\mathbf{u}_2(\boldsymbol{\theta}_2)'\boldsymbol{\Xi}\mathbf{u}_2(\boldsymbol{\theta}_2)]^2.\end{aligned}$$

The estimation results are reported in Tables 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.

[Insert Tables 1-4 here]

- With homoskedastic disturbances, all the estimators are essentially unbiased when $n = 500$. The QML estimator for $(\lambda_{11,0}, \lambda_{21,0})$ has the smallest SD with the GMM-2 estimator being a close runner up. As reported in Table 1, when $n = 500$ and $\sigma_{12} = 0.5$, the QML estimator of $\lambda_{11,0}$ reduces the SDs of the 2SLS, 3SLS, GMM-1 and GMM-2 estimators by, respectively, 33.9%, 34.5%, 15.9% and 5.1%; and the QML estimator of $\lambda_{21,0}$ reduces the SDs of the 2SLS, 3SLS, GMM-1 and GMM-2 estimators by, respectively, 24.1%, 24.7%, 22.1% and 7.4%. On the other hand, the SDs of all the estimators for $\gamma_{21,0}$ and $\beta_{1,0}$ are largely the same.

- With heteroskedastic disturbances, the QML estimator for $(\lambda_{11,0}, \lambda_{21,0})$ is downwards biased. The bias remains as sample size increases. The GMM-2 estimator for $(\lambda_{11,0}, \lambda_{21,0})$ has the smallest SD. As reported in Table 2, when $n = 500$ and $\sigma_{12} = 0.5$, the GMM-2 estimator of $\lambda_{11,0}$ reduces the SDs of the 2SLS, 3SLS, GMM-1 and QML estimators by, respectively, 22.5%, 21.6%, 5.5% and 13.8%; and the GMM-2 estimator of $\lambda_{21,0}$ reduces the SDs of the 2SLS, 3SLS, GMM-1 and QML estimators by, respectively, 15.6%, 13.8%, 11.0% and 19.0%. When $\beta_{1,0} = \beta_{2,0} = 0.5$, the IV matrix $\mathbf{Q}$ is less informative and the efficiency improvement of the GMM-2 estimator relative to the other estimators is more prominent. As reported in Table 4, when $n = 500$ and $\sigma_{12} = 0.5$, the GMM-2 estimator of $\lambda_{11,0}$ reduces the SDs of the 2SLS, 3SLS, GMM-1 and QML estimators by, respectively, 35.4%, 35.0%, 12.0% and 67.4%; and the GMM-2 estimator of $\lambda_{21,0}$ reduces the SDs of the 2SLS, 3SLS, GMM-1 and QML estimators by, respectively, 24.9%, 24.9%, 17.3% and 69.0%. In this case, the GMM-2 estimators for $\gamma_{21,0}$ and $\beta_{1,0}$ also reduce the SDs of the QML estimators by, respectively, 9.6% and 7.8%.

- The computational cost of the GMM estimator is much lower than that of the QML estimator. For example, when $n = 250$, $\beta_{1,0} = \beta_{2,0} = 1$, $\sigma_{12} = 0.5$ and disturbances are homoskedastic, the average computation time for GMM-1, GMM-2 and QML estimators are, respectively, 0.06, 0.10 and 9.38 seconds.[9]

6 Summary

In this paper, we propose a general GMM framework for the estimation of SAR models in a system of simultaneous equations with unknown heteroskedasticity. We introduce a new set of quadratic moment conditions to construct the GMM estimator based on the correlation structure of model disturbances within and across equations. We establish the consistency and asymptotic normality of the proposed the GMM estimator and discuss the optimal choice of moment conditions. We also provide heteroskedasticity-robust estimators for the optimal GMM weighting matrix and the asymptotic variance-covariance matrix of the GMM estimator.

The Monte Carlo experiments show that the proposed GMM estimator perform well in finite samples. In particular, the optimal GMM estimator with some simple moment functions is robust under heteroskedasticity with no apparent loss in efficiency under homoskedasticity, whereas the

[9]The Monte Carlo experiments are performed on a computer with Intel (R) Xeon (R) CPU X5450 @ 3.00 GHz and 32.0 GB RAM.

QML estimator for the spatial lag coefficient is biased with a large standard deviation in the presence of heteroskedasticity. Furthermore, the computational cost of the GMM approach is drastically smaller than that of the QML.

References

Allers, M. A. and Elhorst, J. P. (2011). A simultaneous equations model of fiscal policy interactions, *Journal of Regional Science* **51**: 271–291.

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.

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

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

Cohen-Cole, E., Liu, X. and Zenou, Y. (2016). Multivariate choices and identification of social interactions. Working paper, University of Colorado Boulder.

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

Hauptneier, S., Mittermaier, F. and Rincke, J. (2012). Fiscal competition over taxes and public inputs, *Regional Science and Urban Economics* **42**: 407–419.

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.

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

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

Liu, X. (2015). GMM identification and estimation of peer effects in a system of simultaneous equations. Working paper, Univeristy of Colorado Boulder.

