GMM Identification and Estimation of Peer Effects
in a System of Simultaneous Equations

Xiaodong Liu*
Department of Economics, University of Colorado, Boulder, CO 80309, USA
August, 2016

Abstract

This paper considers the identification and estimation of network models with agents interacting in multiple activities. We establish the model identification using both linear and quadratic moment conditions. The quadratic moment conditions exploit the covariance structure of individuals’ choices in the same and related activities, and facilitate the identification of peer effects when exclusion restrictions based on intransitivity of network connections or variation in network sizes are not available. Combining linear and quadratic moment conditions, we propose a general GMM framework for the estimation of simultaneous equations network models. The GMM estimator improves the estimation efficiency of the existing IV-based linear estimators in the literature. Simulation experiments show that the GMM estimator performs well in finite samples.

JEL classification: C31, C36

Key words: social networks, quadratic moment conditions, efficiency, many-instrument bias

*Phone: 1-303-492-7414. Fax: 1-303-492-8960. Email: xiaodong.liu@colorado.edu.
1 Introduction

Tremendous progress has been made in understanding the identification of peer effects since the seminal work by Manski (1993) (see Blume et al., 2011, for a review). However, until recently, little attention has been paid to the modeling and identification of peer effects when economic agents interact in multiple activities. To investigate how peers influence decision-making involving multiple activities, Cohen-Cole et al. (2014) develop a simultaneous equations network model and suggest an IV-based estimation procedure by extending the generalized spatial 2SLS and 3SLS estimators proposed by Kelejian and Prucha (2004). Following the insight by Bramoullé et al. (2009), the identification strategy in Cohen-Cole et al. (2014) exploits exclusion restrictions based on intransitivity of network connections. Liu (2014) considers the identification of the simultaneous equations network model when the adjacency matrix that represents network topology has non-constant row sums. In this case, Liu (2014) shows that the variation in Bonacich centrality (Bonacich, 1987) across network nodes can facilitate the identification of peer effects and proposes to use the Bonacich centrality as an additional IV to improve estimation efficiency. As Cohen-Cole et al. (2014) and Liu (2014) focus on IV-based linear estimators, the corresponding identification strategy only utilizes linear moment conditions of model disturbances.

For single-equation spatial econometric models, quadratic moment conditions based on the covariance structure of model disturbances are often used to achieve identification when the model cannot be identified through linear moment conditions (see, e.g., Kelejian and Prucha, 1999; Lee, 2007b). In this paper, we propose to utilize quadratic moment conditions based on the covariance structure of disturbances both within and across equations for the identification of simultaneous equations network models. The idea of identifying peer effects by the covariance structure of model disturbances traces back to Glaeser et al. (1996) and is later developed to the method of variance contrasts by Graham (2008). In the method of variance contrasts, identification is achieved through the differences in intergroup outcome variances when there are at least two groups with different sizes (Durlauf and Tanaka, 2008). By contrast, the identification strategy in this paper exploits the covariance structure of individuals’ choices in the same and related activities within a group.
Hence, peer effects can be identified through quadratic moment conditions even if there is neither variation in group sizes nor intransitivity in network connections (see Examples 1 and 4).\(^1\)

Combining linear and quadratic moment conditions, we propose a generalized method of moments (GMM) framework for the identification and estimation of simultaneous equations network models. The GMM estimator improves the estimation efficiency of the IV-based linear estimators proposed by Cohen-Cole et al. (2014) and Liu (2014). Compared to the quasi-maximum likelihood estimator proposed by Yang and Lee (2014) for the simultaneous equations spatial autoregressive model, the GMM estimator is computationally simple and can be easily modified to incorporate group fixed effects.\(^2\) Monte Carlo simulations show that the proposed GMM estimator performs well in finite samples. The efficiency improvement through quadratic moment conditions is more significant when the linear moment conditions are less informative.

The rest of the paper is organized as follows. Sections 2 and 3 consider different specifications of the econometric model motivated by the network game described in Appendix A and provide sufficient conditions for the identification of peer effects. Section 4 proposes a general GMM framework for the estimation of the simultaneous equations network model. Monte Carlo evidence on the finite sample performance of the proposed estimator is given in Section 5. Section 6 briefly concludes. The proofs are collected in the Appendix.

Throughout the paper, we adopt the following notation. For an \(n \times n\) matrix \(A = [a_{ij}]\), let \(A^{(s)} = A + A'\), \(\text{vec}_D(A) = (a_{11}, \cdots, a_{nn})'\), and \(\rho(A)\) denote the spectral radius of \(A\). The row (or column) sums of an \(n \times n\) matrix \(A\) are uniformly bounded in absolute value if \(\max_{i=1,\ldots,n} \sum_{j=1}^n |a_{ij}|\) (or \(\max_{j=1,\ldots,n} \sum_{i=1}^n |a_{ij}|\)) is bounded as \(n \to \infty\). For an \(n \times m\) matrix \(A = [a_{ij}]\), the vectorization of \(A\) is denoted by \(\text{vec}(A) = (a_{11}, \cdots, a_{n1}, a_{12}, \cdots, a_{nm})'\), and the Euclidean matrix norm of \(A\) is

\(^1\)In this paper, we focus on the case that model disturbances are i.i.d. across individuals for exposition purpose. For the basic simultaneous equations network model in Section 2, quadratic moment conditions can be constructed with cross-sectionally correlated disturbances in the presence of an unknown form of heteroskedasticity as in Liu and Lee (2010) and Lin and Lee (2010).

\(^2\)When group fixed effects are introduced into the network model, it is usually desirable to eliminate the fixed effect parameters from the objective function of the estimator to avoid the potential incidental parameter problem (Neyman and Scott, 1948). When the adjacency matrix has non-constant row sums, the fixed effect parameters cannot be easily eliminated from the likelihood function, which makes the quasi-maximum likelihood approach less desirable in the presence of group fixed effects.

\(^3\)If \(A, B, C\) are conformable matrices, then \(\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)\), where \(\otimes\) denotes the Kronecker product.
denoted by $|A| = \sqrt{\text{tr}(A^tA)}$. Let $\text{diag}(A_j)$ denote a “generalized” block diagonal matrix with a typical diagonal block being an $n_j \times m_j$ matrix $A_j$. Finally, let $I_n$ denote the $n \times n$ identity matrix and $\mathbf{1}_n$ denote an $n \times 1$ vector of ones.

## 2 Basic Econometric Model

### 2.1 Model and GMM estimator

Suppose $n$ individuals interact in $m$ activities through a network. The topology of the network is captured by an $n \times n$ (weighted) adjacency matrix $W = [w_{ij}]$, where the $(i,j)$-th element $w_{ij}$ is a known nonnegative constant that represents the proximity of individuals $i$ and $j$ in the network. As a normalization, $w_{ii} = 0$ for all $i$. The peers of individual $i$ in the network is the set of individuals $\{j : w_{ij} > 0\}$.

The choices (or outcomes) of the individuals in the $m$ activities are given by a system of $m$ equations:

$$
\mathbf{Y} = \mathbf{Y}\Phi_0 + \bar{\mathbf{Y}}\Lambda_0 + \mathbf{X}\mathbf{B}_0 + \bar{\mathbf{X}}\Gamma_0 + \mathbf{U}. 
$$

In this model, $\mathbf{Y} = [y_1, \cdots, y_m]$ is an $n \times m$ matrix of observations on $m$ endogenous variables, $\mathbf{X} = [x_1, \cdots, x_K]$ is an $n \times K$ matrix of observations on $K$ exogenous variables, and $\mathbf{U} = [u_1, \cdots, u_m]$ is an $n \times m$ matrix of disturbances. $\bar{\mathbf{Y}} = \mathbf{W}\mathbf{Y}$ and $\bar{\mathbf{X}} = \mathbf{W}\mathbf{X}$ represent, respectively, the choices and exogenous characteristics of the peers. $\Phi_0 = [\phi_{lk,0}]$, $\Lambda_0 = [\lambda_{lk,0}]$, $\mathbf{B}_0$ and $\Gamma_0$ are, respectively, $m \times m$, $m \times m$, $K \times m$ and $K \times m$ matrices of true parameters in the data generating process (DGP).\(^4\)

We assume that each row of $\mathbf{U}$ is an i.i.d. random vector with a zero mean and an $m \times m$ covariance matrix $\Sigma = [\sigma_{kl}]$. Thus, the disturbances of the same individual are allowed to be correlated across different activities.

The econometric model can be motivated by the best response function of a multiple-activity network game introduced by Cohen-Cole et al. (2014) (see Appendix A). As discussed in Cohen-Cole et al. (2014), in model (2.1), $\phi_{lk,0}$ represents the *simultaneity effect*, wherein an individual’s

\(^4\)As a normalization, $\phi_{kk,0} = 0$ for all $k$. 

4
choice in a certain activity could be affected by her own choices in related activities; \( \lambda_{kk,0} \) represents the *endogenous peer effect*, wherein an individual’s choice in a certain activity could be affected by her peers’ choices in that activity; \( \lambda_{lk,0} \) \((l \neq k)\) represents the *cross-activity peer effect*, wherein an individual’s choice in a certain activity could be affected by her peers’ choices in related activities; and \( \Gamma_0 \) represents the *contextual effect*, wherein an individual’s choice could be affected by the exogenous characteristics of her peers. The purpose of this paper is to discuss the identification and estimation of these different effects.

In general, the identification of simultaneous equations models requires exclusion restrictions. Let \( \phi_{k,0}, \lambda_{k,0}, \beta_{k,0}, \) and \( \gamma_{k,0} \) denote vectors of nonzero elements of the \( k \)-th columns of \( \Phi_0, \Lambda_0, \beta_0, \) and \( \Gamma_0 \) respectively under some exclusion restrictions. Let \( Y_k, \tilde{Y}_k, X_k \) and \( \tilde{X}_k \) denote the corresponding matrices containing columns of \( Y, \tilde{Y}, X \) and \( \tilde{X} \) that appear in the \( k \)-th equation. Then, the \( k \)-th equation of model (2.1) is

\[
y_k = Y_k \phi_{k,0} + \tilde{Y}_k \lambda_{k,0} + X_k \beta_{k,0} + \tilde{X}_k \gamma_{k,0} + u_k. \tag{2.2}
\]

Let

\[
u_k(\theta_k) = y_k - Y_k \phi_k - \tilde{Y}_k \lambda_k - X_k \beta_k - \tilde{X}_k \gamma_k,
\]

where \( \theta_k = (\phi_k', \lambda_k', \beta_k', \gamma_k')' \). Inspired by the GMM estimator proposed by Lee (2007b) for the single-equation spatial autoregressive model, we consider both linear and quadratic moment functions of \( u_k(\theta_k) \) to construct the GMM estimator. The linear moment functions are given by

\[
f_{1,k}(\theta_k) = Q' u_k(\theta_k), \quad \text{for } k = 1, \ldots, m.
\]

where \( Q \) is an \( n \times q \) nonstochastic IV matrix. The quadratic moment functions are given by

\[
f_{2,kl}(\theta_k, \theta_l) = \Xi^i [u_k(\theta_k), \ldots, \Xi^j [u_k(\theta_k)]' u_l(\theta_l)], \quad \text{for } k, l = 1, \ldots, m,
\]

where \( \Xi_r \) is an \( n \times n \) nonstochastic matrix with \( \text{tr}(\Xi_r) = 0 \), for \( r = 1, \ldots, p. \)

5In this section, we treat \( q \) as a fixed constant that does not grow with the sample size. A possible candidate for
parameter value, $E[f_{1,k}(\theta_{k,0})] = Q'E(u_k) = 0$ and $E[f_{2,kl}(\theta_{k,0}, \theta_{l,0})] = E[\Xi'u_k, \cdots, \Xi'_p u_k]'u_l] = \sigma_{kl}[tr(\Xi_1), \cdots, tr(\Xi_p)]' = 0$. The above quadratic moment functions generalize those proposed by Lee (2007b) as they utilize the covariance structure of disturbances both within and across equations.

Let $f_1(\theta) = [f_{1,1}(\theta_1)', \cdots, f_{1,m}(\theta_m)']'$, $f_2(\theta) = [f_{2,11}(\theta_1, \theta_1)', \cdots, f_{2,1m}(\theta_1, \theta_m)', f_{2,21}(\theta_2, \theta_1)', \cdots, f_{2,mm}(\theta_m, \theta_m)']'$, and

$$f(\theta) = [f_1(\theta)', f_2(\theta)']'$$

where $\theta = (\theta_1', \cdots, \theta_m')'$. Let $\Omega_f = \text{Var}[f(\theta_0)]$. Suppose $n^{-1}\hat{\Omega}_f$ is a consistent estimator of $n^{-1}\Omega_f$ such that $n^{-1}\hat{\Omega}_f - n^{-1}\Omega_f = o_p(1)$. The GMM estimator for $\theta_0$ is given by $\hat{\theta}_{gmm} = \text{arg min} f(\theta)'\hat{\Omega}_f^{-1}f(\theta)$.

### 2.2 Identification

In the following, we discuss the model identification through the linear and quadratic moment conditions utilized by the GMM estimator under different exclusion restrictions. In general, for $\theta_0$ to be identified, the moment equations $\lim_{n \to \infty} n^{-1}E[f(\theta)] = 0$ need to have a unique solution at $\theta = \theta_0$ (Hansen, 1982).

#### 2.2.1 SUR network model

First, we impose the exclusion restrictions that $\phi_{lk,0} = \lambda_{lk,0} = 0$ for all $k \neq l$. In this case, model (2.2) is a system of seemingly unrelated regressions (SUR) with the $k$-th equation given by

$$y_k = \lambda_{kk,0} y_k + X_\beta_{k,0} + X_\gamma_{k,0} + u_k.$$  

(2.5)

If $|\lambda_{kk,0}| < 1/\rho(W)$, where $\rho(W)$ is the spectral radius of $W$, then $(I_n - \lambda_{kk,0} W)$ is nonsingular, and thus the reduced-form equation of (2.5) is given by

$$y_k = (I_n - \lambda_{kk,0} W)^{-1}(X_\beta_{k,0} + X_\gamma_{k,0}) + \gamma_{k,0} u_k,$$

(2.6)

$Q$ could be $[W^2 X, W^2 X]$ and possible candidates for $\Xi_p$ could be $W$ and $W^2 - n^{-1}\text{tr}(W^2)I_n$. 6
which implies

$$\tilde{y}_k = G_k (X \beta_k + \bar{X} \gamma_k) + G_k u_k,$$

(2.7)

where $G_k = W(I_n - \lambda_{kk,0} W)^{-1}$. From (2.5) and (2.7), we have

$$u_k(\theta_k) = y_k - \lambda_{kk} \tilde{y}_k - X \beta_k - \bar{X} \gamma_k = d_k(\theta_k) + u_k + (\lambda_{kk,0} - \lambda_{kk}) G_k u_k,$$

where

$$d_k(\theta_k) = (\lambda_{kk,0} - \lambda_{kk}) E(\bar{y}_k) + X(\beta_k - \bar{X} \gamma_k),$$

and $E(\bar{y}_k) = G_k (X \beta_{k,0} + \bar{X} \gamma_{k,0}).$

For the linear moment functions, we have

$$\lim_{n \to \infty} n^{-1} E[f_{1,k}(\theta_k)] = \lim_{n \to \infty} n^{-1} E[Q'u_k(\theta_k)] = \lim_{n \to \infty} n^{-1} Q'd_k(\theta_k),$$

for $k = 1, \cdots, m$. Therefore, $\lim_{n \to \infty} n^{-1} E[f_{1,k}(\theta_k)] = 0$ has a unique solution at the true parameter value, if $Q'[E(\bar{y}_k), X, \bar{X}]$ has full column rank for large enough $n$. This sufficient rank condition implies the necessary rank condition that $[E(\bar{y}_k), X, \bar{X}]$ has full column rank and the rank of $Q$ is at least $2K+1$, for large enough $n$. By a similar argument as in Cohen-Cole et al. (2014), $[E(\bar{y}_k), X, \bar{X}]$ has full column rank if $\lambda_{kk,0} \beta_{k,0} + \gamma_{k,0} \neq 0$ and $[X, WX, W^2X]$ has full column rank.\(^6\)

On the other hand, the necessary rank condition for identification does not hold if $E(\bar{y}_k)$ is linearly dependent on $[X, \bar{X}]$ such that $E(\bar{y}_k) = Xc_1 + \bar{X}c_2$, where $c_1, c_2$ are $K \times 1$ constant vectors. In this case,

$$d_k(\theta_k) = X[ (\lambda_{kk,0} - \lambda_{kk}) c_1 + (\beta_{k,0} - \bar{X} \gamma_k) ] + \bar{X}[ (\lambda_{kk,0} - \lambda_{kk}) c_2 + (\gamma_{k,0} - \bar{X} \gamma_k)]$$

\(^6\)Bramoullé et al. (2009) show that, in a single-equation network model in the form of (2.5) with a stochastic (but extremely exogenous) regressor, if $\lambda_{kk,0} \beta_{k,0} + \gamma_{k,0} \neq 0$, endogenous and contextual effects are identified if $I_n, W, W^2$ are linearly independent. This identification condition essentially requires that intransitivity exists in network connections.
and the solutions of $\lim_{n \to \infty} n^{-1} Q'_k d_k(\theta_k) = 0$ are characterized by

$$\beta_k = (\lambda_{kk,0} - \lambda_{kk}) c_1 + \beta_{kk,0}, \quad \text{and} \quad \gamma_k = (\lambda_{kk,0} - \lambda_{kk}) c_2 + \gamma_{kk,0},$$

(2.8)

as long as $Q[X, \bar{X}]$ has full column rank for large enough $n$. Hence, $(\beta'_{kk,0}, \gamma'_{kk,0})$ can be identified if $\lambda_{kk,0}$ can be identified from the quadratic moment equations

$$\lim_{n \to \infty} n^{-1} E[f_{2,kl}(\theta_k, \theta_l)] = 0, \quad \text{for } l = 1, \ldots, m.$$

To wrap up, the following proposition summarizes sufficient identification conditions for the SUR network model.

**Proposition 2.1.** For the SUR network model (2.5), suppose $|\lambda_{kk,0}| < 1/\rho(W)$. Then $\theta_{kk,0} = (\lambda_{kk,0}, \beta'_{kk,0}, \gamma'_{kk,0})'$ is identified if either

(i) $\lim_{n \to \infty} n^{-1} Q[E(\bar{y}_k), X, \bar{X}]$ is finite with full column rank; or

(ii) $\lim_{n \to \infty} n^{-1} Q[X, \bar{X}]$ is finite with full column rank and the equations

$$\sigma_{kl} \lim_{n \to \infty} n^{-1} [(\lambda_{kk,0} - \lambda_{kk}) \text{tr}(\Xi'_k G) + (\lambda_{ll,0} - \lambda_{ll}) \text{tr}(\Xi'_l G_l) + (\lambda_{kk,0} - \lambda_{kk})(\lambda_{ll,0} - \lambda_{ll}) \text{tr}(G'_k \Xi'_l G_l)] = 0,$$

(2.9)

for $r = 1, \cdots, p$ and $l = 1, \ldots, m$, have a unique solution for $\lambda_{kk}$ at the true parameter value.