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem, *The Review of Economic Studies* **60**: 531–542.

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

Schmidt, P. (1976). *Econometrics*, Marcel Dekker, New York.

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.

Yang, K. and Lee, L. F. (2017). Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models, *Journal of Econometrics* **196**: 196–214.

A Lemmas

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

Lemma A.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* $\boldsymbol{\Sigma}_{kl} = \mathrm{E}(\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_l')$ *for* $k, l = 1, 2, 3, 4$. *Then,*

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

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 A.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 A.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 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 A.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 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 A.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 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 (2017). ■

Lemma A.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 A.7 *Let* $\mathbf{f}(\boldsymbol{\theta}) = [\mathbf{f}_1(\boldsymbol{\theta})', \mathbf{f}_2(\boldsymbol{\theta})']'$ *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). ■

B Proofs

Proof of Proposition 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 (3),

$$\begin{aligned}
\mathbf{u}_k(\boldsymbol{\theta}_k) &= \mathbf{y}_k - \mathbf{Y}_k\boldsymbol{\gamma}_k - \bar{\mathbf{Y}}_k\boldsymbol{\lambda}_k - \mathbf{X}_k\boldsymbol{\beta}_k \\
&= \mathbf{Y}_k(\boldsymbol{\gamma}_{k,0} - \boldsymbol{\gamma}_k) + \bar{\mathbf{Y}}_k(\boldsymbol{\lambda}_{k,0} - \boldsymbol{\lambda}_k) + \mathbf{X}_k(\boldsymbol{\beta}_{k,0} - \boldsymbol{\beta}_k) + \mathbf{u}_k \\
&= \mathbf{Y}(\bar{\boldsymbol{\gamma}}_{k,0} - \bar{\boldsymbol{\gamma}}_k) + \mathbf{W}\mathbf{Y}(\bar{\boldsymbol{\lambda}}_{k,0} - \bar{\boldsymbol{\lambda}}_k) + \mathbf{X}_k(\boldsymbol{\beta}_{k,0} - \boldsymbol{\beta}_k) + \mathbf{u}_k \\
&= \mathbf{d}_k(\boldsymbol{\theta}_k) + \mathbf{u}_k + \sum_{l=1}^{m}[(\bar{\gamma}_{lk,0} - \bar{\gamma}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{I}_n) + (\bar{\lambda}_{lk,0} - \bar{\lambda}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{W})]\mathbf{S}^{-1}\mathbf{u}
\end{aligned} \tag{33}$$

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 $\mathbf{S} = \mathbf{I}_{mn} - \boldsymbol{\Gamma}'_0 \otimes \mathbf{I}_n - \boldsymbol{\Lambda}'_0 \otimes \mathbf{W}$. 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 A.3 and A.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 A.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 (3), 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 A.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 (33), it follows by Lemmas A.3 and A.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: 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 (33), 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 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 A.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 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 1, $n^{-1}\mathbf{g}(\boldsymbol{\theta})'\mathbf{\Omega}^{-1}\mathbf{g}(\boldsymbol{\theta}) - n^{-1}\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})'\mathbf{\Omega}^{-1}\mathbf{g}(\boldsymbol{\theta})] = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta}$. Hence, it suffices to show that $n^{-1}\mathbf{g}(\boldsymbol{\theta})'(\widetilde{\mathbf{\Omega}}^{-1} - \mathbf{\Omega}^{-1})\mathbf{g}(\boldsymbol{\theta}) = \mathrm{o_p}(1)$ uniformly in $\boldsymbol{\theta}$. Let $\|\cdot\|$ denote the Euclidian norm for vectors and matrices. Then,

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

From the proof of Proposition 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 $\mathbf{G}_k(\boldsymbol{\theta}_k) = \sum_{l=1}^{m}[(\bar{\gamma}_{lk,0} - \tilde{\bar{\gamma}}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{I}_n) + (\bar{\lambda}_{lk,0} - \tilde{\bar{\lambda}}_{lk})(\mathbf{i}'_{m,l} \otimes \mathbf{W})]\mathbf{S}^{-1}$, it follows that $n^{-1}\|\mathbf{g}(\boldsymbol{\theta})\| = \mathrm{O}_\mathrm{p}(1)$ uniformly in $\boldsymbol{\theta}$. Therefore, $n^{-1}\mathbf{g}(\boldsymbol{\theta})'(\widetilde{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\mathbf{g}(\boldsymbol{\theta}) = \mathrm{o}_\mathrm{p}(1)$ 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 $\overline{\boldsymbol{\theta}}$,