**Example 1 (Complete Networks).** Suppose there are $n$ individuals that are randomly assigned into $G$ equal-sized groups with $\bar{n} = n/G$ individuals in each group.\(^7\) Suppose each individual is connected with all the others in the same group but with no one from a different group. The corresponding (row-normalized) adjacency matrix is given by $W_C = I_G \otimes [(\bar{n} - 1)^{-1}(t_n^t - t_{\bar{n}}) - I_{\bar{n}}]$. In this case, $[E(\bar{y}_k), X, \bar{X}]$ does not have full column rank as $I_{\bar{n}}, W_C, W^2_C$ are linearly dependent (Bramoullé et al., 2009). Consequently, condition (i) of Proposition 2.1 does not hold and the SUR network model cannot be identified by linear moment conditions alone. However, the model may still be identifiable through additional quadratic moment conditions. To illustrate the idea, suppose

\(^7\)For simplicity, in this example, we assume $\bar{n}$ is a fixed constant that does not increase with the sample size.
\( m = 2 \) and \( \sigma_{kl} \neq 0 \) for \( k, l = 1, 2 \). Then, from (2.9) we have

\[
\lim_{n \to \infty} n^{-1}[(\lambda_{11,0} - \lambda_{11})\text{tr}(\Xi^{(s)}_r G_1) + (\lambda_{11,0} - \lambda_{11})^2\text{tr}(G'_1 \Xi_r G_1)] = 0 \quad (2.10)
\]

\[
\lim_{n \to \infty} n^{-1}[(\lambda_{22,0} - \lambda_{22})\text{tr}(\Xi^{(s)}_r G_2) + (\lambda_{22,0} - \lambda_{22})^2\text{tr}(G'_2 \Xi_r G_2)] = 0 \quad (2.11)
\]

\[
\lim_{n \to \infty} n^{-1}[(\lambda_{11,0} - \lambda_{11})\text{tr}(\Xi'_r G_1) + (\lambda_{22,0} - \lambda_{22})\text{tr}(\Xi_r G_2) + (\lambda_{11,0} - \lambda_{11})(\lambda_{22,0} - \lambda_{22})\text{tr}(G'_1 \Xi_r G_2)] = 0 \quad (2.12)
\]

\[
\lim_{n \to \infty} n^{-1}[(\lambda_{22,0} - \lambda_{22})\text{tr}(\Xi'_r G_2) + (\lambda_{11,0} - \lambda_{11})\text{tr}(\Xi_r G_1) + (\lambda_{11,0} - \lambda_{11})(\lambda_{22,0} - \lambda_{22})\text{tr}(G'_2 \Xi_r G_1)] = 0 \quad (2.13)
\]

for \( r = 1, \cdots, p \). If \( \lim n^{-1}\text{tr}(\Xi^{(s)}_r G_1) \neq 0 \) and \( \lim n^{-1}\text{tr}(G'_1 \Xi_r G_1) \neq 0 \), then the \( r \)-th equation in (2.10) has two roots \( \lambda_{11} = \lambda_{11,0} \), and \( \lambda_{11} = \lambda_{11,0} + \lim n^{-1}\text{tr}(\Xi^{(s)}_r G_1)/\lim n^{-1}\text{tr}(G'_1 \Xi_r G_1) \). In general, if, for \( r \neq s \) \((r, s \in \{1, \cdots, p\})\),

\[
\frac{\lim n^{-1}\text{tr}(\Xi^{(s)}_r G_1)}{\lim n^{-1}\text{tr}(G'_1 \Xi_r G_1)} \neq \frac{\lim n^{-1}\text{tr}(\Xi^{(s)}_r G_1)}{\lim n^{-1}\text{tr}(G'_2 \Xi_r G_1)}
\]

then the system of equations (2.10) has a unique root \( \lambda_{11} = \lambda_{11,0} \). Similarly, \( \lambda_{22,0} \) can be identified from (2.11).

Unfortunately, for equal-sized complete networks, identification of \((\lambda_{11,0}, \lambda_{22,0})\) cannot be established by the above argument. For \( W_C = I_G \otimes [(\bar{n} - 1)^{-1}(\nu_n \nu'_n - I_n)] \), it can be shown that \( G_k = W_C(I_n - \lambda_{kk,0} W_C)^{-1} = a_k I_n + b_k I_G \otimes (\nu_n \nu'_n) \), where \( a_k = (1 - \bar{n} - \lambda_{kk,0})^{-1} \) and \( b_k = -(1 - \lambda_{kk,0})^{-1}(1 - \bar{n} - \lambda_{kk,0})^{-1} \). Hence, the inconsistent roots of (2.10) and (2.11) are given by

\[
\lambda_{kk} = \lambda_{kk,0} + \frac{\lim n^{-1}\text{tr}(\Xi^{(s)}_r G_k)}{\lim n^{-1}\text{tr}(G'_1 \Xi_r G_k)} = \lambda_{kk,0} + \frac{2b_k}{2a_k b_k + \bar{n}b'_k},
\]

for \( k = 1, 2 \), which do not depend on \( \Xi_r \). Therefore, the inconsistent root cannot be ruled out by adding more quadratic moment equations in the form of (2.10) and (2.11). Furthermore, it can be shown that the inconsistent root \((\lambda_{11}, \lambda_{22}) = (\lambda_{11,0} + \frac{2b_1}{2a_1 b_1 + \bar{n}b'_1}, \lambda_{22,0} + \frac{2b_2}{2a_2 b_2 + \bar{n}b'_2})\) is also a solution of (2.12) and (2.13) for any \( \Xi_r \). Hence, quadratic moment equations in the form of (2.12) and (2.13) are not helpful for identification in this particular case either.
Here, we can achieve identification through quadratic moment conditions using the idea of root estimators (see Lee, 2001; Jin and Lee, 2012). As (2.10) is quadratic in $\lambda_{11}$, its two roots can be written as

$$
\lambda_{11} = \lambda_{11,0} + \frac{\lim n^{-1}\text{tr}(\Xi_r^{(s)}G_1) \pm \sqrt{[\lim n^{-1}\text{tr}(\Xi_r^{(s)}G_1)]^2}}{2 \lim n^{-1}\text{tr}(G'_1\Xi_rG_1)}. \tag{2.14}
$$

Hence, to distinguish the consistent root from the inconsistent one, we need to know the sign of $\lim n^{-1}\text{tr}(\Xi_r^{(s)}G_1)$. If $\lim n^{-1}\text{tr}(\Xi_r^{(s)}G_1) > 0$, the consistent root is

$$
\lambda_{11} = \lambda_{11,0} + \frac{\lim n^{-1}\text{tr}(\Xi_r^{(s)}G_1) - \sqrt{[\lim n^{-1}\text{tr}(\Xi_r^{(s)}G_1)]^2}}{2 \lim n^{-1}\text{tr}(G'_1\Xi_rG_1)} = \lambda_{11,0}.
$$

If $\lim n^{-1}\text{tr}(\Xi_r^{(s)}G_1) < 0$, the consistent root is

$$
\lambda_{11} = \lambda_{11,0} + \frac{\lim n^{-1}\text{tr}(\Xi_r^{(s)}G_1) + \sqrt{[\lim n^{-1}\text{tr}(\Xi_r^{(s)}G_1)]^2}}{2 \lim n^{-1}\text{tr}(G'_1\Xi_rG_1)} = \lambda_{11,0}.
$$

To be more specific, consider the following empirical moment functions for the estimation of $\theta_{1,0}$

$$
\begin{align*}
f_{1,1}(\theta_1) &= Q'u_1(\theta_1) \\
f_{2,11}(\theta_1, \theta_1) &= u_1(\theta_1)'\Xi u_1(\theta_1),
\end{align*}
$$

where $Q = [X, \bar{X}]$, $\Xi = W_C$, and $u_1(\theta_1) = y_1 - \lambda_{11}\bar{y}_1 - X\beta_1 - \bar{X}\gamma_1$. Then, from $f_{1,1}(\theta_1) = 0$, we have

$$
(\beta'_1, \gamma'_1)' = (Q'Q)^{-1}Q'(y_1 - \lambda_{11}\bar{y}_1). \tag{2.15}
$$

Let $M = I_n - Q(Q'Q)^{-1}Q'$. Substitution of (2.15) into $f_{2,11}(\theta_1, \theta_1) = 0$ gives

$$
(y_1 - \lambda_{11}\bar{y}_1)'\Xi M(y_1 - \lambda_{11}\bar{y}_1) = 0
$$

10
which has two explicit roots:

\[ \hat{\lambda}_{11} = \frac{y_1'M\Xi M_y_1 \pm \sqrt{(y_1'M\Xi M_y_1)^2 - (y_1'M\Xi M_y_1)(y_1'M\Xi M_y_1)}}{y_1'M\Xi M_y_1}. \]

Since \( \lim_n n^{-1}\text{tr}(\Xi G_1) = 2 \lim_n n^{-1}\text{tr}(WCG_1) = 2(1 - \lambda_{11,0}^{-1})(\bar{n} - 1 + \lambda_{11,0})^{-1} > 0 \) (as long as \( |\lambda_{11,0}| < 1/\rho(WC) = 1 \) and \( \bar{n} \geq 2 \)), Lemma B.1 in the Appendix suggests the consistent root is

\[ \hat{\lambda}_{11} = \frac{y_1'M\Xi M\bar{y}_1 - \sqrt{(y_1'M\Xi M\bar{y}_1)^2 - (y_1'M\Xi M\bar{y}_1)(y_1'M\Xi M\bar{y}_1)}}{y_1'M\Xi M\bar{y}_1}. \]

With \( \lambda_{11,0} \) consistently estimated, \( \beta_{1,0} \) and \( \gamma_{1,0} \) can be consistently estimated by (2.15). \( \theta_{2,0} \) can be identified in a similar manner.

**Remark 1.** The above example shows that even if there is no intransitivity in network connections or variation in group sizes, identification of endogenous and contextual peer effects can still be achieved via quadratic moment conditions. Although the network structure considered in the example is very specific, data with equal-sized complete networks are not rare in the real world. For example, in sports games, the number of players in the starting line up of a team is usually fixed. Without information on how specific players interact with each other within a line up, the starting line ups of all teams can be regarded as equal-sized complete networks. As another example, suppose the researcher observes a complete networks of size \( \bar{n} \) repeatedly over \( G \) time periods. Then, the data would have the same structure as the one described in the above example. The above identification strategy could be helpful in such situations.

### 2.2.2 Multivariate network model

Next, we consider model (2.2) under the exclusion restrictions that \( \phi_{l,k,0} = 0 \) for all \( k \neq l \). The restricted model is

\[ y_k = \sum_{l=1}^m \lambda_{l,k,0}\bar{y}_l + X\beta_{k,0} + \bar{X}\gamma_{k,0} + u_k. \tag{2.16} \]

For the \( m \) activities, we have

\[ Y = \bar{Y}A_0 + XB_0 + \bar{X}\Gamma_0 + U \]
or, equivalently,
\[ y = (\Lambda_0' \otimes W)y + (B'_0 \otimes I_n + \Gamma'_0 \otimes W)x + u, \]
where \( y = \text{vec}(Y) \), \( x = \text{vec}(X) \), and \( u = \text{vec}(U) \). If \( \rho(\Lambda_0) < 1/\rho(W) \), then \( (I_{mn} - \Lambda_0' \otimes W) \) is nonsingular, and thus the reduced-form equation of the multivariate network model is
\[ y = (I_{mn} - \Lambda_0' \otimes W)^{-1}(B'_0 \otimes I_n + \Gamma'_0 \otimes W)x + (I_{mn} - \Lambda_0' \otimes W)^{-1}u, \]  
(2.17)
which implies
\[ \bar{y}_k = H_k(B'_0 \otimes I_n + \Gamma'_0 \otimes W)x + H_ku, \]  
(2.18)
where \( H_k = (i_{m,k} \otimes W)(I_{mn} - \Lambda_0' \otimes W)^{-1} \), with \( i_{m,k} \) denoting the \( k \)-th column of \( I_m \). From (2.16) and (2.18), we have
\[ u_k(\theta_k) = y_k - \sum_{l=1}^m \lambda_{lk}\bar{y}_l - X\beta_k - \bar{X}\gamma_k = d_k(\theta_k) + u_k + \sum_{l=1}^m (\lambda_{lk,0} - \lambda_{lk})H_lu, \]
where
\[ d_k(\theta_k) = \sum_{l=1}^m (\lambda_{lk,0} - \lambda_{lk})E(\bar{y}_l) + X(\beta_k - \beta_k) + \bar{X} (\gamma_k - \gamma_k), \]
with \( \theta_k = (\lambda_{1k}, \cdots, \lambda_{mk}, \beta_k', \gamma_k')' \) and \( E(\bar{y}_l) = H_l(B'_0 \otimes I_n + \Gamma'_0 \otimes W)x \).

For the linear moment functions, we have
\[ \lim_{n \to \infty} n^{-1}E[f_{1,k}(\theta_k)] = \lim_{n \to \infty} n^{-1}E[Q'u_k(\theta_k)] = \lim_{n \to \infty} n^{-1}Q'd_k(\theta_k), \]
for \( k = 1, \cdots, m \). Therefore, \( \lim_{n \to \infty} n^{-1}E[f_{1,k}(\theta_k)] = 0 \) has a unique solution at \( \theta_k = \theta_{k,0} \), if \( Q'[E(\bar{y}_1), \cdots, E(\bar{y}_m), X, \bar{X}] \) has full column rank for large enough \( n \). This sufficient rank condition implies the necessary rank condition that \( [E(\bar{y}_1), \cdots, E(\bar{y}_m), X, \bar{X}] \) has full column rank and the rank of \( Q \) is at least \( 2K + m \), for large enough \( n \). When \( m = 2 \), by a similar argument as in Cohen-Cole et al. (2014) we can show that, under some regularity conditions, \( [E(\bar{y}_1), E(\bar{y}_2), X, \bar{X}] \) has full column rank if \( [X, WX, W^2X, W^3X] \) has full column rank. The full rank condition of
\([X, WX, W^2X, W^3X]\) requires the matrices \(I_n, W, W^2, W^3\) to be linearly independent.

On the other hand, the above necessary rank condition for identification does not hold if, for some \(0 \leq \bar{m} \leq m - 1\), \(E(\bar{y}_l)\) is linearly dependent on \([E(\bar{y}_1), \cdots, E(\bar{y}_m), X, \bar{X}]\) for \(l = \bar{m} + 1, \cdots, m\).\(^8\)

In this case, \(\theta_0\) can be identified if \(\Lambda_0\) can be identified from the quadratic moment equations

\[
\lim_{n \to \infty} n^{-1}E[f_{2,kl}(\theta_k, \theta_l)] = 0, \quad \text{for } k, l = 1, \ldots, m.
\]

The following proposition gives sufficient identification conditions for the multivariate network model. Let \(\sigma_k\) denote the \(k\)-th column of \(\Sigma\).

**Proposition 2.2.** For the multivariate network model (2.16), suppose \(\rho(\Lambda_0) < 1/\rho(W)\). Then \(\theta_0\) is identified if either

(i) \(\lim_{n \to \infty} n^{-1}Q'[E(\bar{y}_1), \cdots, E(\bar{y}_m), X, \bar{X}] \) is finite with full column rank; or

(ii) \(\lim_{n \to \infty} n^{-1}Q'[E(\bar{y}_1), \cdots, E(\bar{y}_m), X, \bar{X}] \) is finite with full column rank for some \(0 \leq \bar{m} \leq m - 1\), and the equations

\[
\lim_{n \to \infty} n^{-1}\{\sum_{s=1}^{m}(\lambda_{sk,0} - \lambda_{sk})\mathrm{tr}[\Xi_s H_s(\sigma_l \otimes I_n)] + \sum_{t=1}^{m}(\lambda_{lt,0} - \lambda_{lt})\mathrm{tr}[\Xi_t H_t(\sigma_k \otimes I_n)]
+ \sum_{s=1}^{m} \sum_{t=1}^{m}(\lambda_{sk,0} - \lambda_{sk})(\lambda_{lt,0} - \lambda_{lt})\mathrm{tr}[\Xi_s H_s(\Xi_t H_t(\Sigma \otimes I_n))]\} = 0, \quad (2.19)
\]

for \(r = 1, \cdots, p\) and \(k, l = 1, \cdots, m\), have a unique solution at \(\Lambda_0\).

**Example 2.** When \(m = 2\), the multivariate network model (2.16) can be written as

\[
\begin{align*}
y_1 &= \lambda_{11,0} \bar{y}_1 + \lambda_{21,0} \bar{y}_2 + X\beta_{1,0} + X\gamma_{1,0} + u_1 \\
y_2 &= \lambda_{12,0} \bar{y}_1 + \lambda_{22,0} \bar{y}_2 + X\beta_{2,0} + X\gamma_{2,0} + u_2.
\end{align*}
\]

Suppose \(\beta_{1,0} = \beta_{2,0} = \gamma_{1,0} = \gamma_{2,0} = 0\) in the DGP. This corresponds to the scenario where none of the exogenous regressors is relevant and thus IVs such as \(W^2X, W^3X\) are not informative. From (2.18), we have \(E(\bar{y}_1) = E(\bar{y}_2) = 0\), which implies \(d_k(\theta_k) = X(\beta_{k,0} - \beta_k) + \bar{X}(\gamma_{k,0} - \gamma_k)\). Hence, $8\)If \(\bar{m} = 0\), then \(E(\bar{y}_1)\) and \([X, \bar{X}]\) are linearly dependent for \(l = 1, \cdots, m\).
from the linear moment equations \( \lim_{n \to \infty} n^{-1}Q'd_k(\theta_k) = 0 \), only \( \beta_{k,0} \) and \( \gamma_{k,0} \) can be identified. Identification of \( \lambda_{k,0} \) has to be achieved through the quadratic moment equations.

It follows from (2.19) that, for \( k = l = 1 \),

\[
(\lambda_{11,0} - \lambda_{11}) \lim_{n \to \infty} n^{-1} \text{tr}[\Xi^{(s)}_r H_1(\sigma_1 \otimes I_n)] + (\lambda_{21,0} - \lambda_{21}) \lim_{n \to \infty} n^{-1} \text{tr}[\Xi^{(s)}_r H_2(\sigma_1 \otimes I_n)] \\
+ (\lambda_{11,0} - \lambda_{11})^2 \lim_{n \to \infty} n^{-1} \text{tr}[H'_2 \Xi H_1(\Sigma \otimes I_n)] + (\lambda_{21,0} - \lambda_{21})^2 \lim_{n \to \infty} n^{-1} \text{tr}[H'_2 \Xi H_2(\Sigma \otimes I_n)] \\
+ 2(\lambda_{11,0} - \lambda_{11})(\lambda_{21,0} - \lambda_{21}) \lim_{n \to \infty} n^{-1} \text{tr}[H'_2 \Xi^{(s)}_r H_2(\Sigma \otimes I_n)] = 0, \\
\tag{2.20}
\]

for \( r = 1, \ldots, p \). If the matrix

\[
\lim_{n \to \infty} n^{-1} \begin{bmatrix}
\text{tr}[\Xi^{(s)}_1 H_1(\sigma_1 \otimes I_n)] & \cdots & \text{tr}[\Xi^{(s)}_p H_1(\sigma_1 \otimes I_n)] \\
\text{tr}[\Xi^{(s)}_1 H_2(\sigma_1 \otimes I_n)] & \cdots & \text{tr}[\Xi^{(s)}_p H_2(\sigma_1 \otimes I_n)] \\
\text{tr}[H'_1 \Xi H_1(\Sigma \otimes I_n)] & \cdots & \text{tr}[H'_1 \Xi H_2(\Sigma \otimes I_n)] \\
\text{tr}[H'_2 \Xi H_1(\Sigma \otimes I_n)] & \cdots & \text{tr}[H'_2 \Xi H_2(\Sigma \otimes I_n)] \\
\end{bmatrix}
\]

has full column rank, (2.20) has a unique solution at \( (\lambda_{11,0}, \lambda_{21,0}) \) and hence \( (\lambda_{11,0}, \lambda_{21,0}) \) can be identified. Similarly, \( (\lambda_{12,0}, \lambda_{22,0}) \) can be identified from (2.19) with \( k = l = 2 \). \( \square \)

2.2.3 Simultaneous equations network model

Now, we consider the identification of the simultaneous equations network model (2.1). As pointed by Yang and Lee (2014), (2.1) has a “pseudo” reduced-form equation

\[
Y = \bar{Y} \Lambda_0^* + XB_0^* + \bar{X} \Gamma_0^* + U^*, \\
\tag{2.21}
\]

where

\[
\Lambda_0^* = \Lambda_0(I_m - \Phi_0)^{-1}, \quad B_0^* = B_0(I_m - \Phi_0)^{-1}, \quad \Gamma_0^* = \Gamma_0(I_m - \Phi_0)^{-1}, \\
\tag{2.22}
\]

\( ^9 \)In this example, we only consider quadratic moment conditions based on the correlation of outcomes within an activity. We could also use quadratic moment conditions based on the correlation of outcomes across different activities, i.e. (2.19) for \( k \neq l \), to improve identification and estimation efficiency.
and $U^* = U(I_m - \Phi_0)^{-1}$. (2.21) can be considered as a multivariate network model with $\text{vec}(U^*) = [(I_m - \Phi_0')^{-1} \otimes I_n]u \sim (0, [(I_m - \Phi_0')^{-1}\Sigma(I_m - \Phi_0)^{-1}] \otimes I_n)$.

Suppose the “pseudo” reduced-form parameters $\Lambda_0^*, B_0^*, \Gamma_0^*$ can be identified by the GMM approach as discussed in Section 2.2.2. Then, the identification problem of the structural parameters $\Theta_0 = [(I_m - \Phi_0)', -\Lambda_0', -B_0', -\Gamma_0']'$ through the linear restrictions (2.22) is essentially the same one as in the classical linear simultaneous equations model (see, e.g., Schmidt, 1976). Let $\vartheta_{k,0}$ denote the $k$-th column of $\Theta_0$. Suppose there are $R_k$ restrictions on $\vartheta_{k,0}$ of the form $R_k \vartheta_{k,0} = 0$ where $R_k$ is a $R_k \times (2m + 2K)$ matrix of known constants. As shown in Yang and Lee (2014), the sufficient and necessary rank condition for $\vartheta_{k,0}$ to be identified by the restrictions $R_k \vartheta_{k,0} = 0$ is that $\text{rank}(R_k \Theta_0) = m - 1$, and the necessary order condition is $R_k \geq m - 1$.

**Example 3.** Consider the model

\[
\begin{align*}
y_1 &= \phi_{21,0}y_2 + \lambda_{11,0}y_1 + X\beta_{1,0} + \bar{X}\gamma_{1,0} + u_1 \\
y_2 &= \phi_{12,0}y_1 + \lambda_{22,0}y_2 + X\beta_{2,0} + \bar{X}\gamma_{2,0} + u_2.
\end{align*}
\]  

Model (2.23) has “pseudo” reduced-form equations

\[
\begin{align*}
y_1 &= \lambda_{11,0}^*y_1 + \lambda_{21,0}^*y_2 + X\beta_{1,0}^* + \bar{X}\gamma_{1,0}^* + u_1^* \\
y_2 &= \lambda_{12,0}^*y_1 + \lambda_{22,0}^*y_2 + X\beta_{2,0}^* + \bar{X}\gamma_{2,0}^* + u_2^*,
\end{align*}
\]  

where

\[
\begin{bmatrix}
\lambda_{11,0}^* & \lambda_{12,0}^* \\
\lambda_{21,0}^* & \lambda_{22,0}^*
\end{bmatrix} = (1 - \phi_{12,0}\phi_{21,0})^{-1}
\begin{bmatrix}
\lambda_{11,0} & \phi_{12,0}\lambda_{11,0} \\
\phi_{21,0}\lambda_{22,0} & \lambda_{22,0}
\end{bmatrix}
\]  

and

\[
\begin{align*}
[\beta_{1,0}^*, \beta_{2,0}^*] &= (1 - \phi_{12,0}\phi_{21,0})^{-1}[\beta_{1,0} + \phi_{21,0}\beta_{2,0}, \beta_{2,0} + \phi_{12,0}\beta_{1,0}] \\
[\gamma_{1,0}^*, \gamma_{2,0}^*] &= (1 - \phi_{12,0}\phi_{21,0})^{-1}[\gamma_{1,0} + \phi_{21,0}\gamma_{2,0}, \gamma_{2,0} + \phi_{12,0}\gamma_{1,0}].
\end{align*}
\]
Suppose the multivariate network model (2.24) can be identified as discussed in Section 2.2.2. Then, the structural parameters
\[
\Theta_0 = [(I_m - \Phi_0)', -\Lambda_0', -B_0', -\Gamma_0]'
\]
can be identified through (2.25)-(2.27) if the rank condition holds. The exclusion restriction for the first equation of model (2.23) can be represented by
\[
R_1 = [0, 0, 0, 0_{1 \times K}, 0_{1 \times K}]
\]
Then \(R_1 \Theta_0 = [0, -\lambda_{22,0}],\) which has rank 1 if \(\lambda_{22,0} \neq 0.\)
Similarly, the exclusion restriction for the second equation can be represented by
\[
R_2 = [0, 0, 1, 0_{1 \times K}, 0_{1 \times K}]
\]
Then \(R_2 \Theta_0 = [-\lambda_{11,0}, 0]\) which has rank 1 if \(\lambda_{11,0} \neq 0.\) Indeed, if \(\lambda_{11,0} = \lambda_{22,0} = 0,\) (2.23) becomes a classical linear simultaneous equations model, which cannot be identified without imposing exclusion restrictions on \(\beta_{k,0}\) and \(\gamma_{k,0}.\)

3 Group Fixed Effects

3.1 Model and fixed-effect GMM estimator

Suppose there are \(G\) groups in the sample, with \(n_g\) individuals in the \(g\)-th group and \(n = \sum_{g=1}^{G} n_g.\)

For the \(g\)-th group, the econometric model is given by
\[
Y_g = Y_g \Phi_0 + W_g Y_g A_0 + X_g B_0 + W_g X_g \Gamma_0 + \eta_{g1} \eta_g + U_g,
\]
where \(W_g\) is the adjacency matrix for the \(g\)-th network and \(\eta_g = (\eta_{g1}, \cdots, \eta_{gm})\) is a \(1 \times m\) vector of group- and activity-specific parameters.

Let \(Y = [Y_1', \cdots, Y_G]', X = [X_1', \cdots, X_G]', U = [U_1', \cdots, U_G]', W = \text{diag}(W_g),\) and \(L = \text{diag}(\eta_{g}).\) Then, for all the \(G\) groups,
\[
Y = Y \Phi_0 + \bar{Y} A_0 + X B_0 + \bar{X} \Gamma_0 + L \eta + U,
\]
where \(\bar{Y} = WY, \bar{X} = WX,\) and \(\eta = [\eta_{1}', \cdots, \eta_{G}'].\)
As in Section 2.1, \( Y_k, \bar{Y}_k, X_k \) and \( \bar{X}_k \) are matrices containing columns of \( Y, \bar{Y}, X \) and \( \bar{X} \) that appear in the \( k \)-th equation under some exclusion restrictions. Then, the \( k \)-th equation of model (3.1) is
\[
y_k = Y_k \phi_{k,0} + \bar{Y}_k \lambda_{k,0} + X_k \beta_{k,0} + \bar{X}_k \gamma_{k,0} + L \eta_k + u_k,
\]
where \( \eta_k \) denotes the \( k \)-th column of \( \eta \).

We allow \( \eta \) to depend on \( W \) and \( X \) by treating \( \eta \) as a \( G \times m \) matrix of unknown parameters. When the number of network \( G \) is large, we may have the “incidental parameter” problem (Neyman and Scott, 1948). To avoid this problem, we transform (3.1) with a deviation from group mean projector \( J = \text{diag}(J_g) \) where \( J_g = I_{n_g} - n_g^{-1} l_{n_g} l_{n_g}' \). This transformation is analogous to the “within” transformation for the \( \text{fixed-effect panel data model} \). As \( JL = 0 \), the \( k \)-th equation of the within-transformed model is
\[
Jy_k = JY_k \phi_{k,0} + J\bar{Y}_k \lambda_{k,0} + JX_k \beta_{k,0} + J\bar{X}_k \gamma_{k,0} + Ju_k.
\]

As discussed in Section 2.1, the GMM estimator utilizes both linear and quadratic moment functions. The linear moment functions are given by
\[
g_{1,k}(\theta_k) = \bar{Q}'u_k(\theta_k), \quad \text{for } k = 1, \ldots, m,
\]
where \( u_k(\theta_k) \) is given by (2.3) and \( \bar{Q} = JQ \) is the within-transformed IV matrix. The quadratic moment functions are given by
\[
g_{2,kl}(\theta_k, \theta_l) = [\hat{z}_1' u_k(\theta_k), \ldots, \hat{z}_p' u_k(\theta_k)]' u_l(\theta_l), \quad k, l = 1, \ldots, m,
\]
where \( \hat{z}_r = J\Xi_r J - \text{tr}(J\Xi_r) J / \text{tr}(J) \), for \( r = 1, \ldots, p \). By construction, \( \text{tr}(\hat{z}_r) = 0 \) and \( J\hat{z}_r J = \hat{z}_r \).

At the true parameter value, \( E[g_{1,k}(\theta_{k,0})] = \bar{Q}'E(u_k) = 0 \) and \( E[g_{2,kl}(\theta_{k,0}, \theta_{l,0})] = E[(\hat{z}_1' u_k, \ldots, \hat{z}_p' u_k)' u_l] = \sigma_{kl}[\text{tr}(\hat{z}_1), \ldots, \text{tr}(\hat{z}_p)]' = 0 \).

Let \( g_1(\theta) = [g_{1,1}(\theta_1)', \ldots, g_{1,m}(\theta_m)']' \), \( g_2(\theta) = [g_{2,11}(\theta_1, \theta_1)', \ldots, g_{2,1m}(\theta_1, \theta_m)', g_{2,21}(\theta_2, \theta_1)', \ldots] \).
\[
\cdots, g_{2,mm}(\theta_m, \theta_m)', \text{ and}
\]
\[
g(\theta) = [g_1(\theta)', g_2(\theta)']'.
\] (3.3)

Let \( \Omega_g = \text{Var}[g(\theta_0)] \). Suppose \( n^{-1} \hat{\Omega}_g \) is a consistent estimator of \( n^{-1} \Omega_g \) such that \( n^{-1} \hat{\Omega}_g - n^{-1} \Omega_g = o_p(1) \). The fixed-effect GMM estimator for \( \theta_0 \) is given by
\[
\hat{\theta}_{gmm} = \arg \min g(\theta)' \hat{\Omega}_g^{-1} g(\theta).
\] (3.4)

3.2 Identification

The group fixed effect captures the correlated effect, wherein agents in the same network may behave similarly as they have similar unobserved individual characteristics or they face similar institutional environment (Manski, 1993). In particular, the group fixed effect may serve as a partial remedy for the selection bias that originates from the possible sorting of individuals with similar unobserved characteristics into a network.

In the following, we revisit the identification results discussed in Section 2.2 in the presence of group fixed effects. In general, for \( \theta_0 \) to be identified by the fixed-effect GMM approach, the moment equations \( \lim_{n \to \infty} n^{-1}E[g(\theta)] = 0 \) need to have a unique solution at \( \theta = \theta_0 \).

3.2.1 SUR network model

First, we consider the SUR network model with group fixed effects. Under the exclusion restrictions that \( \phi_{lk,0} = \lambda_{lk,0} = 0 \) for all \( k \neq l \), the \( k \)-th equation of the SUR network model is given by
\[
y_k = \lambda_{kk,0} \bar{y}_k + X\beta_{k,0} + X\gamma_{k,0} + L\eta_k + u_k.
\] (3.5)

The within-transformed model is given by
\[
Jy_k = \lambda_{kk,0} J\bar{y}_k + JX\beta_{k,0} + JX\gamma_{k,0} + Ju_k.
\] (3.6)
If $|\lambda_{kk,0}| < 1/\rho(W)$, then $(I_n - \lambda_{kk,0}W)$ is nonsingular, and thus the reduced-form equation of (3.5) is given by

$$y_k = (I_n - \lambda_{kk,0}W)^{-1}(X\beta_{k,0} + \bar{X}\gamma_{k,0} + Ln_k) + (I_n - \lambda_{kk,0}W)^{-1}u_k,$$

which implies

$$\bar{y}_k = G_k(X\beta_{k,0} + \bar{X}\gamma_{k,0} + Ln_k) + G_ku_k,$$

where $G_k = W(I_n - \lambda_{kk,0}W)^{-1}$. From (3.6) and (3.7), we have

$$Ju_k(\theta_k) = Jd_k(\theta_k) + Ju_k + (\lambda_{kk,0} - \lambda_{kk})JG_ku_k,$$

where

$$Jd_k(\theta_k) = (\lambda_{kk,0} - \lambda_{kk})E(J\bar{y}_k) + JX(\beta_{k,0} - \beta_k) + J\bar{X}(\gamma_{k,0} - \gamma_k),$$

with $\theta_k = (\lambda_{kk}, \beta_k', \gamma_k')$ and $E(J\bar{y}_k) = JG_k(X\beta_{k,0} + \bar{X}\gamma_{k,0} + Ln_k)$.

For the linear moment functions, we have

$$\lim_{n \to \infty} n^{-1}E[g_{1,k}(\theta_k)] = \lim_{n \to \infty} n^{-1}E[Q'Ju_k(\theta_k)] = \lim_{n \to \infty} n^{-1}E[Q'Jd_k(\theta_k)],$$

for $k = 1, \cdots, m$. Therefore, $\lim_{n \to \infty} n^{-1}E[g_{1,k}(\theta_k)] = 0$ has a unique solution at the true parameter value, if $Q'[E(J\bar{y}_k), JX, J\bar{X}]$ has full column rank for large enough $n$. This sufficient rank condition implies the necessary rank condition that $[E(J\bar{y}_k), JX, J\bar{X}]$ has full column rank and the rank of $Q$ is at least $2K + 1$, for large enough $n$.

The term $G_kL$ in the reduced-form equation (3.7) has information on the Bonacich centrality. If $W_g$ has constant row sums (including the case that $W_g$ is row-normalized) for all $g$, then $JG_kL = 0$, and thus $E(J\bar{y}_k) = JG_k(X\beta_{k,0} + \bar{X}\gamma_{k,0})$. In this case, Cohen-Cole et al. (2014) show that $[E(J\bar{y}_k), JX, J\bar{X}]$ has full column rank if $\lambda_{kk,0}\beta_{k,0} + \gamma_{k,0} \neq 0$ and $[X, WX, W^2X, W^3X]$ has full column rank. As discussed in Section 2.2, this rank condition essentially requires that the $G_kL = \text{diag}(W_g(I_{n_g} - \lambda_{kk,0}W_g)^{-1}t_{n_g})$ where $W_g(I_{n_g} - \lambda_{kk,0}W_g)^{-1}t_{n_g}$ is the Bonacich centrality for the $g$-th group (Bonacich, 1987).
matrices $I, W, W^2, W^3$ to be linearly independent. On the other hand, if $W_g$ has non-constant row sums, then $JG_kL \neq 0$. In this case, the Bonacich centrality measure provides additional information to identify the endogenous peer effect, and hence the identification condition is in general weaker. In particular, Liu (2014) shows that $[E(J\tilde{y}_k), JX, J\tilde{X}]$ has full column rank if $W_g$ is symmetric and has non-constant row sums.

When the above necessary rank condition does not hold, the quadratic moment equations

$$
\lim_{n \to \infty} n^{-1} E[g_{2,kl}(\theta_k, \theta_l)] = 0, \quad \text{for } k, l = 1, \ldots, m,
$$

provide another channel for identification as discussed in Section 2.2.1. The following proposition summarizes sufficient identification conditions for the SUR network model with group fixed effects.

**Proposition 3.1.** For the SUR network model with group fixed effects given by (3.5), suppose $|\lambda_{kk,0}| < 1/\rho(W)$. Then $\theta_{k,0} = (\lambda_{kk,0}, \beta'_{kk,0}, \gamma'_{kk,0})'$ is identified if either

(i) $\lim_{n \to \infty} n^{-1} Q'[E(J\tilde{y}_k), JX, J\tilde{X}]$ is finite with full column rank; or

(ii) $\lim_{n \to \infty} n^{-1} Q'[JX, J\tilde{X}]$ is finite with full column rank and the equations

$$
\sigma_{kl} \lim_{n \to \infty} n^{-1} [(\lambda_{kk,0} - \lambda_{kk}) \text{tr}(\tilde{\Xi}_k G_k) + (\lambda_{ll,0} - \lambda_{ll}) \text{tr}(\tilde{\Xi}_l G_l) + (\lambda_{kk,0} - \lambda_{kk})(\lambda_{ll,0} - \lambda_{ll}) \text{tr}(G_k G_l)] = 0,
$$

for $r = 1, \cdots, p$ and $l = 1, \ldots, m$, have a unique solution for $\lambda_{kk}$ at the true parameter value.

The sufficient conditions in Proposition 3.1 are stronger than those in Proposition 2.1. For equal-sized completed networks considered in Example 1, neither conditions in Proposition 3.1 are satisfied since $J\tilde{X} = JWTX = -(\bar{n} - 1)^{-1}JX$. Lee (2007a) and Davezies et al. (2009) show that, in the presence of group fixed effects, peer effects in complete networks can be identified if there are sufficient variations in group sizes.

**Example 4 (Star Networks).** Suppose $n$ individuals are randomly assigned into $G$ equal-sized groups with $\bar{n} = n/G$ individuals in each group. Suppose the first individual in each group is connected with all the other group members. The other group members are not connected with
each other. The corresponding (row-normalized) adjacency matrix is given by

\[
W_S = I_G \otimes \begin{bmatrix}
0 & (\bar{n} - 1)^{-1} & \cdots & (\bar{n} - 1)^{-1} \\
1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 0 & \cdots & 0
\end{bmatrix}.
\] (3.9)

As \( JW_S^2 = -JW_S \), from the reduced-form equation (3.7), one can show that \( E(J y_k) \) and \( [JX, J\tilde{X}] \) are linearly dependent.\(^{11}\) Hence, it is not possible to identify the peer effects based on the linear moment conditions alone.\(^{12}\) However, the model may still be identifiable through additional quadratic moment conditions. To illustrate the idea, consider the case \( m = 2 \) and \( \sigma_{kl} \neq 0 \) for \( k, l = 1, 2 \). Then, from (3.8) we have

\[
\lim_{n \to \infty} n^{-1} \left[ (\lambda_{11,0} - \lambda_{11}) \text{tr}(\tilde{\Xi}_{1}^{(s)} G_1) + (\lambda_{11,0} - \lambda_{11})^2 \text{tr}(G_1' \tilde{\Xi}_r G_1) \right] = 0 \tag{3.10}
\]

\[
\lim_{n \to \infty} n^{-1} \left[ (\lambda_{22,0} - \lambda_{22}) \text{tr}(\tilde{\Xi}_{2}^{(s)} G_2) + (\lambda_{22,0} - \lambda_{22})^2 \text{tr}(G_2' \tilde{\Xi}_r G_2) \right] = 0 \tag{3.11}
\]

and

\[
\lim_{n \to \infty} n^{-1} \left[ (\lambda_{11,0} - \lambda_{11}) \text{tr}(\tilde{\Xi}_{1}^{(s)} G_1) + (\lambda_{22,0} - \lambda_{22}) \text{tr}(\tilde{\Xi}_r G_2) + (\lambda_{11,0} - \lambda_{11})(\lambda_{22,0} - \lambda_{22}) \text{tr}(G_1' \tilde{\Xi}_r G_2) \right] = 0 \tag{3.12}
\]

\[
\lim_{n \to \infty} n^{-1} \left[ (\lambda_{22,0} - \lambda_{22}) \text{tr}(\tilde{\Xi}_{2}^{(s)} G_2) + (\lambda_{11,0} - \lambda_{11}) \text{tr}(\tilde{\Xi}_r G_1) + (\lambda_{11,0} - \lambda_{11})(\lambda_{22,0} - \lambda_{22}) \text{tr}(G_2' \tilde{\Xi}_r G_1) \right] = 0 \tag{3.13}
\]

for \( r = 1, \cdots, p \). If \( \lim n^{-1} \text{tr}(\tilde{\Xi}_{1}^{(s)} G_1) \neq 0 \) and \( \lim n^{-1} \text{tr}(G_1' \tilde{\Xi}_r G_1) \neq 0 \), the \( r \)-th equation in (3.10) has two roots \( \lambda_{11} = \lambda_{11,0} \) and \( \lambda_{11} = \lambda_{11,0} + \lim n^{-1} \text{tr}(\tilde{\Xi}_{1}^{(s)} G_1)/\lim n^{-1} \text{tr}(G_1' \tilde{\Xi}_r G_1) \). In general, if,

\(^{11}\) On the other hand, if \( W_S \) is not row-normalized with all of its non-zero elements being ones, then it follows by Corollary 1 in Liu (2014) that \( \text{E}(J y_k), JX, J\tilde{X} \) has full column rank.