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

where the asymptotic distribution statement is implied by Lemma A.5. ■

Proof of Proposition 4: To show that the moment functions $\mathbf{g}^*(\boldsymbol{\theta})$ with $\mathbf{Q}^*$ and $\boldsymbol{\Xi}_r^*$'s given in (32) are the most efficient, it suffices to show that any additional moment conditions $\mathbf{g}(\boldsymbol{\theta})$ in the form of (10) are redundant. By Lemma A.7, it suffices to find a matrix $\mathbf{A}$ such that $-\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta}_0)] = \mathrm{E}\left[\mathbf{g}(\boldsymbol{\theta}_0)\mathbf{g}^*(\boldsymbol{\theta}_0)'\right]\mathbf{A}$ and $-\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}^*(\boldsymbol{\theta}_0)] = \mathrm{E}\left[\mathbf{g}^*(\boldsymbol{\theta}_0)\mathbf{g}^*(\boldsymbol{\theta}_0)'\right]\mathbf{A}$. Let $\mathbf{J}_1$ and $\mathbf{J}_2$ be selection matrices such that $\mathbf{X}_1 = \mathbf{X}\mathbf{J}_1$ and $\mathbf{X}_2 = \mathbf{X}\mathbf{J}_2$. For any linear moment function

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

it follows from (29) and (30) that

$$-\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_1(\boldsymbol{\theta}_0)] = \begin{bmatrix} \mathbf{Q}'[\mathrm{E}(\mathbf{y}_2), \mathrm{E}(\mathbf{W}\mathbf{y}_1), \mathrm{E}(\mathbf{W}\mathbf{y}_2), \mathbf{X}\mathbf{J}_1] & \\ & \mathbf{Q}'[\mathrm{E}(\mathbf{y}_1), \mathrm{E}(\mathbf{W}\mathbf{y}_1), \mathrm{E}(\mathbf{W}\mathbf{y}_2), \mathbf{X}\mathbf{J}_2] \end{bmatrix}$$

where

$$\begin{aligned}
\mathrm{E}(\mathbf{y}_1) &= (\mathbf{S}^{-1})_{11}\mathbf{X}_1\boldsymbol{\beta}_{1,0} + (\mathbf{S}^{-1})_{12}\mathbf{X}_2\boldsymbol{\beta}_{2,0} \\
\mathrm{E}(\mathbf{y}_2) &= (\mathbf{S}^{-1})_{21}\mathbf{X}_1\boldsymbol{\beta}_{1,0} + (\mathbf{S}^{-1})_{22}\mathbf{X}_2\boldsymbol{\beta}_{2,0} \\
\mathrm{E}(\mathbf{W}\mathbf{y}_1) &= \mathbf{G}_{11}\mathbf{X}_1\boldsymbol{\beta}_{1,0} + \mathbf{G}_{12}\mathbf{X}_2\boldsymbol{\beta}_{2,0} \\
\mathrm{E}(\mathbf{W}\mathbf{y}_2) &= \mathbf{G}_{21}\mathbf{X}_1\boldsymbol{\beta}_{1,0} + \mathbf{G}_{22}\mathbf{X}_2\boldsymbol{\beta}_{2,0}.
\end{aligned}$$

On the other hand,

$$\mathrm{E}\left[\mathbf{g}_1(\boldsymbol{\theta}_0)\mathbf{g}_1^*(\boldsymbol{\theta}_0)'\right] = \begin{bmatrix} \sigma_{11}\mathbf{Q}'\mathbf{Q}^* & \sigma_{12}\mathbf{Q}'\mathbf{Q}^* \\ \sigma_{12}\mathbf{Q}'\mathbf{Q}^* & \sigma_{22}\mathbf{Q}'\mathbf{Q}^* \end{bmatrix}$$

Hence, $-\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}_1(\boldsymbol{\theta}_0)] = \mathrm{E}\left[\mathbf{g}_1(\boldsymbol{\theta}_0)\mathbf{g}_1^*(\boldsymbol{\theta}_0)'\right]\mathbf{A}_1$, where

$$\mathbf{A}_1 = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\begin{bmatrix} \sigma_{22}\boldsymbol{\Phi}_1 & -\sigma_{12}\boldsymbol{\Phi}_2 \\ -\sigma_{12}\boldsymbol{\Phi}_1 & \sigma_{11}\boldsymbol{\Phi}_2 \end{bmatrix}$$

with

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\beta}_{1,0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\beta}_{2,0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Phi}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_2 \\ \boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}_{1,0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}_{2,0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\beta}_{1,0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\beta}_{2,0} & \mathbf{0} \end{bmatrix}.$$

Next, for any quadratic moment function

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

it follows from (29) and (31) that $-\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_2(\boldsymbol{\theta}_0)] = [\mathbf{D}_{2,1}, \mathbf{D}_{2,2}]$, where

$$\mathbf{D}_{2,1} = \begin{bmatrix} \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_3^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_4^*) & \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_5^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_6^*) & \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_7^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_8^*) & \mathbf{0} \\ \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_3^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_4^*) & \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_5^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_6^*) & \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_7^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_8^*) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$
$$+ \begin{bmatrix} \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_3^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_4^*) & \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_5^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_6^*) & \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_7^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_8^*) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_3^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_4^*) & \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_5^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_6^*) & \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_7^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_8^*) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

and

$$\mathbf{D}_{2,2} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_1^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_2^*) & \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_5^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_6^*) & \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_7^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_8^*) & \mathbf{0} \\ \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_1^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_2^*) & \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_5^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_6^*) & \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_7^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_8^*) & \mathbf{0} \end{bmatrix}$$
$$+ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_1^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_2^*) & \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_5^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_6^*) & \sigma_{11}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_7^*) + \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_8^*) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_1^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_2^*) & \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_5^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_6^*) & \sigma_{12}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_7^*) + \sigma_{22}\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_8^*) & \mathbf{0} \end{bmatrix}$$