\(^{12}\) Bramoullé et al. (2009) have a similar non-identification result. According to Proposition 5 in Bramoullé et al. (2009), the star network cannot be identified since \( W_S^2 = W_S \) and \( \text{rank}(I_n - W_S) = G(\bar{n} - 1) \).
for \( r \neq s \ (r, s \in \{1, \ldots, p\}) \),
\[
\frac{\lim n^{-1} \text{tr}(\mathbf{E}_r^{(s)} \mathbf{G}_1)}{\lim n^{-1} \text{tr}(\mathbf{G}_1^{(s)} \mathbf{G}_1)} \neq \frac{\lim n^{-1} \text{tr}(\mathbf{E}_r^{(s)} \mathbf{G}_1)}{\lim n^{-1} \text{tr}(\mathbf{G}_1^{(s)} \mathbf{G}_1)}
\]
then the system of equations (3.10) has a unique root \( \lambda_{11} = \lambda_{11,0} \). Similarly, \( \lambda_{22,0} \) can be identified from (3.11).

However, for star networks, identification of \((\lambda_{11,0}, \lambda_{22,0})\) cannot be established by the above argument. For \( \mathbf{W}_S \) given by (3.9), we have \( \mathbf{JG}_k = \mathbf{JW}_S(\mathbf{I}_n - \lambda_{kk,0} \mathbf{W}_S)^{-1} = (1 + \lambda_{kk,0})^{-1} \mathbf{JW}_S \).
Hence, the inconsistent roots of (3.10) and (3.11) are given by
\[
\lambda_{kk} = \lambda_{kk,0} + \frac{\lim n^{-1} \text{tr}(\mathbf{E}_r^{(s)} \mathbf{G}_k)}{\lim n^{-1} \text{tr}(\mathbf{G}_1^{(s)} \mathbf{G}_k)} = -\lambda_{kk,0} - 2,
\]
for \( k = 1, 2 \), which do not depends on \( \mathbf{E}_r \). Therefore, the inconsistent root cannot be ruled out by adding more quadratic moment equations in the form of (3.10) and (3.11). Furthermore, it can be shown that the inconsistent root \((\lambda_{11}, \lambda_{22}) = (-\lambda_{11,0} - 2, -\lambda_{22,0} - 2)\) is also a solution of (3.12) and (3.13) for any \( \mathbf{E}_r \). Hence, quadratic moment equations in the form of (3.12) and (3.13) are not helpful for identification in this particular case either.

Similar to the identification strategy considered in Example 1, to identify the consistent root, we need to know the sign of \( \lim n^{-1} \text{tr}(\mathbf{E}_r^{(s)} \mathbf{G}_k) \). Suppose we estimate \( \mathbf{\theta}_{1,0} \) with the following empirical moment functions
\[
g_{1,1}(\mathbf{\theta}_1) = \mathbf{Q}' \mathbf{u}_1(\mathbf{\theta}_1) \\
g_{2,11}(\mathbf{\theta}_1, \mathbf{\theta}_1) = \mathbf{u}_1(\mathbf{\theta}_1)' \mathbf{E}_1 \mathbf{u}_1(\mathbf{\theta}_1),
\]
where \( \mathbf{Q} = \mathbf{J}[\mathbf{X}, \mathbf{\tilde{X}}], \mathbf{E} = \mathbf{JW}_S \mathbf{J} - \text{tr}(\mathbf{JW}_S) \mathbf{J}/\text{tr}(\mathbf{J}), \) and \( \mathbf{u}_1(\mathbf{\theta}_1) = \mathbf{y}_1 - \lambda_{11} \mathbf{\tilde{y}}_1 - \mathbf{X} \beta_1 - \mathbf{\tilde{X}} \gamma_1. \) Then, from \( g_{1,1}(\mathbf{\theta}_1) = 0 \), we have
\[
(\beta_1', \gamma_1')' = (\mathbf{Q}' \mathbf{Q})^{-1} \mathbf{Q}'(\mathbf{y}_1 - \lambda_{11} \mathbf{\tilde{y}}_1).
\]
(3.14)
Let $\bar{M} = I_n - \bar{Q}(\bar{Q}'\bar{Q})^{-1}\bar{Q}'$. Substitution of (3.14) into $g_{2,11}(\theta_1, \theta_1) = 0$ gives

$$(y_1 - \lambda_{11}\bar{y}_1)'\bar{M}\bar{M}(y_1 - \lambda_{11}\bar{y}_1) = 0$$

which has two explicit roots:

$$\hat{\lambda}_{11} = y_1'\bar{M}\bar{M}\bar{y}_1 \pm \sqrt{(y_1'\bar{M}\bar{M}\bar{y}_1)^2 - (y_1'\bar{M}\bar{M}\bar{y}_1)(y_1'\bar{M}\bar{M}\bar{y}_1)^{1/2}}.$$  

Since $\lim n^{-1}\text{tr}(\bar{S}^{(n)}G_1) = 2(1 + \lambda_{11,0})^{-1}\frac{n-2}{n-1} \rho(S) > 0$ (as long as $|\lambda_{11,0}| < 1/\rho(W_S) = 1$ and $n > 2$), Lemma B.1 in the Appendix suggests the consistent root is

$$\hat{\lambda}_{11} = y_1'\bar{M}\bar{M}\bar{y}_1 - \sqrt{(y_1'\bar{M}\bar{M}\bar{y}_1)^2 - (y_1'\bar{M}\bar{M}\bar{y}_1)(y_1'\bar{M}\bar{M}\bar{y}_1)^{1/2}}.$$  

With $\lambda_{11,0}$ consistently estimated, $\beta_{1,0}$ and $\gamma_{1,0}$ can be consistently estimated by (3.14). $\theta_{2,0}$ can be identified in a similar manner.

**3.2.2 Multivariate network model**

Next, we consider the multivariate network model with group fixed effects. Under the exclusion restrictions that $\phi_{ik,0} = 0$ for all $k \neq l$, the $k$-th equation of the restricted model is given by

$$y_k = \sum_{l=1}^m \lambda_{lk,0}\bar{y}_l + X\beta_{k,0} + \bar{X}\gamma_{k,0} + L\eta_k + \bar{u}_k.$$  

(3.15)

The within-transformed model is given by

$$Jy_k = \sum_{l=1}^m \lambda_{lk,0}\bar{y}_l + JX\beta_{k,0} + J\bar{X}\gamma_{k,0} + Ju_k.$$  

(3.16)

For the $m$ activities, model (3.15) can be written as

$$Y = \bar{Y}A_0 + X\beta_0 + \bar{X}\Gamma_0 + L\eta + U$$
or, equivalently,

\[ y = (A'_0 \otimes W)y + (B'_0 \otimes I_n + \Gamma'_0 \otimes W)x + (I_m \otimes L)\text{vec}(\eta) + u, \]

where \( y = \text{vec}(Y) \), \( x = \text{vec}(X) \), and \( u = \text{vec}(U) \). If \( \rho(A_0) < 1/\rho(W) \), then \((I_{mn} - A'_0 \otimes W)\) is nonsingular, and thus the reduced-form equation is

\[ y = (I_{mn} - A'_0 \otimes W)^{-1}[(B'_0 \otimes I_n + \Gamma'_0 \otimes W)x + (I_m \otimes L)\text{vec}(\eta)] + (I_{mn} - A'_0 \otimes W)^{-1}u, \]

which implies

\[ \bar{y}_k = H_k[(B'_0 \otimes I_n + \Gamma'_0 \otimes W)x + (I_m \otimes L)\text{vec}(\eta)] + H_ku, \quad (3.17) \]

where \( H_k = (I_{m,k} \otimes W)(I_{mn} - A'_0 \otimes W)^{-1} \). From (3.15) and (3.17), we have

\[ Ju_k(\theta_k) = Jd_k(\theta_k) + Ju_k + \sum_{l=1}^{m} (\lambda_{lk,0} - \lambda_{lk})JH_lu, \]

where

\[ Jd_k(\theta_k) = \sum_{l=1}^{m} (\lambda_{lk,0} - \lambda_{lk})E(J\bar{y}_l) + JX(\beta_{l,k,0} - \beta_{l,k}) + J\bar{X}(\gamma_{l,k,0} - \gamma_{l,k}), \]

with \( \theta_k = (\lambda_{1k}, \ldots, \lambda_{mk}, \beta'_{k, \gamma'_k})' \) and \( E(J\bar{y}_l) = JH_l[(B'_0 \otimes I_n + \Gamma'_0 \otimes W)x + (I_m \otimes L)\text{vec}(\eta)] \).

For the linear moment functions, we have

\[ \lim_{n \to \infty} n^{-1}E[g_{1k}(\theta_k)] = \lim_{n \to \infty} n^{-1}E[\bar{Q}'u_k(\theta_k)] = \lim_{n \to \infty} n^{-1}E[Q'Ju_k(\theta_k)] = \lim_{n \to \infty} n^{-1}Q'Jd_k(\theta_k), \]

for \( k = 1, \ldots, m \). Therefore, \( \lim_{n \to \infty} n^{-1}E[g_{1k}(\theta_k)] = 0 \) has a unique solution at \( \theta_k = \theta_{k,0} \), if \( Q'[E(J\bar{y}_1), \ldots, E(J\bar{y}_m), JX, J\bar{X}] \) has full column rank for large enough \( n \). This sufficient rank condition implies the necessary rank condition that \( [E(J\bar{y}_1), \ldots, E(J\bar{y}_m), JX, J\bar{X}] \) has full column rank and the rank of \( Q \) is at least \( 2K + m \), for large enough \( n \).

The term \( H_k(I_m \otimes L) \) in the reduced-form equation (3.17) can be interpreted as a centrality measure that takes into account interactions in different activities. If \( W_g \) has constant row sums
for all $g$, then $JH_k(I_m \otimes L) = 0$, and thus $E(Jy_k) = JH_k(B'_0 \otimes I_n + \Gamma'_0 \otimes W)x$. When $m = 2$, Cohen-Cole et al. (2014) show that, under some regularity conditions, $[E(Jy_1), E(Jy_2), JX, JX]$ has full column rank if $[X, WX, W^2X, W^3X, W^4X]$ has full column rank. This rank condition essentially requires that the matrices $I_n, W, W^2, W^3, W^4$ to be linearly independent. On the other hand, if $W_g$ has non-constant row sums, then $JH_k(I_m \otimes L) \neq 0$. In this case, $H_k(I_m \otimes L)$ provides additional information to identify the endogenous peer effect. When $m = 2$, Liu (2014) provides some sufficient conditions for $[E(Jy_1), E(Jy_2), JX, JX]$ to have full column rank.

When the above necessary rank condition for identification does not hold, identification may still be possible through the quadratic moment equations

$$\lim_{n \to \infty} n^{-1} E[g_{2,kl}(\theta_k, \theta_l)] = 0, \quad l = 1, \ldots, m,$$

as discussed in Section 2.2.2. The following proposition gives sufficient identification conditions for the multivariate network model with group fixed effects.

**Proposition 3.2.** For the multivariate network model with group fixed effects given by (3.15), suppose $\rho(A_0) < 1/\rho(W)$. Then $\theta_0$ is identified if either

(i) $\lim_{n \to \infty} n^{-1} Q'[E(Jy_1), \ldots, E(Jy_m), JX, JX]$ is finite with full column rank; or

(ii) $\lim_{n \to \infty} n^{-1} Q'[E(Jy_1), \ldots, E(Jy_{\bar{m}}), JX, JX]$ is finite with full column rank for some $0 \leq \bar{m} \leq m - 1$, and the equations

$$\lim_{n \to \infty} n^{-1} \left\{ \sum_{s=1}^{m} (\lambda_{sk,0} - \lambda_{sk}) \text{tr}[\tilde{\Sigma} H_s(\sigma_l \otimes I_n)] + \sum_{t=1}^{m} (\lambda_{lt,0} - \lambda_{lt}) \text{tr}[\tilde{\Sigma} H_t(\sigma_k \otimes I_n)] ight\} + \sum_{s=1}^{m} \sum_{t=1}^{m} (\lambda_{sk,0} - \lambda_{sk})(\lambda_{lt,0} - \lambda_{lt}) \text{tr}[H_s' \tilde{\Sigma} H_t(\Sigma \otimes I_n)] = 0,$$

for $r = 1, \ldots, p$ and $k, l = 1, \ldots, m$, have a unique solution at $A_0$.

### 3.2.3 Simultaneous equations network model

Finally, we consider the identification of the simultaneous equations network model with group fixed effects. As discussed in Section 2.2.3, if $(I_m - \Phi_0)$ is nonsingular, then (3.1) has a “pseudo”
reduced-form equation
\[ Y = \bar{Y}A_0^* + XB_0^* + \bar{X}\Gamma_0^* + L\eta^* + U^*, \quad (3.18) \]
where \( A_0^*, B_0^*, \Gamma_0^* \) are given by (2.22), \( \eta^* = \eta(I_m - \Phi_0)^{-1} \) and \( U^* = U(I_m - \Phi_0)^{-1} \). Premultiplication both sides of (3.18) by \( J \), we have
\[ JY = J\bar{Y}A_0^* + JXB_0^* + J\bar{X}\Gamma_0^* + JU^*. \quad (3.19) \]
(3.19) is a within-transformed multivariate network model, and its parameters \( A_0^*, B_0^*, \Gamma_0^* \) can be identified as discussed above. Thus, the identification problem of the structural parameters \( \Theta_0 = [(I_m - \Phi_0)', -A_0^*, -B_0^*, -\Gamma_0^*]' \) through the linear restrictions (2.22) is essentially the same one as in the classical linear simultaneous equations model. Suppose there are \( R_k \) restrictions of the form \( R_k \theta_{k,0} = 0 \) where \( R_k \) is an \( R \times (2m + 2K) \) matrix of known constants and \( \theta_{k,0} \) is the \( k \)-th column of \( \Theta_0 \). The sufficient and necessary rank condition for \( \theta_{k,0} \) to be identified by the restrictions \( R_k \theta_{k,0} = 0 \) is that \( \text{rank}(R_k \Theta_0) = m - 1 \), and the necessary order condition is \( R_k \geq m - 1 \).

4 Estimation

4.1 Fixed-effect IV estimator

The \( k \)-th equation of the simultaneous equations network model given by (3.2) can be written more compactly as
\[ y_k = Z_k \theta_{k,0} + L\eta_{k,0} + u_k, \quad (4.1) \]
where \( Z_k = [Y_k, \bar{Y}_k, X_k, \bar{X}_k] \), for \( k = 1, \cdots, m \). As \( JL = 0 \), the within transformation of (4.1) with the projector \( J \) gives
\[ Jy_k = JZ_k \theta_{k,0} + Ju_k. \]

Since \( E(JZ_k) \) is an optimal IV matrix for \( JZ_k \), obviously \( E[JY, J\bar{Y}, JX, J\bar{X}] \) is also an optimal IV matrix for \( JZ_k \). As pointed out by Kelejian and Prucha (2004), \( E(Y) = \sum_{j=0}^{\infty} W^j[X, \bar{X}, L]C_j \),
where \( C_j \) is a \((2K + G) \times m\) matrix whose elements are functions of the elements of \( \Phi_0, \Lambda_0, B_0, \Gamma_0 \) and \( \eta \). Therefore, the optimal IV matrix for \( JZ_k \) can be considered as a linear combination of\(^{13}\)

\[
\bar{Q}_\infty = [JX, JW^2X, \cdots, JLW, JW^2L, \cdots].
\]

For the GMM estimation, the IV matrix \( \bar{Q} \) contains a subset of the linearly independent columns of \( \bar{Q}_\infty \). We allow \( q \), the column rank of \( \bar{Q} \), to increase with the sample size in the asymptotic analysis so that the proposed estimator can be asymptotically efficient.\(^{14}\)

First, we consider the GMM estimator using only linear moment functions, \( g_1(\theta) = [g_{1,1}(\theta_1)', \cdots, g_{1,m}(\theta_m)]', \) where \( g_{1,k}(\theta_k) = \bar{Q}'u_k(\theta_k) \) for \( k = 1, \cdots, m \). As \( \Omega_{11} = \text{Var}[g_1(\theta_0)] = \Sigma \otimes (\bar{Q}'\bar{Q}) \), we let \( \Omega_{11} = \tilde{\Sigma} \otimes (\bar{Q}'\bar{Q}) \), where \( \tilde{\Sigma} \) is a \( \sqrt{n} \)-consistent estimator of \( \Sigma \). Let \( Z = \text{diag}(Z_k) \). Thus, the fixed-effect GMM estimator reduces to the 3SLS estimator given by

\[
\hat{\theta}_{3sls} = \arg \min g_1(\theta)'\tilde{\Sigma}_{11}^{-1}g_1(\theta) = \arg \min u(\theta)'(\tilde{\Sigma}^{-1} \otimes \tilde{P})u(\theta)
= [Z'(\tilde{\Sigma}^{-1} \otimes \tilde{P})Z]^{-1}Z'(\tilde{\Sigma}^{-1} \otimes \tilde{P})y,
\]

where \( u(\theta) = [u_1(\theta_1)', \cdots, u_m(\theta_m)]' = y - Z\theta \) and \( \tilde{P} = \bar{Q}(\bar{Q}'\bar{Q})^{-1}\bar{Q}' \).\(^{15}\) To study the asymptotic properties of the fixed-effect 3SLS estimator, we maintain the following assumptions. Let \( u_{ik} \) denote the \((i, k)\)-th element of \( U \).

**Assumption 1.** \((u_{i1}, \cdots, u_{im})' \sim i.i.d.(0, \Sigma)\), where \( \Sigma \) is an \( m \times m \) nonsingular matrix. For some \( \delta > 0 \), \( E|u_{ik}u_{il}u_{is}u_{it}|^{1+\delta} \) is bounded by some finite constant for any \( i = 1, \cdots, n \) and \( k, l, s, t = 1, \cdots, m \).

**Assumption 2.** \((I_m - \Phi_0)\) is nonsingular and \( \rho(\Lambda_0(I_m - \Phi_0)^{-1}) < 1/\rho(W) \).

\(^{13}\)If \( W_g \) has constant row sums for all \( g \), then \( JW^lL = 0 \) for any positive integer \( l \). In this case, \( \bar{Q}_\infty = [JX, JW^2X, JW^2X, \cdots] \).

\(^{14}\)Another important reason to consider many.instrument asymptotics (Bekker, 1994) in this paper is as follows. If \( W_g \) has non-constant row sums, then \( WL \) is the leading order term of the Bonacich centrality (Bonacich, 1987). \( WL \) have \( G \) columns. If \( JWL \) is included in the IV matrix \( \bar{Q} \), then the number of IVs would increase with \( G \), the number of groups in the data.

\(^{15}\)If we consider the less efficient equation-by-equation estimation using the linear moment function \( g_{1,k}(\theta_k) \), the fixed-effect GMM estimator reduces to the 2SLS estimator given by \( \hat{\theta}_{k,2sls} = \arg \min g_{1,k}(\theta_k)'\text{Var}[g_{1,k}(\theta_{k,0})]^{-1}g_{1,k}(\theta_k)(Z_k'\tilde{P}Z_k)^{-1}Z_k'\tilde{P}y_k \).
Assumption 3. The row and column sums of $W$ and $(I_{mn} - \Phi_0 \otimes I_n - \Lambda_0' \otimes W)^{-1}$ are uniformly bounded in absolute value.

Assumption 4. The matrix of exogenous regressors $X$ has full column rank for $n$ sufficiently large. The elements of $X$ are uniformly bounded constants.

Assumption 1-4 are from Kelejian and Prucha (2004) and Yang and Lee (2014). In particular, Assumption 2 imposes a restriction on the parameter space so that the underlying network game has a unique interior Nash equilibrium (see Proposition A.1). Assumption 3 limits the interdependence between individuals’ actions to a tractable degree. If $W$ is specified as a binary indicator matrix such that $w_{ij} = 1$ if and only if individuals $i$ and $j$ are directly connected, then Assumption 3 requires the number of every individual’s direct connections to be bounded.

Let $F_k = E(JZ_k)$ and $F = \text{diag}(F_k)$. Following the discussion in Section 3.2, identification can be achieved based on linear moment conditions alone if $\bar{Q}'F_k$ has full column rank for large enough $n$, for $k = 1, \ldots, m$. If we consider the (infeasible) optimal IV matrix for the whole system given by $(\Sigma^{-1} \otimes I_n)F$, this sufficient identification can be written as follows.

Assumption 5. $\lim_{n \to \infty} n^{-1}F'(\Sigma^{-1} \otimes I_n)F$ is finite and nonsingular.

Finally, for asymptotic efficiency, Assumption 6 assumes that the (infeasible) optimal IV matrix $F_k$ can be approximated by some linear combination of the columns of the feasible IV matrix $\bar{Q}$, with an approximation error diminishes as the number of IVs increases with the sample size. This assumption is commonly made in the many-instrument literature (see, e.g., Donald and Newey, 2001; Hansen et al., 2008).

Assumption 6. For each $q$, there exists a $q \times m$ constant matrix $C_{k,q}$ such that $||F_k - \bar{Q}C_{k,q}||^2 \to 0$ as $q \to \infty$, for $k = 1, \ldots, m$.

Under the above regularity assumptions, the following proposition establishes the consistency and asymptotic normality of the fixed-effect 3SLS estimator. Let $V_k = JZ_k - F_k$ and $V = \text{diag}(V_k)$. 

28
Proposition 4.1. Under Assumptions 1-6, if \( q/n \to 0 \), then

\[
\sqrt{n}(\hat{\theta}_{3sls} - \theta_0 - b_{3sls}) \overset{d}{\to} N(0, \lim_{n \to \infty} n^{-1} F^t (\Sigma^{-1} \otimes I_n) F^{-1})
\]

where \( b_{3sls} = [Z'(\Sigma^{-1} \otimes \bar{P})Z]^{-1} E[V'(\Sigma^{-1} \otimes \bar{P})u] = O_p(q/n) \).

From Proposition 4.1, we can see that, when the number of IVs \( q \) grows at a rate slower than the sample size \( n \), the fixed-effect 3SLS estimator is consistent and asymptotically normal. However, the asymptotic distribution of the fixed-effect 3SLS estimator may not center around \( \theta_0 \) due to the presence of the many-instrument bias (see, e.g., Bekker, 1994). If \( q^2/n \to 0 \), then \( \sqrt{n} b_{3sls} = o_p(1) \) and the fixed-effect 3SLS estimator is properly centered.

When \( W_g \) has non-constant row sums for \( g = 1, \cdots, G \), the column rank of \( JWL \) is \( G \). If \( JWL \) is included in the IV matrix \( Q \), then \( q/n \to 0 \) implies \( G/n = 1/\bar{n} \to 0 \), where \( \bar{n} = n/G \) is the average group size. So for the fixed-effect 3SLS estimator to be consistent, the average group size needs to be large. On the other hand, \( q^2/n \to 0 \) implies \( G^2/n = G/\bar{n} \to 0 \). So for the fixed-effect 3SLS estimator to be properly centered, the average group size needs to be large relative to the number of groups.

The leading-order asymptotic bias of the fixed-effect 3SLS estimator given in Proposition 4.1 can be estimated to correct for the many-instrument bias. Suppose \( \sqrt{n} b_{3sls} \) is a consistent estimator of \( \sqrt{n} b_{3sls} \). The bias-corrected 3SLS (BC3SLS) estimator is given by \( \hat{\theta}_{bc3sls} = \hat{\theta}_{3sls} - \hat{b}_{3sls} \). From Proposition 4.1, if \( q/n \to 0 \) then

\[
\sqrt{n}(\hat{\theta}_{bc3sls} - \theta_0) \overset{d}{\to} N(0, \lim_{n \to \infty} n^{-1} F^t (\Sigma^{-1} \otimes I_n) F^{-1}).
\]

Example 5. Suppose \( m = 2 \) and \( K = 2 \) with \( X = [x_1, x_2] \). Consider the model

\[
\begin{align*}
    y_1 &= \phi_{21,0} y_2 + \lambda_{11,0} y_1 + \lambda_{21,0} y_2 + x_1 \gamma_{1,0} + \bar{x}_1 \gamma_{1,0} + L \eta_1 + u_1 \\
    y_2 &= \phi_{12,0} y_1 + \lambda_{12,0} y_1 + \lambda_{22,0} y_2 + x_2 \gamma_{2,0} + \bar{x}_2 \gamma_{2,0} + L \eta_2 + u_2.
\end{align*}
\]

(4.2)
Let

$$S = (1 - \phi_{12,0}\phi_{21,0})I_n - (\lambda_{11,0} + \lambda_{22,0} + \phi_{12,0}\lambda_{21,0} + \phi_{21,0}\lambda_{12,0})W + (\lambda_{11,0}\lambda_{22,0} - \lambda_{12,0}\lambda_{21,0})W^2. \quad (4.3)$$

The reduced-form equations of model (4.2) are

$$y_1 = E(y_1) + \epsilon_1 \quad \text{and} \quad y_2 = E(y_2) + \epsilon_2,$$

where

$$E(y_1) = S^{-1}[x_1\beta_{1,0} + Wx_1(\gamma_{1,0} - \lambda_{12,0}\beta_{1,0}) - W^2x_1\lambda_{22,0}\gamma_{1,0}$$

$$+ x_2\phi_{21,0}\beta_{2,0} + Wx_2(\lambda_{21,0}\beta_{2,0} + \phi_{21,0}\gamma_{2,0}) + W^2x_2\lambda_{21,0}\gamma_{2,0}$$

$$+ L(\eta_{1} + \phi_{21,0}\eta_{2}) + WL(\lambda_{21,0}\eta_{2} - \lambda_{22,0}\eta_{1})]$$

$$E(y_2) = S^{-1}[x_2\beta_{2,0} + Wx_2(\gamma_{2,0} - \lambda_{11,0}\beta_{2,0}) - W^2x_2\lambda_{11,0}\gamma_{2,0}$$

$$+ x_1\phi_{12,0}\beta_{1,0} + Wx_1(\lambda_{12,0}\beta_{1,0} + \phi_{12,0}\gamma_{1,0}) + W^2x_1\lambda_{12,0}\gamma_{1,0}$$

$$+ L(\eta_{2} + \lambda_{12,0}\eta_{1}) + WL(\lambda_{12,0}\eta_{1} - \lambda_{11,0}\eta_{2})]$$

and

$$\epsilon_1 = (I_n - \lambda_{22,0}W)S^{-1}u_1 + (\phi_{21,0}I_n + \lambda_{21,0}W)S^{-1}u_2$$

$$\epsilon_2 = (I_n - \lambda_{11,0}W)S^{-1}u_2 + (\phi_{12,0}I_n + \lambda_{12,0}W)S^{-1}u_1.$$

Let $Z_1 = [y_2, \bar{y}_1, \bar{y}_2, x_1, \bar{x}_1]$ and $Z_2 = [y_1, \bar{y}_1, \bar{y}_2, x_2, \bar{x}_2]$. Then,

$$F_1 = E(JZ_1) = J[E(y_2), WE(y_1), WE(y_2), x_1, \bar{x}_1]$$

$$F_2 = E(JZ_2) = J[E(y_1), WE(y_1), WE(y_2), x_2, \bar{x}_2]$$
and

\[ V_1 = JZ_1 - F_1 = J[\epsilon_2, W\epsilon_1, W\epsilon_2, 0_{n\times 2}] \]
\[ V_2 = JZ_2 - F_2 = J[\epsilon_1, W\epsilon_1, W\epsilon_2, 0_{n\times 2}] . \]

When \( W \) has non-constant row sums, Liu (2014) provides sufficient conditions for \( F_1 \) and \( F_2 \) to have full column ranks. From Proposition 4.1, the leading-order asymptotic bias of the fixed-effect 3SLS estimator is given by

\[ b_{3sels} = [Z'(\Sigma^{-1} \otimes \hat{P})Z]^{-1}E[V'(\Sigma^{-1} \otimes \hat{P})u] \]

with

\[ E[V'(\Sigma^{-1} \otimes \hat{P})u] = \begin{bmatrix}
\phi_{12,0} \text{tr} (\hat{P}S^{-1}) + \lambda_{12,0} \text{tr} (\hat{P}WS^{-1}) \\
\text{tr}(\hat{P}WS^{-1}) - \lambda_{22,0} \text{tr} (\hat{P}W^2S^{-1}) \\
\phi_{12,0} \text{tr} (\hat{P}WS^{-1}) + \lambda_{12,0} \text{tr} (\hat{P}W^2S^{-1}) \\
\text{tr}(\hat{P}WS^{-1}) - \lambda_{11,0} \text{tr} (\hat{P}W^2S^{-1}) \\
0_{2 \times 1}
\end{bmatrix}. \]

(4.5)

Let \( \hat{E}[V'(\Sigma^{-1} \otimes \hat{P})u] \) denote the estimated \( E[V'(\Sigma^{-1} \otimes \hat{P})u] \) with \( \phi_{lk,0} \)’s and \( \lambda_{lk,0} \)’s in (4.5) replaced by their \( \sqrt{n} \)-consistent preliminary estimates.\(^\text{16}\) Liu (2014) shows that, if \( q/n \to 0 \), the BC3SLS estimator given by \( \hat{\theta}_{bc3sels} = \hat{b}_{3sels} - \hat{b}_{3sels} \), where \( \hat{b}_{3sels} = [Z'(\hat{\Sigma}^{-1} \otimes \hat{P})Z]^{-1}\hat{E}[V'(\Sigma^{-1} \otimes \hat{P})u] \), has an asymptotically normal distribution around \( \theta_0 \).

\[ \square \]

4.2 Fixed-effect GMM estimator

Now, we consider the GMM estimator with both linear and quadratic moment functions \( g(\theta) = [g_1(\theta)', g_2(\theta)']' \) given by (3.3). The matrices \( \Xi_r \)’s used to construct quadratic moment functions satisfy the following regularity condition.

\(^\text{16}\)For example, \( \Phi_0 \) and \( \Lambda_0 \) can be consistently estimated by a less efficient equation-by-equation 2SLS estimator with a fixed number of IVs \( \hat{Q} = [JX, JWX, JW^2X, JW^3X] \).

31
Assumption 7. Let $\tilde{\Xi}_r = J \Xi_r J - \text{tr}(J \Xi_r J)/\text{tr}(J)$, where $\Xi_r$ is an $n \times n$ matrix with $\text{tr}(\Xi_r) = 0$ for $r = 1, \cdots, p$. The row and column sums of $\Xi_r$'s are uniformly bounded in absolute value.

As discussed in Section 3.2.3, identification of the simultaneous network model can be achieved if the “pseudo” reduced-form parameters can be identified in the multivariate network model and the structural parameters can be identified from the “pseudo” reduced-form parameters through proper exclusion restrictions. Let $F_k^*$ be a matrix containing all the linearly independent columns of $F_k$ and $F^* = \text{diag}(F_k^*)$. The sufficient identification condition is summarized in the following assumption.

Assumption 5'. (i) $\lim_{n \to \infty} n^{-1} F^*(\Sigma^{-1} \otimes I_n) F^*$ is finite and nonsingular. (ii) The simultaneous network model has a “pseudo” reduced-form model in the form of (3.15). The “pseudo” reduced-form model satisfies either condition in Proposition 3.2, and the structural parameters can be identified from the “pseudo” reduced-form parameters.

Let $\omega = [\text{vec}_D(\tilde{\Xi}_1), \cdots, \text{vec}_D(\tilde{\Xi}_p)]$, and let $\mu_3$ be an $m \times m^2$ matrix with its $(k, l)$-th element being $E(u_{ik}u_{il}/m)_{u_{i,l-(l/m)-1}m}$.\(^\dagger\) Then,

$$
\Omega_g = \begin{bmatrix}
\Omega_{11} & \Omega_{12} \\
\Omega_{12}' & \Omega_{22}
\end{bmatrix}
$$

where $\Omega_{11} = \text{Var}[g_1(\theta_0)] = \Sigma \otimes (\tilde{Q}' \tilde{Q})$, $\Omega_{12} = E[g_1(\theta_0)g_2(\theta_0)'] = \mu_3 \otimes (\tilde{Q}' \omega)$, and $\Omega_{22} = \text{Var}[g_2(\theta_0)]$. As the dimensions of $g(\theta)$ and $\Omega_g$ increase with the number of IVs $q$, the limits of $g(\theta)$ and $\Omega_g$ are not well defined. To study the asymptotic properties of the fixed-effect GMM estimator, we rewrite the (infeasible) GMM objective function as

$$
g(\theta)' \Omega_g^{-1} g(\theta) = g_1(\theta)' \Omega_{11}^{-1} g_1(\theta) + g_2(\theta)' \Omega_{22}^{-1} g_2(\theta) = u(\theta)'(\Sigma^{-1} \otimes \tilde{P})u(\theta) + g_2(\theta)' \Omega_{22}^{-1} g_2(\theta),
$$

where $g_2(\theta) = g_2(\theta) - \Omega_{12} \Omega_{11}^{-1} g_1(\theta) = g_2(\theta) - [(\mu_3 \Sigma^{-1}) \otimes (\omega' \tilde{P})]u(\theta)$ and $\Omega_{22} = \text{Var}[g_2^*(\theta)] = \Omega_{22} - \Omega_{12} \Omega_{11}^{-1} \Omega_{12} = \Omega_{22} - (\mu_3 \Sigma^{-1} \mu_3) \otimes (\omega' \tilde{P} \omega)$. As $E[g_1(\theta)g_2^*(\theta)'] = 0$, $g_2^*(\theta)$ can be considered

\(^\dagger\)The ceiling function $[x]$ gives the smallest integer that is no less than $x$. 

32
as the projection residual of \( g_2(\theta) \) on \( g_1(\theta) \). The following two assumptions are standard for nonlinear GMM estimators.

**Assumption 8.** (i) \( \lim_{n \to \infty} n^{-1} \omega^F \) is a finite matrix with full column rank, for \( k = 1, \ldots, m \).

(ii) \( \lim_{n \to \infty} n^{-1} \Omega^{*}_{22} \) is finite and nonsingular.

**Assumption 9.** \( \theta_0 \) is in the interior of a compact and convex parameter space.

Let \( \hat{\Sigma} \) and \( \hat{\mu}_3 \) be \( \sqrt{n} \)-consistent estimators of \( \Sigma \) and \( \mu_3 \) respectively. Let \( \hat{g}_2^*(\theta) = g_2(\theta) - ([\hat{\mu}_3' \hat{\Sigma}^{-1}] \otimes (\omega' \hat{P}))u(\theta) \) and \( \hat{\Omega}^*_{22} = \hat{\Omega}_{22} - ([\hat{\mu}_3' \hat{\Sigma}^{-1} \hat{\mu}_3] \otimes (\omega' \hat{P}) \omega) \), where \( n^{-1} \hat{\Omega}_{22} \) is a \( \sqrt{n} \)-consistent estimator of \( n^{-1} \Omega_{22} \). The fixed-effect GMM estimator is given by

\[
\hat{\theta}_{gmm} = \arg \min_{\theta} u(\theta)' (\hat{\Sigma}^{-1} \otimes \hat{\Omega}) u(\theta) + \hat{g}_2^*(\theta)' \hat{\Omega}^*_{22} \hat{g}_2^*(\theta).
\]

The next proposition establishes the consistency of the fixed-effect GMM estimator.

**Proposition 4.2.** Under Assumptions 1-4,5',6-9, if \( q/n \to 0 \), then the fixed-effect GMM estimator \( \hat{\theta}_{gmm} \) is consistent.

Let \( \Upsilon_{1,kl} = [E(V_k' \hat{\Xi}_1 u_k), \ldots, E(V_k' \hat{\Xi}_p u_k)]', \Upsilon_{2,kl} = [E(V_k' \hat{\Xi}_1 u_k), \ldots, E(V_k' \hat{\Xi}_p u_k)]', \)

\[
D_2 = -E \left[ \frac{\partial}{\partial \theta} g_2(\theta_0) \right] = \begin{bmatrix} \Upsilon_{1,11} \\
\vdots \\
\Upsilon_{1,1m} \\
\Upsilon_{1,1m} \\
\vdots \\
\Upsilon_{1,mm} \end{bmatrix}, \quad \text{and} \quad \hat{D}_2^* = -E \left[ \frac{\partial}{\partial \theta} g_2^*(\theta_0) \right] = D_2 - ([\mu_3' \Sigma^{-1}] \otimes \omega') \cdot \hat{F}.
\]

**Assumption 10.** \( \lim_{n \to \infty} n^{-1} \hat{D}_2^* \) is a finite matrix with full column rank.
The following proposition gives the asymptotic distribution of the fixed-effect GMM estimator.

**Proposition 4.3.** Under Assumptions 1-4, 5’, 6-10, if \( q^{3/2}/n \to 0 \), then

\[
\sqrt{n}(\hat{\theta}_{gmm} - \theta_0 - b_{gmm}) \xrightarrow{d} N(0, \lim_{n \to \infty} n[F'((\Sigma^{-1} \otimes I_n))F + D_2^*\Omega_{22}^{-1}D_2^*]^{-1}),
\]

where \( b_{gmm} = [F'((\Sigma^{-1} \otimes I_n))F + D_2^*\Omega_{22}^{-1}D_2^*]^{-1}E[V'((\Sigma^{-1} \otimes \bar{P}))u] = O(q/n) \).

The asymptotic covariance matrix of the GMM estimator can be compared with that of the 3SLS estimator. As \( D_2^*\Omega_{22}^{-1}D_2^* \) is positive semi-definite, the asymptotic covariance of \( \hat{\theta}_{gmm} \) is smaller than that of \( \hat{\theta}_{3sls} \). Thus, the additional quadratic moment functions improve the asymptotic efficiency.