Furthermore,

$$\mathrm{E}[\mathbf{g}_2(\boldsymbol{\theta}_0)\mathbf{g}_2^*(\boldsymbol{\theta}_0)'] = \boldsymbol{\Sigma}_1 \otimes [\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_1^*), \cdots, \mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}_8^*)] + \boldsymbol{\Sigma}_2 \otimes [\mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_1^*), \cdots, \mathrm{tr}(\boldsymbol{\Xi}'\boldsymbol{\Xi}_8^*)],$$

where

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{22}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & \sigma_{22}\sigma_{12} \\ \sigma_{12}^2 & \sigma_{22}\sigma_{12} & \sigma_{22}\sigma_{12} & \sigma_{22}^2 \end{bmatrix} \text{ and } \boldsymbol{\Sigma}_2 = \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & \sigma_{22}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{22}\sigma_{12} \\ \sigma_{12}^2 & \sigma_{22}\sigma_{12} & \sigma_{22}\sigma_{12} & \sigma_{22}^2 \end{bmatrix}.$$

Hence, $-\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}_2(\boldsymbol{\theta}_0)] = \mathrm{E}\left[\mathbf{g}_2(\boldsymbol{\theta}_0)\mathbf{g}_2^*(\boldsymbol{\theta}_0)'\right]\mathbf{A}_2$, where

$$\begin{aligned} \mathbf{A}_2 &= [\aleph_1 \otimes \mathbf{i}_{8,3}, \aleph_1 \otimes \mathbf{i}_{8,5}, \aleph_1 \otimes \mathbf{i}_{8,7}, \mathbf{0}, \aleph_3 \otimes \mathbf{i}_{8,1}, \aleph_3 \otimes \mathbf{i}_{8,5}, \aleph_3 \otimes \mathbf{i}_{8,7}, \mathbf{0}] \\ &\quad + [\aleph_2 \otimes \mathbf{i}_{8,4}, \aleph_2 \otimes \mathbf{i}_{8,6}, \aleph_2 \otimes \mathbf{i}_{8,8}, \mathbf{0}, \aleph_4 \otimes \mathbf{i}_{8,2}, \aleph_4 \otimes \mathbf{i}_{8,6}, \aleph_4 \otimes \mathbf{i}_{8,8}, \mathbf{0}], \end{aligned}$$

with

$$\aleph_1 = \boldsymbol{\Sigma}_1^{-1}\begin{bmatrix} \sigma_{11} \\ 0 \\ \sigma_{12} \\ 0 \end{bmatrix} = \boldsymbol{\Sigma}_2^{-1}\begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ 0 \\ 0 \end{bmatrix}$$

$$\aleph_2 = \boldsymbol{\Sigma}_1^{-1}\begin{bmatrix} \sigma_{12} \\ 0 \\ \sigma_{22} \\ 0 \end{bmatrix} = \boldsymbol{\Sigma}_2^{-1}\begin{bmatrix} \sigma_{12} \\ \sigma_{22} \\ 0 \\ 0 \end{bmatrix}$$

$$\aleph_3 = \boldsymbol{\Sigma}_1^{-1}\begin{bmatrix} 0 \\ \sigma_{11} \\ 0 \\ \sigma_{12} \end{bmatrix} = \boldsymbol{\Sigma}_2^{-1}\begin{bmatrix} 0 \\ 0 \\ \sigma_{11} \\ \sigma_{12} \end{bmatrix}$$

and

$$\aleph_4 = \boldsymbol{\Sigma}_1^{-1}\begin{bmatrix} 0 \\ \sigma_{12} \\ 0 \\ \sigma_{22} \end{bmatrix} = \boldsymbol{\Sigma}_2^{-1}\begin{bmatrix} 0 \\ 0 \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}.$$

In summary, the desired result follows since $-\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\boldsymbol{\theta}_0)] = \mathrm{E}\left[\mathbf{g}(\boldsymbol{\theta}_0)\mathbf{g}^*(\boldsymbol{\theta}_0)'\right]\mathbf{A}$ and $-\mathrm{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}^*(\boldsymbol{\theta}_0)] = \mathrm{E}\left[\mathbf{g}^*(\boldsymbol{\theta}_0)\mathbf{g}^*(\boldsymbol{\theta}_0)'\right]\mathbf{A}$ for $\mathbf{A} = [\mathbf{A}_1', \mathbf{A}_2']'$. ■

Table 1: Estimation under Homoskedasticity ($\beta_{1,0} = \beta_{2,0} = 1$)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.4$	$\lambda_{21,0} = 0.2$	$\beta_{1,0} = 1$
$n = 250$				
$\sigma_{12} = 0.3$				
2SLS	0.203(0.093)[0.093]	0.391(0.153)[0.154]	0.193(0.166)[0.166]	0.995(0.094)[0.094]
3SLS	0.199(0.097)[0.097]	0.392(0.153)[0.153]	0.197(0.168)[0.168]	0.996(0.096)[0.096]
GMM-1	0.205(0.092)[0.092]	0.405(0.113)[0.114]	0.194(0.161)[0.161]	0.994(0.093)[0.093]
GMM-2	0.199(0.095)[0.095]	0.408(0.097)[0.098]	0.195(0.128)[0.128]	0.997(0.093)[0.094]
QML	0.194(0.092)[0.092]	0.404(0.096)[0.096]	0.198(0.124)[0.124]	1.003(0.093)[0.093]
$\sigma_{12} = 0.5$				
2SLS	0.204(0.093)[0.093]	0.393(0.154)[0.154]	0.187(0.167)[0.167]	0.995(0.094)[0.094]
3SLS	0.197(0.097)[0.097]	0.395(0.154)[0.154]	0.194(0.170)[0.170]	0.998(0.095)[0.095]
GMM-1	0.208(0.091)[0.092]	0.405(0.125)[0.125]	0.191(0.161)[0.161]	0.994(0.093)[0.093]
GMM-2	0.196(0.095)[0.095]	0.409(0.109)[0.110]	0.195(0.135)[0.135]	0.998(0.093)[0.093]
QML	0.193(0.093)[0.093]	0.405(0.107)[0.107]	0.198(0.129)[0.129]	1.003(0.093)[0.093]
$\sigma_{12} = 0.7$				
2SLS	0.205(0.093)[0.093]	0.395(0.154)[0.154]	0.183(0.168)[0.169]	0.995(0.094)[0.094]
3SLS	0.194(0.098)[0.098]	0.397(0.155)[0.155]	0.191(0.173)[0.173]	1.000(0.096)[0.096]
GMM-1	0.212(0.090)[0.091]	0.405(0.133)[0.133]	0.187(0.157)[0.157]	0.994(0.093)[0.093]
GMM-2	0.194(0.095)[0.096]	0.411(0.123)[0.123]	0.196(0.143)[0.143]	0.999(0.094)[0.094]
QML	0.191(0.093)[0.094]	0.406(0.118)[0.118]	0.197(0.135)[0.135]	1.002(0.094)[0.094]
$n = 500$				
$\sigma_{12} = 0.3$				
2SLS	0.204(0.063)[0.063]	0.390(0.111)[0.112]	0.199(0.115)[0.115]	0.998(0.065)[0.065]
3SLS	0.201(0.064)[0.064]	0.391(0.112)[0.112]	0.201(0.116)[0.116]	0.999(0.065)[0.065]
GMM-1	0.205(0.063)[0.063]	0.398(0.079)[0.079]	0.198(0.111)[0.111]	0.997(0.064)[0.064]
GMM-2	0.202(0.064)[0.064]	0.399(0.070)[0.070]	0.199(0.090)[0.090]	0.999(0.064)[0.064]
QML	0.200(0.062)[0.062]	0.397(0.067)[0.067]	0.199(0.084)[0.084]	1.002(0.063)[0.063]
$\sigma_{12} = 0.5$				
2SLS	0.204(0.063)[0.063]	0.392(0.112)[0.112]	0.196(0.116)[0.116]	0.998(0.065)[0.065]
3SLS	0.200(0.065)[0.065]	0.392(0.113)[0.113]	0.199(0.117)[0.117]	1.000(0.065)[0.065]
GMM-1	0.206(0.062)[0.063]	0.398(0.088)[0.088]	0.197(0.113)[0.113]	0.997(0.064)[0.064]
GMM-2	0.200(0.064)[0.064]	0.400(0.078)[0.078]	0.199(0.095)[0.095]	1.000(0.063)[0.063]
QML	0.200(0.062)[0.062]	0.398(0.074)[0.074]	0.199(0.088)[0.088]	1.002(0.063)[0.063]
$\sigma_{12} = 0.7$				
2SLS	0.205(0.063)[0.063]	0.392(0.112)[0.113]	0.194(0.116)[0.116]	0.998(0.064)[0.064]
3SLS	0.199(0.065)[0.065]	0.393(0.113)[0.114]	0.198(0.118)[0.118]	1.000(0.065)[0.065]
GMM-1	0.208(0.062)[0.062]	0.397(0.095)[0.095]	0.196(0.111)[0.111]	0.998(0.063)[0.064]
GMM-2	0.199(0.064)[0.064]	0.400(0.087)[0.087]	0.200(0.100)[0.100]	1.000(0.063)[0.063]
QML	0.199(0.062)[0.062]	0.398(0.082)[0.082]	0.199(0.093)[0.093]	1.002(0.063)[0.063]

Mean(SD)[RMSE]