Similar to the 3SLS estimator, the leading-order asymptotic bias of the GMM estimator given in Proposition 4.3 can be estimated to correct for the many-instrument bias. Suppose \( \sqrt{n}b_{gmm} \) is a consistent estimator of \( \sqrt{n}b_{gmm} \). The bias-corrected GMM (BCGMM) estimator is given by \( \hat{\theta}_{bcgmm} = \hat{\theta}_{gmm} - \hat{b}_{gmm} \). From Proposition 4.3, if \( q^{3/2}/n \to 0 \) then

\[
\sqrt{n}(\hat{\theta}_{bcgmm} - \theta_0) \xrightarrow{d} N(0, \lim_{n \to \infty} n[F'((\Sigma^{-1} \otimes I_n))F + D_2^*\Omega_{22}^{-1}D_2^*]^{-1}).
\]

**Example 6 (continued).** Model (4.2) has “pseudo” reduced-form equations

\[
\begin{align*}
\mathbf{y}_1 &= \lambda^*_{11,0} \bar{y}_1 + \lambda^*_{21,0} \bar{y}_2 + \mathbf{X}\beta^*_{1,0} + \bar{\mathbf{X}}\gamma^*_{1,0} + \mathbf{L}\eta^*_{1} + \mathbf{u}_{1}^* \\
\mathbf{y}_2 &= \lambda^*_{12,0} \bar{y}_1 + \lambda^*_{22,0} \bar{y}_2 + \mathbf{X}\beta^*_{2,0} + \bar{\mathbf{X}}\gamma^*_{2,0} + \mathbf{L}\eta^*_{2} + \mathbf{u}_{2}^*,
\end{align*}
\]

where

\[
\begin{bmatrix}
\lambda^*_{11,0} & \lambda^*_{12,0} \\
\lambda^*_{21,0} & \lambda^*_{22,0}
\end{bmatrix} = (1 - \phi_{12,0}\phi_{21,0})^{-1}
\begin{bmatrix}
\lambda_{11,0} + \phi_{21,0}\lambda_{12,0} & \lambda_{12,0} + \phi_{12,0}\lambda_{11,0} \\
\lambda_{21,0} + \phi_{21,0}\lambda_{22,0} & \lambda_{22,0} + \phi_{12,0}\lambda_{21,0}
\end{bmatrix}
\]
and

\[
[\beta^*_1, \beta^*_2] = (1 - \phi_{12,0}\phi_{21,0})^{-1} \begin{bmatrix} \beta_{1,0} & \phi_{12,0}\beta_{1,0} \\ \phi_{21,0}\beta_{2,0} & \beta_{2,0} \end{bmatrix}
\]

(4.8)

\[
[\gamma^*_1, \gamma^*_2] = (1 - \phi_{12,0}\phi_{21,0})^{-1} \begin{bmatrix} \gamma_{1,0} & \phi_{12,0}\gamma_{1,0} \\ \phi_{21,0}\gamma_{2,0} & \gamma_{2,0} \end{bmatrix}
\]

(4.9)

Suppose the multivariate network model (4.6) satisfies the identification conditions in Proposition 3.2. Then, the structural parameters

\[
\Theta_0 = [(I_m - \Phi_0)', -A_0', -B_0', -\Gamma_0']' = -\begin{bmatrix} -1 & \phi_{21,0} & \lambda_{11,0} & \lambda_{21,0} & \beta_{1,0} & 0 & \gamma_{1,0} & 0 \\ \phi_{12,0} & -1 & \lambda_{12,0} & \lambda_{22,0} & 0 & \beta_{2,0} & 0 & \gamma_{2,0} \end{bmatrix}'
\]

can be identified from the “pseudo” reduced-form parameters through (4.7)-(4.9) if the rank condition holds. The exclusion restriction for the first equation of model (4.2) can be represented by \(R_1 = [0, 0, 0, 0, 0, 1, 0, 1]\). Then \(R_1\Theta_0 = [0, -\beta_{2,0} - \gamma_{2,0}]\), which has rank 1 if \(\beta_{2,0} + \gamma_{2,0} \neq 0\). Similarly, the exclusion restriction for the second equation can be represented by \(R_2 = [0, 0, 0, 1, 0, 1, 0, 1]\). Then \(R_2\Theta_0 = [-\beta_{1,0} - \gamma_{1,0}, 0]\) which has rank 1 if \(\beta_{1,0} + \gamma_{1,0} \neq 0\).

For the GMM estimation, the linear moment functions are given by

\[
g_1(\theta) = (I_2 \otimes \tilde{Q})'[u_1(\theta_1)', u_2(\theta_2)']',
\]

where \(u_k(\theta_k) = y_k - Z_k\theta_k\) for \(k = 1, 2\). The quadratic moment functions are given by

\[
g_2(\theta) = [g_{2,11}(\theta_1, \theta_1)', g_{2,12}(\theta_1, \theta_2)', g_{2,21}(\theta_2, \theta_1)', g_{2,22}(\theta_2, \theta_2)']',
\]

where \(g_{2,kl}(\theta_k, \theta_l) = [\tilde{\Xi}_1' u_k(\theta_k), \ldots, \tilde{\Xi}_p' u_k(\theta_k)]'u_l(\theta_l)\) for \(k, l = 1, 2\). Let \(\mu_{s,t} = E(u_{s1}'u_{t2})\), for
\[ \Omega_{22} = K_4 \otimes (\omega' \omega) \]

\[
= \begin{bmatrix}
\sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\
* & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{22} \\
* & * & \sigma_{12}^2 & \sigma_{12}\sigma_{22} \\
* & * & * & \sigma_{22}^2 
\end{bmatrix} \otimes \Delta_1 + \begin{bmatrix}
\sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\
* & \sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{22} & \sigma_{12}\sigma_{22} \\
* & * & \sigma_{12}\sigma_{22} & \sigma_{12}\sigma_{22} \\
* & * & * & \sigma_{22}^2 
\end{bmatrix} \otimes \Delta_2
\]

(4.10)

where \( \omega = [\text{vec}_D(\hat{\tau}_1), \ldots, \text{vec}_D(\hat{\tau}_p)] \),

\[
K_4 = \begin{bmatrix}
\mu_{4,0} - 3\sigma_{11}^2 & \mu_{3,1} - 3\sigma_{11}\sigma_{12} & \mu_{3,1} - 3\sigma_{11}\sigma_{12} & \mu_{2,2} - \sigma_{11}\sigma_{22} - 2\sigma_{12}^2 \\
* & \mu_{2,2} - \sigma_{11}\sigma_{12} - 2\sigma_{12}^2 & \mu_{2,2} - \sigma_{11}\sigma_{12} - 2\sigma_{12}^2 & \mu_{1,3} - 3\sigma_{12}\sigma_{22} \\
* & * & \mu_{2,2} - \sigma_{11}\sigma_{22} - 2\sigma_{12}^2 & \mu_{1,3} - 3\sigma_{12}\sigma_{22} \\
* & * & * & \mu_{0,4} - 3\sigma_{22}^2 
\end{bmatrix}
\]

\[
\Delta_1 = \begin{bmatrix}
\text{tr}(\hat{\tau}_1\hat{\tau}_1) & \cdots & \text{tr}(\hat{\tau}_1\hat{\tau}_p) \\
\vdots & \ddots & \vdots \\
\text{tr}(\hat{\tau}_p\hat{\tau}_1) & \cdots & \text{tr}(\hat{\tau}_p\hat{\tau}_p) 
\end{bmatrix}, \quad \text{and} \quad \Delta_2 = \begin{bmatrix}
\text{tr}(\hat{\tau}_1\hat{\tau}_1') & \cdots & \text{tr}(\hat{\tau}_1\hat{\tau}_p') \\
\vdots & \ddots & \vdots \\
\text{tr}(\hat{\tau}_p\hat{\tau}_1') & \cdots & \text{tr}(\hat{\tau}_p\hat{\tau}_p') 
\end{bmatrix}
\]

Let \( \hat{g}_2^*(\theta) = g_2(\theta) - [(\hat{\mu}_g^T \hat{\Sigma}^{-1}) \otimes (\omega' \hat{P})]u(\theta) \) and \( \hat{\Omega}_{22} = \hat{\Omega}_{22} - (\hat{\mu}_g^T \hat{\Sigma}^{-1} \hat{\mu}_g) \otimes (\omega' \hat{P} \omega) \), where

\[
\hat{\Sigma} = \begin{bmatrix}
\sigma_{11} & \sigma_{12} \\
\sigma_{12} & \sigma_{22} 
\end{bmatrix} \quad \text{and} \quad \hat{\mu}_3 = \begin{bmatrix}
\hat{\mu}_{3,0} & \hat{\mu}_{2,1} & \hat{\mu}_{1,2} \\
\hat{\mu}_{2,1} & \hat{\mu}_{1,2} & \hat{\mu}_{0,3} 
\end{bmatrix}
\]

are \( \sqrt{n} \)-consistent estimators of \( \Sigma \) and \( \mu_g \), and \( \hat{\Omega}_{22} \) is the estimated \( \Omega_{22} \) with \( \mu_{s,t} \)'s and \( \sigma_{s,t} \)'s in (4.10) replaced by their \( \sqrt{n} \)-consistent estimators. Then, it follows from Proposition 4.3 that, if \( q^{3/2}/n \to 0 \), the fixed-effect GMM estimator \( \hat{\theta}_{gmm} = \arg \min u(\theta)'(\hat{\Sigma}^{-1} \otimes \hat{P})u(\theta) + \hat{g}_2^*(\theta)' \hat{\Omega}_{22}^{-1} \hat{g}_2^*(\theta) \) is asymptotically normal with an asymptotic bias

\[
b_{gmm} = [F'(\Sigma^{-1} \otimes I_n)F + D_2^T \hat{\Omega}_{22}^{-1} D_2]^{-1}E[V'(\Sigma^{-1} \otimes \hat{P})u], \quad (4.11)
\]

36
where $E[V'(\Sigma^{-1} \otimes \tilde{P})u]$ is given by (4.5). Let $\tilde{D}_2 = \tilde{D}_2 - [(\tilde{\mu}_3' \tilde{\Sigma}^{-1}) \otimes \omega' \tilde{P}]Z$, where $n^{-1}\tilde{D}_2$ is a consistent estimator of $n^{-1}D_2$.\(^{18}\) If follows by Lemma C.1 in the appendix that, if $q^{3/2}/n \to 0$, the BCGMM estimator $\tilde{\theta}_{bcgmm} = \tilde{\theta}_{gmm} - \tilde{b}_{gmm}$, where

$$
\tilde{b}_{gmm} = [Z'(\Sigma^{-1} \otimes \tilde{P})Z + \tilde{D}_2' \tilde{\Omega}_2^{-1} \tilde{D}_2]^{-1} E[V'(\Sigma^{-1} \otimes \tilde{P})u], 
$$

has an asymptotically normal distribution around $\theta_0$.\(^{19}\)

## 5 Monte Carlo Experiments

To investigate the finite sample performance of the proposed GMM estimator, we conduct a limited simulation study based on model (4.2) considered in Example 5. The DGP of the Monte Carlo experiment follows that in Liu (2014). Specifically, the adjacency matrix $W_g$ is generated as follows. First, for the $i$-th row of $W_g$, we generate an integer $c_{g,i}$ uniformly at random from the set of integers $\{1, 2, 3\}$. Then, if $i + c_{g,i} \leq n_g$, we set the $(i+1)$-th, $\ldots$, $(i+c_{g,i})$-th elements of the $i$-th row of $W_g$ to be ones and the other elements in that row to be zeros; otherwise, the elements of ones will be wrapped around such that the first $(i + c_{g,i} - n_g)$ elements of the $i$-th row will be ones. We experiment with different numbers of groups $G$ and different group sizes $n_g$.

We conduct 1000 repetitions for each specification in this Monte Carlo experiment. In each repetition, $x_k$ is generated from $N(0, I_n)$ and $\eta_k$ is generated from $N(0, I_G)$ for $k = 1, 2$. The error term $u = (u_1', u_2')'$ is generated from $N(0, \Sigma \otimes I_n)$. We set $\sigma_1^2 = \sigma_2^2 = 1$, $\phi_{21,0} = \phi_{12,0} = 0.2$, and $\lambda_{11,0} = \lambda_{21,0} = \lambda_{12,0} = \lambda_{22,0} = 0.1$.\(^{20}\) We experiment with different values for $\sigma_1$, $(\beta_{1,0}, \beta_{2,0})$ and $(\gamma_{1,0}, \gamma_{2,0})$.

We consider the following estimators in the experiment. (i) 3SLS-1: the 3SLS estimator with the IV matrix $\tilde{Q}_1 = [JX, JW^2X]$, where $X = [x_1, x_2]$; (ii) 3SLS-2: the 3SLS estimator with the IV matrix $\tilde{Q}_2 = [\tilde{Q}_1, JW L]$; (iii) BC3SLS: the bias-corrected 3SLS-2; (iv) GMM-1: the GMM estimator with the IV matrix $\tilde{Q}_1$ and quadratic moment functions $g_2(\theta) =$

\(^{18}\)The explicit expression of $n^{-1}D_2$ and its estimator can be found in the proof of Lemma C.1.

\(^{19}\)When $u = (u_1', u_2')' \sim N(0, \Sigma \otimes I_n)$, the above GMM estimator can be simplified as $\mu_3 = 0$ and $K_4 = 0$.

\(^{20}\)The matrix $S$ given by (4.3) is invertible with the chosen parameters.
\[u_1(\theta_1)\varepsilon_1 u_1(\theta_1), u_2(\theta_2)\varepsilon_2 u_2(\theta_2), u_1(\theta_1)\varepsilon_1^\prime u_2(\theta_2), u_2(\theta_2)\varepsilon_2^\prime u_2(\theta_2)]',\] where \(\tilde{\Xi} = JWJ – \text{tr}(JWJ)/\text{tr}(J)\);

(v) GMM-2: the GMM estimator with the IV matrix \(\tilde{\Phi}_2\) and the same set of quadratic moment functions used by GMM-1; and (vi) BCGMM: the bias-corrected GMM-2. The IV matrix \(\tilde{\Phi}_1\) is based on the exogenous attributes of direct and indirect connections. \(\tilde{\Phi}_2\) also uses additional IVs \(JWL\) based on the numbers of (direct) connections to improve estimation efficiency. As \(WL\) has \(G\) columns, the number of IVs in \(\tilde{\Phi}_2\) increases with the number of groups.

[Insert Tables 1-6 here]

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). Due to symmetry of the two equations in model (4.2), we only report the estimation result for the first equation of model (4.2) in Tables 1-6. The main findings from the simulation experiment are summarized as follows.

(a) The quadratic moment conditions improves the estimation efficiency of the peer effect parameters. When \(\beta_{1,0} = \beta_{2,0} = \gamma_{1,0} = \gamma_{2,0} = 0.6\) and the correlation across equations is moderate \((\sigma_{12} = 0.5)\), for the sample with \(n_g = 10\) and \(G = 30\) in Table 1, GMM-1 reduces the SD of 3SLS-1 estimates of \(\lambda_{11,0}\) and \(\lambda_{21,0}\) by 11.4\% and 13.3\% respectively. The efficiency improvement is more significant when the IV matrix \(\tilde{\Phi}_1\) is less informative. When \(\beta_{1,0} = \beta_{2,0} = \gamma_{1,0} = \gamma_{2,0} = 0.3\) and the correlation across equations is moderate \((\sigma_{12} = 0.5)\), for the sample with \(n_g = 10\) and \(G = 30\) in Table 4, GMM-1 reduces the SD of 3SLS-1 estimates of \(\lambda_{11,0}\) and \(\lambda_{21,0}\) by 32.3\% and 35.8\% respectively.

(b) The additional IVs \(JWL\) in \(\tilde{\Phi}_2\) also improve the estimation efficiency of the peer effect parameters. When \(\beta_{1,0} = \beta_{2,0} = \gamma_{1,0} = \gamma_{2,0} = 0.6\) and the correlation across equations is moderate \((\sigma_{12} = 0.5)\), for the sample with \(n_g = 10\) and \(G = 30\) in Table 1, GMM-2 reduces the SD of GMM-1 estimates of \(\lambda_{11,0}\) and \(\lambda_{21,0}\) by 15.4\% and 17.9\% respectively. The efficiency improvement is more significant when the IV matrix \(\tilde{\Phi}_1\) is less informative. When \(\beta_{1,0} = \beta_{2,0} = \gamma_{1,0} = \gamma_{2,0} = 0.3\) and the correlation across equations is moderate \((\sigma_{12} = 0.5)\), for the sample with \(n_g = 10\) and \(G = 30\)
in Table 4, GMM-2 reduces the SD of GMM-1 estimates of $\lambda_{11,0}$ and $\lambda_{21,0}$ by 30.8% and 27.9% respectively.

(c) The additional IVs $JWL$ in $Q_2$ introduce biases into the estimators. The size of the bias increases as the correlation across equations $\sigma_{12}$ increases and as $Q_1$ becomes less informative. The size of the bias reduces as the network size increases. The proposed bias-correction procedure substantially reduces the bias. When the sample size is relatively large (Tables 3 and 6), the bias corrected estimates are essentially unbiased.

6 Summary and Future Work

In the literature of linear social interaction models, identification of peers effects is usually achieved through (i) intransitivity in network connections (e.g., Bramoullé et al., 2009; Blume et al., 2013; Cohen-Cole et al., 2014), (ii) variation in Bonacich centrality when the adjacency matrix has non-constant row sums (e.g., Liu and Lee, 2010; Liu et al., 2014; Liu, 2014), and/or (iii) variation in group sizes (e.g., Lee, 2007a; Graham, 2008; Davezies et al., 2009). In this paper, we propose a new set of quadratic moment conditions based on the covariance structure of the simultaneous equations network model and bring some new insight into the identification of peer effects. We give examples to show that the quadratic moment conditions of model disturbances could be useful for identifying peer effects when the above mentioned identification strategies fail. Furthermore, we develop a general GMM framework for the estimation of a system of simultaneous equations with social interactions. The proposed GMM estimator improves the asymptotic efficiency of the IV-based linear estimators, and performs well in the Monte Carlo experiment.

Some possible extensions of the current work are in order. First, different individuals may participate in different activities. Therefore, it would be interesting to study the sample selection issue (Heckman, 1976) in the context of social networks and multivariate choices. Second, people may form different social networks for different activities they participate. Hence, another thread of future research could be to consider activity-specific networks and to study the formation and evolution of such networks and associated identification problems.
References


A Microfoundation

Suppose a set of individuals \{1, \cdots, n\} interacts in \(m\) activities within a network captured by an adjacency matrix \(W = [w_{ij}]\). Let \(y_k = (y_{1k}, \cdots, y_{nk})'\), where \(y_{ik}\) is individual \(i\)'s choice of action (from a continuous action space) in activity \(k\). The utility of individual \(i\) is a linear-quadratic function of her and her peers’ actions in the \(m\) activities:

\[
U_i(y_1, \cdots, y_m; W) = \sum_{k=1}^{m} \left( \varphi_{ik} \sum_{l=1}^{m} \vartheta_{lk} \sum_{j=1}^{n} w_{ij} y_{jl} \right) y_{ik} - \frac{1}{2} \sum_{k=1}^{m} \sum_{l=1}^{m} \varphi_{lk} y_{ik} y_{il}, \quad (A.1)
\]

where \(\varphi_{kl} = \varphi_{lk} \neq 0\) for all \(k\) and \(l\). As in the single-activity linear-quadratic utility function (Ballester et al., 2006), the utility (A.1) has two components: payoff and cost. The marginal payoff of individual \(i\)'s action in activity \(k\) depends on the (exogenous) productivity of individual \(i\) in activity \(k\) given by \(\varpi_{ik}\) and her peers’ actions in the \(m\) activities.\(^{21}\) The parameter \(\vartheta_{lk}\) captures the strategic complementarity or substitutability (depending on the sign of \(\vartheta_{lk}\)) between individual \(i\)'s own action in activity \(k\) and her peer’s action in activity \(l\). The marginal cost of individual \(i\)'s action in activity \(k\) depends on her actions in the \(m\) activities. The parameter \(\varphi_{lk} (k \neq l)\) measures the strategic complementarity or substitutability (depending on the sign of \(\varphi_{lk}\)) of an individual’s actions in activities \(k\) and \(l\).

Given \(W\), individuals simultaneously choose actions in the \(m\) activities to maximize their utilities.\(^{22}\) From the first order condition of utility maximization, the best response function of individual \(i\) in activity \(k\) is

\[
y_{ik} = \sum_{l=1, l\neq k}^{m} \phi_{lk} y_{il} + \sum_{l=1}^{m} \lambda_{lk} \sum_{j=1}^{n} w_{ij} y_{jl} + \pi_{ik},
\]

where \(\phi_{lk} = -\varphi_{lk}/\varphi_{kk}\) (for \(k \neq l\)), \(\lambda_{lk} = \vartheta_{lk}/\varphi_{kk}\), and \(\pi_{ik} = \varpi_{ik}/\varphi_{kk}\). We set \(\phi_{kk} = 0\) for all \(k\). In

\(^{21}\)Following Ballester et al. (2006), we consider a simultaneous move complete information game where \(\varpi_{ik}\) is common knowledge. By contrast, Blume et al. (2013) consider a single-activity incomplete information game where a component of \(\varpi_{ik}\) can only be privately observed by individual \(i\).