Table 2: Estimation under Heteroskedasticity ($\beta_{1,0} = \beta_{2,0} = 1$)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.4$	$\lambda_{21,0} = 0.2$	$\beta_{1,0} = 1$
$n = 250$				
$\sigma_{12} = 0.3$				
2SLS	0.200(0.096)[0.096]	0.398(0.123)[0.123]	0.194(0.135)[0.135]	0.996(0.097)[0.097]
3SLS	0.199(0.093)[0.093]	0.398(0.122)[0.122]	0.195(0.132)[0.132]	0.998(0.093)[0.093]
GMM-1	0.206(0.087)[0.088]	0.405(0.096)[0.096]	0.191(0.129)[0.130]	0.995(0.090)[0.090]
GMM-2	0.199(0.090)[0.090]	0.406(0.088)[0.088]	0.196(0.110)[0.110]	0.996(0.090)[0.090]
QML	0.203(0.094)[0.094]	0.366(0.110)[0.116]	0.185(0.154)[0.154]	1.013(0.095)[0.096]
$\sigma_{12} = 0.5$				
2SLS	0.200(0.096)[0.096]	0.399(0.124)[0.124]	0.191(0.136)[0.136]	0.997(0.096)[0.097]
3SLS	0.197(0.093)[0.094]	0.399(0.123)[0.123]	0.194(0.133)[0.133]	0.999(0.093)[0.093]
GMM-1	0.208(0.086)[0.086]	0.405(0.103)[0.103]	0.189(0.127)[0.128]	0.995(0.089)[0.089]
GMM-2	0.196(0.090)[0.090]	0.408(0.095)[0.095]	0.196(0.113)[0.113]	0.998(0.090)[0.090]
QML	0.202(0.094)[0.094]	0.368(0.118)[0.122]	0.182(0.148)[0.149]	1.013(0.095)[0.096]
$\sigma_{12} = 0.7$				
2SLS	0.200(0.096)[0.096]	0.400(0.124)[0.124]	0.188(0.137)[0.138]	0.997(0.097)[0.097]
3SLS	0.194(0.094)[0.095]	0.401(0.124)[0.124]	0.193(0.135)[0.135]	1.001(0.093)[0.093]
GMM-1	0.210(0.085)[0.086]	0.404(0.108)[0.108]	0.188(0.124)[0.124]	0.995(0.089)[0.089]
GMM-2	0.193(0.091)[0.091]	0.409(0.103)[0.103]	0.197(0.117)[0.117]	1.000(0.090)[0.090]
QML	0.201(0.094)[0.094]	0.370(0.122)[0.125]	0.179(0.140)[0.141]	1.013(0.096)[0.097]
$n = 500$				
$\sigma_{12} = 0.3$				
2SLS	0.202(0.064)[0.064]	0.396(0.089)[0.089]	0.197(0.096)[0.096]	1.000(0.066)[0.066]
3SLS	0.200(0.061)[0.061]	0.398(0.088)[0.088]	0.198(0.093)[0.094]	0.999(0.063)[0.063]
GMM-1	0.204(0.059)[0.059]	0.399(0.068)[0.068]	0.197(0.091)[0.092]	0.998(0.061)[0.061]
GMM-2	0.201(0.060)[0.060]	0.400(0.063)[0.063]	0.199(0.078)[0.078]	0.999(0.061)[0.061]
QML	0.210(0.062)[0.063]	0.358(0.075)[0.086]	0.187(0.104)[0.104]	1.013(0.064)[0.065]
$\sigma_{12} = 0.5$				
2SLS	0.202(0.064)[0.064]	0.397(0.089)[0.089]	0.196(0.096)[0.096]	1.000(0.066)[0.066]
3SLS	0.199(0.061)[0.061]	0.399(0.088)[0.089]	0.198(0.094)[0.094]	1.000(0.063)[0.063]
GMM-1	0.205(0.059)[0.059]	0.399(0.073)[0.073]	0.196(0.091)[0.091]	0.998(0.061)[0.061]
GMM-2	0.200(0.060)[0.060]	0.401(0.069)[0.069]	0.199(0.081)[0.081]	0.999(0.061)[0.061]
QML	0.209(0.062)[0.063]	0.359(0.080)[0.090]	0.185(0.100)[0.102]	1.013(0.064)[0.065]
$\sigma_{12} = 0.7$				
2SLS	0.202(0.064)[0.064]	0.397(0.089)[0.089]	0.195(0.096)[0.096]	1.000(0.066)[0.066]
3SLS	0.198(0.061)[0.061]	0.399(0.089)[0.089]	0.197(0.094)[0.094]	1.000(0.063)[0.063]
GMM-1	0.206(0.058)[0.058]	0.399(0.077)[0.077]	0.195(0.089)[0.090]	0.998(0.061)[0.061]
GMM-2	0.198(0.060)[0.060]	0.402(0.074)[0.074]	0.199(0.084)[0.084]	1.000(0.061)[0.061]
QML	0.209(0.062)[0.062]	0.362(0.083)[0.091]	0.181(0.095)[0.097]	1.013(0.064)[0.065]

Mean(SD)[RMSE]

Table 3: Estimation under Homoskedasticity ($\beta_{1,0} = \beta_{2,0} = 0.5$)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.4$	$\lambda_{21,0} = 0.2$	$\beta_{1,0} = 0.5$
$n = 250$				
$\sigma_{12} = 0.3$				
2SLS	0.221(0.185)[0.186]	0.366(0.315)[0.317]	0.197(0.346)[0.346]	0.489(0.095)[0.095]
3SLS	0.207(0.207)[0.208]	0.370(0.322)[0.323]	0.204(0.357)[0.357]	0.488(0.099)[0.099]
GMM-1	0.227(0.177)[0.179]	0.405(0.186)[0.186]	0.192(0.299)[0.299]	0.492(0.092)[0.093]
GMM-2	0.216(0.203)[0.203]	0.413(0.150)[0.151]	0.180(0.217)[0.218]	0.495(0.095)[0.095]
QML	0.182(0.190)[0.191]	0.411(0.153)[0.153]	0.203(0.216)[0.216]	0.504(0.093)[0.093]
$\sigma_{12} = 0.5$				
2SLS	0.225(0.186)[0.188]	0.368(0.334)[0.336]	0.184(0.352)[0.352]	0.489(0.096)[0.097]
3SLS	0.199(0.211)[0.211]	0.376(0.328)[0.329]	0.200(0.374)[0.374]	0.492(0.099)[0.099]
GMM-1	0.242(0.173)[0.178]	0.399(0.219)[0.219]	0.182(0.306)[0.307]	0.491(0.091)[0.091]
GMM-2	0.208(0.204)[0.204]	0.412(0.181)[0.181]	0.185(0.242)[0.243]	0.499(0.093)[0.093]
QML	0.173(0.197)[0.199]	0.412(0.180)[0.180]	0.207(0.232)[0.232]	0.505(0.094)[0.094]
$\sigma_{12} = 0.7$				
2SLS	0.232(0.187)[0.190]	0.381(0.323)[0.324]	0.180(0.365)[0.365]	0.490(0.091)[0.092]
3SLS	0.194(0.215)[0.215]	0.394(0.331)[0.331]	0.199(0.391)[0.391]	0.499(0.094)[0.094]
GMM-1	0.258(0.166)[0.176]	0.393(0.238)[0.238]	0.172(0.296)[0.298]	0.490(0.087)[0.087]
GMM-2	0.206(0.202)[0.203]	0.415(0.213)[0.213]	0.185(0.258)[0.259]	0.503(0.089)[0.089]
QML	0.167(0.204)[0.207]	0.416(0.217)[0.217]	0.208(0.252)[0.252]	0.505(0.094)[0.095]
$n = 500$				
$\sigma_{12} = 0.3$				
2SLS	0.215(0.127)[0.128]	0.367(0.215)[0.218]	0.199(0.230)[0.230]	0.496(0.063)[0.063]
3SLS	0.207(0.134)[0.134]	0.370(0.217)[0.219]	0.204(0.235)[0.235]	0.497(0.064)[0.064]
GMM-1	0.221(0.123)[0.125]	0.391(0.129)[0.129]	0.201(0.204)[0.204]	0.497(0.061)[0.061]
GMM-2	0.213(0.132)[0.133]	0.399(0.106)[0.106]	0.194(0.154)[0.154]	0.501(0.062)[0.062]
QML	0.199(0.127)[0.127]	0.398(0.097)[0.097]	0.201(0.138)[0.138]	0.504(0.061)[0.061]
$\sigma_{12} = 0.5$				
2SLS	0.217(0.125)[0.126]	0.378(0.212)[0.213]	0.190(0.234)[0.234]	0.496(0.065)[0.065]
3SLS	0.201(0.134)[0.134]	0.383(0.215)[0.216]	0.198(0.243)[0.243]	0.498(0.065)[0.065]
GMM-1	0.228(0.120)[0.123]	0.392(0.151)[0.151]	0.191(0.210)[0.210]	0.497(0.063)[0.063]
GMM-2	0.209(0.132)[0.132]	0.404(0.128)[0.128]	0.191(0.168)[0.168]	0.502(0.063)[0.063]
QML	0.196(0.128)[0.128]	0.402(0.116)[0.116]	0.199(0.151)[0.151]	0.503(0.063)[0.063]
$\sigma_{12} = 0.7$				
2SLS	0.217(0.125)[0.126]	0.375(0.215)[0.217]	0.193(0.236)[0.236]	0.495(0.064)[0.064]
3SLS	0.197(0.133)[0.133]	0.381(0.220)[0.220]	0.205(0.249)[0.249]	0.500(0.064)[0.064]
GMM-1	0.233(0.116)[0.121]	0.385(0.167)[0.168]	0.192(0.206)[0.206]	0.496(0.061)[0.061]
GMM-2	0.203(0.129)[0.129]	0.398(0.152)[0.152]	0.202(0.184)[0.184]	0.502(0.061)[0.061]
QML	0.191(0.128)[0.129]	0.399(0.136)[0.136]	0.205(0.163)[0.163]	0.502(0.063)[0.063]

Mean(SD)[RMSE]