\(^{22}\)We consider a static network game with a predetermined network in this paper. Although not the focus of this paper, it is worth pointing out the modeling, identification and estimation of the network formation process has recently made some exciting progress (see, e.g., Goldsmith-Pinkham and Imbens, 2013; Mele, 2013; Graham, 2014).
matrix form, the best response function can be written as

$$y_k = \sum_{l=1, l \neq k}^m \phi_{lk} y_l + \sum_{l=1}^m \lambda_{lk} \bar{y}_l + \pi_k,$$  
(A.2)

where $\bar{y}_l = Wy_l$ and $\pi_k = (\pi_{1k}, \cdots, \pi_{nk})'$. Let $Y = [y_1, \cdots, y_m]$, $\bar{Y} = WY$, and $\Pi = [\pi_1, \cdots, \pi_m]$. For all the $m$ activities, it follows from (A.2) that

$$Y = Y\Phi + \bar{Y}\Lambda + \Pi,$$  
(A.3)

where $\Phi = [\phi_{lk}]$ and $\Lambda = [\lambda_{lk}]$ are $m \times m$ parameter matrices. Let $y = \text{vec}(Y)$ and $\pi = \text{vec}(\Pi)$. The following theorem characterizes the unique equilibrium of the network game.

**Proposition A.1.** If $(I_m - \Phi)$ is nonsingular and $\rho(\Lambda(I_m - \Phi)^{-1}) < 1/\rho(W)$, then the network game with utility (A.1) has a unique Nash equilibrium in pure strategies given by

$$y^* = (I_{mn} - \Phi' \otimes I_n - \Lambda' \otimes W)^{-1}\pi.$$

**Proof.** If $(I_m - \Phi)$ is nonsingular and $\rho(\Lambda(I_m - \Phi)^{-1}) < 1/\rho(W)$, then $(I_{mn} - \Phi' \otimes I_n - \Lambda' \otimes W)$ is nonsingular (Yang and Lee, 2014). The existence of a unique Nash equilibrium follows a similar argument as in the proof of Theorem 1 in Ballester et al. (2006). □

Note that, when $m = 1$, we have $\Phi = 0$ and $\Lambda = \lambda_{11}$. It follows from (A.2) that the best response function is $y_1 = \lambda_{11} Wy_1 + \pi_1$. Ballester et al. (2006) have shown that if $|\lambda_{11}| < 1/\rho(W)$ then the single-activity network game has a unique Nash equilibrium in pure strategies given by $y^*_1 = (I_n - \lambda_{11} W)^{-1}\pi_1$.

If the adjacency matrix $W$ is row-normalized such that each row of $W$ adds up to unity, then $\bar{y}_k = Wy_k$ represents peers’ average choice in activity $k$. Liu et al. (2014) show that the single-activity network game may have different equilibrium implications, depending on whether or not the adjacency matrix is row-normalized. The same intuition carries through to network games.

---

23 To be coherent with the econometric model developed in Section 2, we focus on the interior equilibrium of the game. In a recent paper, Bramoullé et al. (2014) characterize the interior and boundary equilibria of the single-activity network game.
with any finite number of activities. In particular, if individuals are ex-ante homogeneous such that \( \pi_k = c_k \epsilon_n \) for \( k = 1, \ldots, m \), where \( c_k \) is a finite constant scalar, then \( y_{i,k}^* = y_{j,k}^* \) for all \( i, j = 1, \ldots, n \) and \( k = 1, \ldots, m \) when \( W \) is row-normalized, while \( y_{i,k}^* \) and \( y_{j,k}^* \) may be different when \( W \) has varying row sums. The intuition is as follows. In the utility function (A.1), there are two layers of heterogeneity, namely, heterogeneity in productivity and heterogeneity in network centrality. When \( W \) is row-normalized, the heterogeneity in network centrality is washed out, and hence individuals will behave the same if they are homogeneous in productivity. On the other hand, when \( W \) has varying row sums, individuals may behave differently if they have different positions in the network even if they are ex-ante homogeneous in productivity.

To obtain the econometric model, we assume that the productivity matrix of the \( n \) individuals in the \( m \) activities is given by

\[
\Pi = X\beta + \Phi + U,
\]

where \( X \) is an \( n \times K \) matrix of observations on \( K \) exogenous variables, \( \Phi = WX \), and \( U = [u_1, \ldots, u_m] \) is an \( n \times m \) matrix of disturbances. Substitution of (A.4) into the best response function (A.3) gives the econometric model (2.1).

**B Identification by Quadratic Moment Conditions**

Consider the following model

\[
y = \lambda_0 Wy + X\delta_0 + \epsilon,
\]

where \( y \) is an \( n \times 1 \) vector of observations on the dependent variable, \( W \) is an \( n \times n \) nonstochastic adjacency matrix with a zero diagonal, \( X \) is \( n \times K \) matrix of observations on \( K \) (nonstochastic) exogenous regressors, and \( \epsilon = (\epsilon_1, \ldots, \epsilon_n)' \) is an \( n \times 1 \) vector of disturbances. It is easy to see that the \( k \)-th equation of the SUR network model (2.5) and the within-transformed model (3.6) can be written in the form of (B.1). Suppose the following regularity assumptions hold.

**A1.** \( \epsilon_i \sim i.i.d. (0, \sigma^2) \) and its moments of order higher than the fourth exist.

**A2.** The row and column sums of \( W \) and \( (I - \lambda_0 W)^{-1} \) are uniformly bounded in absolute value.
A3. \( X \) has full column rank for \( n \) sufficiently large. The elements of \( X \) are uniformly bounded.

A4. Let \( \Xi \) be an \( n \times n \) nonstochastic symmetric matrix with \( \text{tr}(\Xi) = 0 \). The row and column sums of \( \Xi \) are uniformly bounded in absolute value.

Lemma B.1. Suppose assumptions A1-A4 hold. Let \( \bar{y} = Wy \), \( G = W(I_n - \lambda_0W)^{-1} \), and \( M = I_n - X(X'X)^{-1}X' \). If \( \lim_{n \to \infty} n^{-1}\text{tr}(\Xi G) > 0 \), then the consistent estimator of \( \lambda_0 \) in (B.1) is
\[
\hat{\lambda} = \frac{y'\Xi M \bar{y} - [(y'\Xi M \bar{y})^2 - (y'\Xi M \bar{y})(y'\Xi M \bar{y})]^{1/2}}{y'\Xi M \bar{y}}.
\]
If \( \lim_{n \to \infty} n^{-1}\text{tr}(\Xi G) < 0 \), then the consistent estimator of \( \lambda_0 \) in (B.1) is
\[
\hat{\lambda} = \frac{y'\Xi M \bar{y} + [(y'\Xi M \bar{y})^2 - (y'\Xi M \bar{y})(y'\Xi M \bar{y})]^{1/2}}{y'\Xi M \bar{y}}.
\]

Proof of Lemma B.1. Note that the row and column sums of \( M \) and \( I_n - M \) are uniformly bounded in absolute value (Lee, 2004), and \( n^{-1}\text{tr}(A(I_n - M)) = O(K/n) = o(1) \) for any \( n \times n \) nonstochastic matrix \( A \) whose row and column sums are uniformly bounded in absolute value (see Lemma B.2 in Liu and Lee, 2010). The proof follows a similar argument as in the proof of Proposition 2.2 in Lee (2001).

C  Proofs

Proof of Propositions 2.1 and 3.1. We observe that the \( r \)-th element of \( E[f_{2,kl}(\theta_k, \theta_l)] \) is
\[
E[u_k(\theta_k)'\Xi, u_l(\theta_l)] = d_k(\theta_k)'\Xi d_l(\theta_l) + \sigma_{kl}[(\lambda_{kk,0} - \lambda_{kk})\text{tr}(\Xi'G_k) + (\lambda_{ll,0} - \lambda_{ll})\text{tr}(\Xi'G_l) + (\lambda_{kk,0} - \lambda_{ll})(\lambda_{ll,0} - \lambda_{ll})\text{tr}(G_k'\Xi G_l)].
\]

With \( \beta_k \) and \( \gamma_k \) given by (2.8), \( d_k(\theta_k) = 0 \). Proposition 2.1 follows the discussion in Section 2.2.1 and Theorem 2.1 of Hansen (1982). Proposition 3.1 follows by a similar argument.

Proof of Propositions 2.2 and 3.2. Suppose, for some \( 0 \leq \bar{m} \leq m-1 \), \( E(\tilde{y}_1) \) and \( [E(\tilde{y}_1), \cdots, E(\tilde{y}_{\bar{m}}), X, \tilde{X}] \)
are linearly dependent such that $E(\tilde{y}_l) = b_{l,1}E(\tilde{y}_1) + \cdots + b_{l,m}E(\tilde{y}_m) + Xc_{l,1} + \tilde{X}c_{l,2}$, for $l = m+1, \ldots, m$, where $b_{l,1}, \ldots, b_{l,m}$ are constant scalars and $c_{l,1}, c_{l,2}$ are $K \times 1$ constant vectors. In this case,

$$d_k(\theta_k) = E(\tilde{y}_1)[\sum_{l=m+1}^m(\lambda_{lk,0} - \lambda_{lk})b_{l,1} + (\lambda_{lk,0} - \lambda_{lk})] + \cdots + E(\tilde{y}_m)[\sum_{l=m+1}^m(\lambda_{lk,0} - \lambda_{lk})b_{l,m} + (\lambda_{lk,0} - \lambda_{lk})] + X[\sum_{l=m+1}^m(\lambda_{lk,0} - \lambda_{lk})c_{l,1} + (\beta_{lk,0} - \beta_{lk})] + \tilde{X}[\sum_{l=m+1}^m(\lambda_{lk,0} - \lambda_{lk})c_{l,2} + (\gamma_{lk,0} - \gamma_{lk})]$$

and the solutions of $\lim_{n \to \infty} n^{-1} \mathbf{Q}'d_k(\theta_k) = 0$ are characterized by

$$\lambda_{lk} = \sum_{l=m+1}^m(\lambda_{lk,0} - \lambda_{lk})b_{l,1} + (\lambda_{lk,0} - \lambda_{lk}), \quad (C.1)$$

$$\lambda_{lk} = \sum_{l=m+1}^m(\lambda_{lk,0} - \lambda_{lk})b_{l,m} + (\lambda_{lk,0} - \lambda_{lk}),$$

$$\beta_{lk} = \sum_{l=m+1}^m(\lambda_{lk,0} - \lambda_{lk})c_{l,1} + (\beta_{lk,0} - \beta_{lk}),$$

$$\gamma_{lk} = \sum_{l=m+1}^m(\lambda_{lk,0} - \lambda_{lk})c_{l,2} + (\gamma_{lk,0} - \gamma_{lk}),$$

as long as $\mathbf{Q}'[E(\tilde{y}_1), \ldots, E(\tilde{y}_m), X, \tilde{X}]$ has full column rank for large enough $n$. Then, $\theta_{k,0}$ can be identified if the quadratic moment equations

$$\lim_{n \to \infty} n^{-1} E[f_{2,kl}(\theta_k, \theta_l)] = 0, \quad \text{for } k, l = 1, \ldots, m,$$

where the $r$-th element of $E[f_{2,kl}(\theta_k, \theta_l)]$ is

$$E[u_k(\theta_k)'\Xi_r u_l(\theta_l)] = d_k(\theta_k)'\Xi_r d_l(\theta_l) + \sum_{s=1}^m (\lambda_{sk,0} - \lambda_{sk}) tr[\Sigma_r H_s (\sigma_l \otimes I_n)]$$

$$+ \sum_{t=1}^m (\lambda_{tl,0} - \lambda_{tl}) tr[\Xi_r H_t (\sigma_k \otimes I_n)]$$

$$+ \sum_{s=1}^m \sum_{t=1}^m (\lambda_{sk,0} - \lambda_{sk})(\lambda_{tl,0} - \lambda_{tl}) tr[H_s \Xi_r H_t (\Sigma \otimes I_n)].$$
have a unique solution at $\mathbf{A}_0$. With $(\lambda_{1k}, \ldots, \lambda_{mk}, \beta_k', \gamma_k')$ given by (C.1), $d_k(\theta_k) = 0$. Then, Proposition 2.2 follows by the discussion in Section 2.2.2 and Theorem 2.1 of Hansen (1982). Proposition 3.2 follows by a similar argument.

Proof of Proposition 4.1. First, we consider the infeasible fixed-effect 3SLS estimator $\tilde{\theta}_{3sls} = [\mathbf{Z}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{Z}]^{-1} \mathbf{Z}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{y}$.

$$\sqrt{n}(\tilde{\theta}_{3sls} - \theta_0 - \mathbf{b}_{3sls}) = \left[\frac{1}{n} \mathbf{Z}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{Z}\right]^{-1} \frac{1}{\sqrt{n}} \{\mathbf{Z}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{u} - \mathbb{E}[\mathbf{V}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{u}]\}.$$ 

As $(\mathbf{I}_m \otimes \mathbf{J})\mathbf{Z} = \mathbf{F} + \mathbf{V}$, we have

$$\frac{1}{n} \mathbf{Z}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{Z} = \frac{1}{n} \mathbf{F}'(\Sigma^{-1} \otimes \mathbf{I}_n)\mathbf{F} - \frac{1}{n} \mathbf{F}'(\Sigma^{-1} \otimes (\mathbf{I}_n - \mathbf{P}))\mathbf{F}$$

$$+ \frac{1}{n} \mathbf{F}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{V} + \frac{1}{n} \mathbf{V}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{F} + \frac{1}{n} \mathbf{V}(\Sigma^{-1} \otimes \mathbf{P})\mathbf{V}$$

and

$$\frac{1}{\sqrt{n}} \left\{\mathbf{Z}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{u} - \mathbb{E}[\mathbf{V}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{u}]\right\}$$

$$= \frac{1}{\sqrt{n}} \mathbf{F}'(\Sigma^{-1} \otimes \mathbf{I}_n)\mathbf{u} - \frac{1}{\sqrt{n}} \mathbf{F}'(\Sigma^{-1} \otimes (\mathbf{I}_n - \mathbf{P}))\mathbf{u} + \frac{1}{\sqrt{n}} \{\mathbf{V}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{u} - \mathbb{E}[\mathbf{V}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{u}]\}.$$ 

By Lemma C.3 of Liu (2014), if $q/n \to 0$, $n^{-1} \mathbf{Z}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{Z} = n^{-1} \mathbf{F}'(\Sigma^{-1} \otimes \mathbf{I}_n)\mathbf{F} + o_p(1)$ and $n^{-1/2} \{\mathbf{Z}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{u} - \mathbb{E}[\mathbf{V}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{u}]\} = n^{-1/2} \mathbf{F}'(\Sigma^{-1} \otimes \mathbf{I}_n)\mathbf{u} + o_p(1)$. By Theorem A of Kelejian and Prucha (1999) and Slutsky’s Theorem, $\sqrt{n}(\tilde{\theta}_{3sls} - \theta_0 - \mathbf{b}_{3sls}) \overset{d}{\to} N(0, \lim_{n \to \infty} n^{-1} \mathbf{F}'(\Sigma^{-1} \otimes \mathbf{I}_n)\mathbf{F}^{-1})$. By Lemma C.2 of Liu (2014), $\mathbb{E}[\mathbf{V}'(\Sigma^{-1} \otimes \mathbf{P})\mathbf{u}] = O(q)$, which implies $\mathbf{b}_{3sls} = O_p(q/n)$.

---

24 A weaker identification condition requires that the equations (C.1) and $\lim_{n \to \infty} n^{-1} \mathbb{E}[\mathbf{f}_{2,l}(\theta_k, \theta_l)] = 0$, for $l = 1, \ldots, m$, have a unique solution at $\theta_{k,0}$. 

---

48
To obtain the desired result, it is sufficient to show that $\sqrt{n}(\hat{\theta}_{3als} - \tilde{\theta}_{3als}) = o_p(1)$.

$$\sqrt{n}(\hat{\theta}_{3als} - \tilde{\theta}_{3als})$$

$$= \left[ \frac{1}{n} Z'(\Sigma^{-1} \otimes \tilde{P})Z \right]^{-1} \left[ \frac{1}{n} Z'[\sqrt{n}(\Sigma^{-1} - \Sigma^{-1}) \otimes \tilde{P}]u \right]$$

$$- \frac{1}{n} Z'(\Sigma^{-1} \otimes \tilde{P})Z \left( \frac{1}{n} Z'[\sqrt{n}(\Sigma^{-1} - \Sigma^{-1}) \otimes \tilde{P}]Z \right)^{-1} \frac{1}{n} Z'(\Sigma^{-1} \otimes \tilde{P})u.$$ 

As $\sqrt{n}(\Sigma^{-1} - \Sigma^{-1}) = o_p(1)$, it follows by Lemma C.3 of Liu (2014) that, if $q/n \to 0$, $n^{-1}Z'(\Sigma^{-1} \otimes \tilde{P})Z = o_p(1)$, $n^{-1}Z'[\sqrt{n}(\Sigma^{-1} - \Sigma^{-1}) \otimes \tilde{P}]Z = o_p(1)$, and $n^{-1}Z'[\sqrt{n}(\Sigma^{-1} - \Sigma^{-1}) \otimes \tilde{P}]u = o_p(1)$. The desired result follows.

Proof of Proposition 4.2. First, we consider the infeasible fixed-effect GMM estimator $\hat{\theta}_{gmm} = \arg \min u(\theta)'(\Sigma^{-1} \otimes \tilde{P})u(\theta) + g_x(\theta)'\Omega_{xx}^{-1}g_x(\theta)$. If $Ju(\theta) = Ju$ and $r(\theta) = V(\theta_\theta - \theta) + Ju$. When $q/n \to 0$, it follows by Lemma C.3 of Liu (2014) that

$$\frac{1}{n} u(\theta)'(\Sigma^{-1} \otimes \tilde{P})u(\theta)$$

$$= \frac{1}{n} d(\theta)'(\Sigma^{-1} \otimes \tilde{P})d(\theta) + 2\frac{1}{n} r(\theta)'(\Sigma^{-1} \otimes \tilde{P})d(\theta) + \frac{1}{n} r(\theta)'(\Sigma^{-1} \otimes \tilde{P})r(\theta)$$

$$= \frac{1}{n} (\theta_0 - \theta)'F'(\Sigma^{-1} \otimes I_n)F(\theta_0 - \theta) - \frac{1}{n} (\theta_0 - \theta)'F'[\Sigma^{-1} \otimes (I_n - \tilde{P})]F(\theta_0 - \theta)$$

$$+ 2\frac{1}{n} r(\theta)'(\Sigma^{-1} \otimes \tilde{P})r(\theta) + \frac{1}{n} r(\theta)'(\Sigma^{-1} \otimes \tilde{P})r(\theta)$$

$$= \frac{1}{n} (\theta_0 - \theta)'F'(\Sigma^{-1} \otimes I_n)F(\theta_0 - \theta) + o_p(1),$$

uniformly in $\theta$. Suppose $F(\theta_0 - \theta) = F^*(\theta_0 - \theta) + F^*C(\theta_0 - \theta)$ for some constant matrix $C$, where the elements of $\theta_0$ correspond to the linearly independent columns of $F$. Then, $n^{-1}(\theta_0 - \theta)'F'(\Sigma^{-1} \otimes I_n)F(\theta_0 - \theta) = n^{-1}(\theta_0 - \theta)'F^*(\Sigma^{-1} \otimes I_n)F^*(\theta_0 - \theta) + n^{-1}F^*[C(\theta_0 - \theta)]$. By Assumption 5, $\lim_{n \to \infty} n^{-1}F^*[\Sigma^{-1} \otimes I_n]F^*$ is nonsingular. Hence, $\lim_{n \to \infty} n^{-1}(\theta_0 - \theta)'F'(\Sigma^{-1} \otimes I_n)F(\theta_0 - \theta) \geq 0$, with equality if and only if $\theta_0 = \theta_0 + C(\theta_0 - \theta)^2$.