Table 4: Estimation under Heteroskedasticity ($\beta_{1,0} = \beta_{2,0} = 0.5$)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.4$	$\lambda_{21,0} = 0.2$	$\beta_{1,0} = 0.5$
$n = 250$				
$\sigma_{12} = 0.3$				
2SLS	0.209(0.194)[0.194]	0.379(0.253)[0.254]	0.196(0.280)[0.280]	0.494(0.098)[0.098]
3SLS	0.203(0.206)[0.206]	0.379(0.253)[0.254]	0.197(0.281)[0.281]	0.492(0.095)[0.095]
GMM-1	0.228(0.172)[0.174]	0.400(0.162)[0.162]	0.190(0.252)[0.253]	0.494(0.089)[0.089]
GMM-2	0.211(0.199)[0.199]	0.407(0.140)[0.140]	0.187(0.200)[0.201]	0.499(0.090)[0.090]
QML	0.183(0.210)[0.211]	0.410(0.400)[0.400]	0.180(0.580)[0.580]	0.513(0.099)[0.100]
$\sigma_{12} = 0.5$				
2SLS	0.213(0.192)[0.193]	0.382(0.256)[0.256]	0.187(0.282)[0.282]	0.493(0.097)[0.097]
3SLS	0.197(0.210)[0.210]	0.386(0.254)[0.254]	0.193(0.290)[0.290]	0.496(0.094)[0.094]
GMM-1	0.240(0.169)[0.174]	0.395(0.184)[0.184]	0.183(0.254)[0.254]	0.493(0.087)[0.087]
GMM-2	0.204(0.201)[0.201]	0.409(0.161)[0.161]	0.189(0.210)[0.210]	0.501(0.086)[0.086]
QML	0.178(0.210)[0.212]	0.405(0.470)[0.470]	0.178(0.602)[0.603]	0.513(0.099)[0.100]
$\sigma_{12} = 0.7$				
2SLS	0.211(0.195)[0.195]	0.386(0.251)[0.252]	0.186(0.282)[0.282]	0.492(0.097)[0.097]
3SLS	0.186(0.213)[0.214]	0.394(0.250)[0.250]	0.198(0.296)[0.296]	0.500(0.093)[0.093]
GMM-1	0.249(0.165)[0.172]	0.393(0.196)[0.197]	0.175(0.240)[0.241]	0.492(0.086)[0.086]
GMM-2	0.195(0.200)[0.200]	0.411(0.184)[0.184]	0.195(0.221)[0.221]	0.504(0.086)[0.086]
QML	0.169(0.214)[0.216]	0.391(0.469)[0.469]	0.186(0.543)[0.543]	0.512(0.101)[0.102]
$n = 500$				
$\sigma_{12} = 0.3$				
2SLS	0.209(0.127)[0.128]	0.386(0.176)[0.177]	0.189(0.194)[0.194]	0.500(0.065)[0.065]
3SLS	0.202(0.126)[0.126]	0.391(0.175)[0.175]	0.192(0.193)[0.193]	0.499(0.062)[0.062]
GMM-1	0.218(0.115)[0.116]	0.398(0.118)[0.118]	0.190(0.180)[0.181]	0.499(0.060)[0.060]
GMM-2	0.208(0.122)[0.123]	0.403(0.100)[0.100]	0.190(0.138)[0.139]	0.502(0.059)[0.059]
QML	0.206(0.138)[0.138]	0.397(0.320)[0.320]	0.160(0.488)[0.490]	0.513(0.065)[0.066]
$\sigma_{12} = 0.5$				
2SLS	0.210(0.128)[0.128]	0.384(0.181)[0.181]	0.189(0.197)[0.197]	0.499(0.065)[0.065]
3SLS	0.199(0.128)[0.128]	0.389(0.180)[0.180]	0.195(0.197)[0.197]	0.500(0.063)[0.063]
GMM-1	0.224(0.113)[0.116]	0.393(0.133)[0.134]	0.189(0.179)[0.180]	0.498(0.059)[0.059]
GMM-2	0.205(0.123)[0.123]	0.401(0.117)[0.117]	0.194(0.148)[0.148]	0.503(0.059)[0.059]
QML	0.203(0.136)[0.136]	0.377(0.359)[0.360]	0.178(0.477)[0.478]	0.512(0.064)[0.065]
$\sigma_{12} = 0.7$				
2SLS	0.209(0.128)[0.128]	0.383(0.179)[0.180]	0.192(0.196)[0.196]	0.498(0.066)[0.066]
3SLS	0.194(0.130)[0.130]	0.390(0.178)[0.178]	0.199(0.198)[0.198]	0.500(0.063)[0.063]
GMM-1	0.229(0.111)[0.114]	0.390(0.141)[0.141]	0.187(0.171)[0.171]	0.497(0.059)[0.059]
GMM-2	0.200(0.123)[0.123]	0.400(0.132)[0.132]	0.199(0.157)[0.157]	0.502(0.060)[0.060]
QML	0.201(0.132)[0.132]	0.355(0.318)[0.321]	0.193(0.381)[0.381]	0.511(0.065)[0.066]

Mean(SD)[RMSE]