On the other hand, it follows by a similar argument as in the proof of Proposition 1 in Lee (2007b) that $n^{-1}g_x(\theta) - n^{-1}E[g_x(\theta)] = o_p(1)$ uniformly in $\theta$. When $q/n \to 0$, it follows by Lemma
As shown above, \( n^{-1}g_2^*(\theta) - n^{-1}\hat{g}_2^*(\theta) = o_p(1) \) uniformly in \( \theta \). Hence, \( n^{-1}g_2^*(\theta) - n^{-1}\hat{g}_2^*(\theta) = o_p(1) \) uniformly in \( \theta \), where \( \hat{g}_2^*(\theta) = E[g_2(\theta)] - [(\mu_2^t \Sigma^{-1}) \otimes \omega] F[\theta_0 - \theta] - \frac{1}{n}([\mu_2^t \Sigma^{-1}) \otimes (\omega \tilde{P})] r(\theta) \). As \( \hat{g}_2^*(\theta) \) is a quadratic function of \( \theta \) and the parameter space of \( \theta \) is bounded, \( n^{-1}\hat{g}_2^*(\theta) \) is uniformly equicontinuous in \( \theta \). The identification condition and uniform equicontinuity of \( n^{-1}\hat{g}_2^*(\theta) \) imply that the identification uniqueness condition for \( n^{-1}\hat{g}_2^*(\theta) \) must be satisfied. The consistency of \( \theta_{gmm} \) follows from the uniform convergence and identification uniqueness condition for \( \lim_{n \to \infty} n^{-1}[u(\theta)'(\Sigma^{-1} \otimes \tilde{P})u(\theta) + g_2^*(\theta)'\Omega_{22}^{-1}g_2^*(\theta)] \) as shown above (White, 1994).

It remains to show that \( n^{-1}u(\theta)'[(\hat{\Sigma}^{-1} - \Sigma^{-1}) \otimes \tilde{P}]u(\theta) = o_p(1) \) and \( n^{-1}\hat{g}_2^*(\theta)'\hat{\Omega}_{22}^{-1}\hat{g}_2^*(\theta) - n^{-1}g_2^*(\theta)'\Omega_{22}^{-1}g_2^*(\theta) = o_p(1) \) uniformly in \( \theta \). By a similar argument as above, when \( q/n \to 0 \),

\[
\frac{1}{n}u(\theta)'[(\hat{\Sigma}^{-1} - \Sigma^{-1}) \otimes \tilde{P}]u(\theta) = \frac{1}{n}[\theta_0^{(1)} - \theta^{(1)} + C(\theta_0^{(2)} - \theta^{(2)}))] F'[\theta_0^{(1)} - \theta^{(1)} + C(\theta_0^{(2)} - \theta^{(2)}) + o_p(1),
\]

uniformly in \( \theta \). As \( \hat{\Sigma}^{-1} - \Sigma^{-1} = o_p(1), n^{-1}u(\theta)'[(\hat{\Sigma}^{-1} - \Sigma^{-1}) \otimes \tilde{P}]u(\theta) = o_p(1) \) uniformly in \( \theta \) by Assumption 5’. On the other hand,

\[
\frac{1}{n} \hat{g}_2^*(\theta)'\hat{\Omega}_{22}^{-1}\hat{g}_2^*(\theta) - \frac{1}{n}g_2^*(\theta)'\Omega_{22}^{-1}g_2^*(\theta) = \frac{1}{n} [\hat{g}_2^*(\theta) - g_2^*(\theta)] (\frac{1}{n}\hat{\Omega}_{22})^{-1} \hat{g}_2^*(\theta) + \frac{1}{n} g_2^*(\theta)'(\frac{1}{n}\hat{\Omega}_{22})^{-1} \frac{1}{n} [\hat{g}_2^*(\theta) - g_2^*(\theta)] \\
+ \frac{1}{n} g_2^*(\theta)'(\frac{1}{n}\hat{\Omega}_{22})^{-1} - (\frac{1}{n}\Omega_{22})^{-1} \frac{1}{n} g_2^*(\theta).
\]

As shown above, \( n^{-1}g_2^*(\theta) - n^{-1}\hat{g}_2^*(\theta) = o_p(1) \) uniformly in \( \theta \). By a similar argument as in the proof
of Proposition 2 in Lee (2007b), \( n^{-1}E[g_2(\theta)] = O(1) \) uniformly in \( \theta \). Therefore, \( n^{-1}\tilde{g}_2^*(\theta) = O(1) \) uniformly in \( \theta \) by Assumption 8, which implies \( n^{-1}g_2(\theta) = O_p(1) \) uniformly in \( \theta \).

\[
\frac{1}{n}[\tilde{g}_2^*(\theta) - g_2^*(\theta)] = -\frac{1}{n}[(\tilde{\mu}_2^*\tilde{\Sigma}^{-1} - \mu_2^*\Sigma^{-1}) \otimes (\omega'\tilde{P})]u(\theta) = -\frac{1}{n}[(\tilde{\mu}_2^*\tilde{\Sigma}^{-1} - \mu_2^*\Sigma^{-1}) \otimes (\omega'\tilde{P})]u(\theta).
\]

When \( q/n \to 0 \), it follows by Lemma C.3 of Liu (2014) that \( n^{-1}[\tilde{g}_2^*(\theta) - g_2^*(\theta)] = O_p(1) \) uniformly in \( \theta \). As \( n^{-1}\tilde{\Omega}_2^* - n^{-1}\Omega_2^* = o_p(1) \), \( n^{-1}\tilde{g}_2^*(\theta)'\tilde{\Omega}_2^*-1\tilde{g}_2^*(\theta) - n^{-1}g_2(\theta)'\Omega_2^*-1g_2(\theta) = o_p(1) \) uniformly in \( \theta \) by Assumption 8. The desired result follows.

\[\Box\]

**Proof of Proposition 4.3.** The Taylor expansion of

\[-Z' (\tilde{\Sigma}^{-1} \otimes \tilde{P}) u(\tilde{\theta}_{gmm}) + \frac{\partial \tilde{g}_2^*(\tilde{\theta}_{gmm})}{\partial \theta}' \tilde{\Omega}_2^* \tilde{g}_2^*(\tilde{\theta}_{gmm}) = 0\]

around \( \theta_0 \) gives \( \sqrt{n}(\tilde{\theta}_{gmm} - \theta_0) = \tilde{A}^{-1}\tilde{b} \), where

\[
\tilde{A} = \frac{1}{n} Z' (\tilde{\Sigma}^{-1} \otimes \tilde{P}) Z + \frac{1}{n} \frac{\partial \tilde{g}_2^*(\tilde{\theta}_{gmm})}{\partial \theta}' \tilde{\Omega}_2^* \frac{\partial \tilde{g}_2^*(\tilde{\theta}_{gmm})}{\partial \theta},
\]

\[
\tilde{b} = \frac{1}{\sqrt{n}} Z' (\tilde{\Sigma}^{-1} \otimes \tilde{P}) u - \frac{1}{\sqrt{n}} \frac{\partial \tilde{g}_2^*(\tilde{\theta}_{gmm})}{\partial \theta}' \tilde{\Omega}_2^* \tilde{g}_2^*(\theta_0),
\]

for some \( \theta^* \) between \( \tilde{\theta}_{gmm} \) and \( \theta_0 \). As \( \tilde{\Sigma} - \Sigma = O_p(n^{-1/2}) \), it follows by Lemma C.3 of Liu (2014) that, if \( q/n \to 0 \),

\[
\frac{1}{n} Z' (\tilde{\Sigma}^{-1} \otimes \tilde{P}) Z = \frac{1}{n} F' (\tilde{\Sigma}^{-1} \otimes I_n) F - \frac{1}{n} F' (\tilde{\Sigma}^{-1} \otimes (I_n - \tilde{P})) F
\]

\[+ \frac{1}{n} F' (\tilde{\Sigma}^{-1} \otimes \tilde{P}) V + \frac{1}{n} V' (\tilde{\Sigma}^{-1} \otimes \tilde{P}) F + \frac{1}{n} V' (\tilde{\Sigma}^{-1} \otimes \tilde{P}) V = \frac{1}{n} F' (\Sigma^{-1} \otimes I_n) F + O_p(\sqrt{q/n}). \]
\[
\frac{\partial}{\partial \theta_k} \hat{g}^*_2(\theta) = \frac{\partial}{\partial \theta} g_2(\theta) + [(\hat{\mu}'_3 \hat{\Sigma}^{-1}) \otimes (\omega' \hat{P})] Z. \]

For a typical element of \( g_2(\theta) \), we have

\[
\frac{\partial u_k(\theta)}{\partial \theta_k} - \hat{F} \hat{F}' u_l = -Z_k \hat{F}' u_l(\theta). \tag{C.2}
\]

It follows from Lemmas A.4 and A.5 of Lee (2007b) that \( n^{-1} \frac{\partial}{\partial \theta} g_2(\tilde{\theta}) = -n^{-1} D_2 + o_p(1) \) for \( \tilde{\theta} = \theta_0 + o_p(1) \). As \( \hat{\Sigma} - \Sigma = O_p(n^{-1/2}) \) and \( \hat{\mu}_3 - \mu_3 = O_p(n^{-1/2}) \), it follows by Lemma C.3 of Liu (2014) that, if \( q/n \to 0 \),

\[
\frac{1}{n} [((\hat{\mu}'_3 \hat{\Sigma}^{-1}) \otimes (\omega' \hat{P})] Z = \frac{1}{n} [((\hat{\mu}'_3 \hat{\Sigma}^{-1}) \otimes \omega')] F - \frac{1}{n} [((\hat{\mu}'_3 \hat{\Sigma}^{-1}) \otimes (\omega') \hat{P})] V + \frac{1}{n} [((\hat{\mu}'_3 \hat{\Sigma}^{-1}) \otimes (\omega' \hat{P})] V
\]

\[
= \frac{1}{n} [((\mu'_3 \Sigma^{-1}) \otimes \omega')] F + o_p(\sqrt{q/n}).
\]

Therefore, \( n^{-1} \frac{\partial}{\partial \theta} \hat{g}^*_2(\tilde{\theta}) = -n^{-1} D_2 + o_p(1) \) for \( \tilde{\theta} = \theta_0 + o_p(1) \), which implies that

\[
n^{-1} \frac{\partial}{\partial \theta} \hat{g}^*_2(\hat{\theta}_{gmm}) - \frac{\partial}{\partial \theta} \hat{g}^*_2(\hat{\theta}_{gmm}) = o_p(1) \text{ since } n^{-1/2}(\hat{\Sigma}_{22} - \Sigma_{22}) = O_p(1).
\]

In summary,

\[
\hat{A} = \frac{1}{n} [F' \Sigma^{-1} \otimes I_n] F + D_2' \Omega_{22}^{-1} D_2^* + o_p(1). \tag{C.3}
\]

As \( \hat{\Sigma} - \Sigma = O_p(n^{-1/2}) \) and \( \hat{\mu}_3 - \mu_3 = O_p(n^{-1/2}) \), it follows by Lemma C.3 of Liu (2014) that, if \( q/n \to 0 \),

\[
\frac{1}{\sqrt{n}} Z' (\hat{\Sigma}^{-1} \otimes \hat{P}) u
\]

\[
= \frac{1}{\sqrt{n}} F' \Sigma^{-1} \otimes I_n u - \frac{1}{\sqrt{n}} F' (\hat{\Sigma}^{-1} \otimes (I_n - \hat{P})) u + \frac{1}{\sqrt{n}} V' (\Sigma^{-1} \otimes \hat{P}) u
\]

\[
= \frac{1}{\sqrt{n}} F' (\Sigma^{-1} \otimes I_n) u + \frac{1}{\sqrt{n}} E[V' (\Sigma^{-1} \otimes \hat{P}) u] + o_p(1),
\]

and \( n^{-1/2} \hat{g}^*_2(\theta_0) - n^{-1/2} g^*_2(\theta_0) = -n^{-1/2} \sqrt{n}(\hat{\mu}'_3 \hat{\Sigma}^{-1} - \mu'_3 \Sigma^{-1}) \otimes (\omega' \hat{P})] u = o_p(1) \). As

\[
\frac{1}{n} E[g^*_2(\theta_0) u' (\Sigma^{-1} \otimes I_n) F] = \frac{1}{n} [(\mu'_3 \Sigma^{-1}) \otimes (\omega' (I_n - \hat{P}))] F = o_p(1),
\]

52
$n^{-1/2} F'(\Sigma^{-1} \otimes I_n) u$ and $n^{-1/2} g^*_2(\theta_0)$ are asymptotically uncorrelated. It follows by Lemma 3 of Yang and Lee (2014) that

$$
\hat{b} - \frac{1}{\sqrt{n}} E[V'(\Sigma^{-1} \otimes \bar{P}) u] \overset{d}{\to} N(0, \lim_{n \to \infty} \frac{1}{n} [F'(\Sigma^{-1} \otimes I_n) F + D'_2 \Omega^{-1}_{22} D'_2]).
$$

(C.4)

As $E[V'(\Sigma^{-1} \otimes \bar{P}) u] = O(q)$, from (C.3) and (C.4), we have $\sqrt{n}(\hat{\theta}_{gmm} - \theta_0) = O_p(q/\sqrt{n})$, or $\hat{\theta}_{gmm} - \theta_0 = O_p(q/n)$.

It follows from (C.2) and Lemmas A.4 and A.5 of Lee (2007b) that, if $q/n \to 0$, we have $n^{-1} \frac{\partial}{\partial \theta} g_2(\tilde{\theta}) = -n^{-1} D_2 + O_p(\max\{1/\sqrt{n}, q/n\})$ for $\tilde{\theta} - \theta_0 = O_p(q/n)$, which implies that $n^{-1} \frac{\partial}{\partial \theta} g^*_2(\tilde{\theta}) = -n^{-1} D_2 + O_p(\sqrt{q/n})$. Hence,

$$
\hat{A} = \frac{1}{n} [F'(\Sigma^{-1} \otimes I_n) F + D'_2 \Omega^{-1}_{22} D'_2] + O_p(\sqrt{q/n}).
$$

(C.5)

From (C.4) and (C.5),

$$
\sqrt{n}(\hat{\theta}_{gmm} - \theta_0 - b_{gmm})
= -[\frac{1}{n} \frac{\partial g(\hat{\theta}_{gmm})}{\partial \theta} \Omega^{-1}_g \frac{\partial g(\theta_0^+)}{\partial \theta^+}]^{-1} \frac{1}{\sqrt{n}} \frac{\partial g(\hat{\theta}_{gmm})}{\partial \theta} \Omega^{-1}_g + E[V'(\Sigma^{-1} \otimes \bar{P}) u] + O_p(\sqrt{q^3/n^2}).
$$

Hence, if $q^3/n \to 0$, $\sqrt{n}(\hat{\theta}_{gmm} - \theta_0 - b_{gmm}) \overset{d}{\to} N(0, \lim_{n \to \infty} n [F'(\Sigma^{-1} \otimes I_n) F + D'_2 \Omega^{-1}_{22} D'_2]^{-1})$.

\[\Box\]

**Lemma C.1.** If $q/n \to 0$ then $\sqrt{n}(\hat{b}_{gmm} - b_{gmm}) = o_p(1)$, where $\hat{b}_{gmm}$ and $b_{gmm}$ are given by (4.12) and (4.11) respectively.

**Proof.** To show the desired result, it is sufficient to show that $n^{-1} Z(\hat{\Sigma}^{-1} \otimes \bar{P}) Z - n^{-1} F'(\Sigma^{-1} \otimes I_n) F = o_p(1)$, $n^{-1} \hat{D}_2 - n^{-1} D_2 = o_p(1)$, $n^{-1} \hat{\Omega}^*_2 - n^{-1} \Omega^*_2 = o_p(1)$, and $n^{-1/2} E[V'(\Sigma^{-1} \otimes \bar{P}) u] - n^{-1/2} E[V'(\Sigma^{-1} \otimes \bar{P}) u] = o_p(1)$. By a similar argument as in the proof of Proposition 4.1, if $q/n \to 0$ then $n^{-1} Z(\hat{\Sigma}^{-1} \otimes \bar{P}) Z - n^{-1} F'(\Sigma^{-1} \otimes I_n) F = o_p(1)$. As $n^{-1} \omega' \omega = O(1)$, $n^{-1} \Delta_1 = O(1)$, $n^{-1} \Delta_2 = O(1)$, and $n^{-1} \omega' \bar{P} \omega = O(1)$, we have $n^{-1} \hat{\Omega}^*_2 - n^{-1} \Omega^*_2 = o_p(1)$. As shown in the proof of Proposition 9 in Liu (2014), if $q/n \to 0$ then $n^{-1/2} E[V'(\Sigma^{-1} \otimes \bar{P}) u] - n^{-1/2} E[V'(\Sigma^{-1} \otimes \bar{P}) u] = 53$


$o_p(1)$. Hence, it only remains to show that $n^{-1}\hat{D}_2 - n^{-1}D_2 = o_p(1)$.

$$D_2 = -E\left[ \frac{\partial}{\partial \theta} g_2(\theta_0) \right] = \begin{bmatrix} \gamma_{1,11} & 0 \\ \gamma_{1,12} & 0 \\ 0 & \gamma_{1,21} \\ 0 & \gamma_{1,22} \end{bmatrix} + \begin{bmatrix} \gamma_{2,11} & 0 \\ 0 & \gamma_{2,12} \\ \gamma_{2,21} & 0 \\ 0 & \gamma_{2,22} \end{bmatrix},$$

where $\gamma_{i,k,l} = [E(V_i^t \hat{Z}_i u_k), \cdots, E(V_i^t \hat{Z}_i u_k)]'$ and $\gamma_{2,k,l} = [E(V_i^t \hat{Z}_i u_k), \cdots, E(V_i^t \hat{Z}_i u_k)]'$ for $k, l = 1, 2$. With $V_1$ and $V_2$ given by (4.4),

$$E(V_1^t A u_1) = \begin{bmatrix} (\sigma_{12} + \phi_{12,0} \sigma_{11}) tr(A'S^{-1}) + (\lambda_{12,0} \sigma_{11} - \lambda_{11,0} \sigma_{12}) tr(A'W S^{-1}) \\ (\sigma_{11} + \phi_{21,0} \sigma_{12}) tr(A'W S^{-1}) + (\lambda_{21,0} \sigma_{12} - \lambda_{22,0} \sigma_{11}) tr(A'W^2 S^{-1}) \\ (\sigma_{12} + \phi_{12,0} \sigma_{11}) tr(A'W S^{-1}) + (\lambda_{12,0} \sigma_{11} - \lambda_{11,0} \sigma_{12}) tr(A'W^2 S^{-1}) \\ 0_{2 \times 1} \end{bmatrix}$$

$$E(V_2^t A u_2) = \begin{bmatrix} (\sigma_{12} + \phi_{21,0} \sigma_{22}) tr(A'S^{-1}) + (\lambda_{21,0} \sigma_{22} - \lambda_{22,0} \sigma_{12}) tr(A'W S^{-1}) \\ (\sigma_{12} + \phi_{21,0} \sigma_{22}) tr(A'W S^{-1}) + (\lambda_{21,0} \sigma_{22} - \lambda_{22,0} \sigma_{12}) tr(A'W^2 S^{-1}) \\ (\sigma_{22} + \phi_{12,0} \sigma_{12}) tr(A'W S^{-1}) + (\lambda_{12,0} \sigma_{12} - \lambda_{11,0} \sigma_{22}) tr(A'W^2 S^{-1}) \\ 0_{2 \times 1} \end{bmatrix}$$

$$E(V_1^t A u_2) = \begin{bmatrix} (\sigma_{22} + \phi_{12,0} \sigma_{12}) tr(A'S^{-1}) + (\lambda_{12,0} \sigma_{12} - \lambda_{11,0} \sigma_{22}) tr(A'W S^{-1}) \\ (\sigma_{12} + \phi_{21,0} \sigma_{22}) tr(A'W S^{-1}) + (\lambda_{21,0} \sigma_{22} - \lambda_{22,0} \sigma_{12}) tr(A'W^2 S^{-1}) \\ (\sigma_{22} + \phi_{12,0} \sigma_{12}) tr(A'W S^{-1}) + (\lambda_{12,0} \sigma_{12} - \lambda_{11,0} \sigma_{22}) tr(A'W^2 S^{-1}) \\ 0_{2 \times 1} \end{bmatrix}$$

$$E(V_2^t A u_1) = \begin{bmatrix} (\sigma_{11} + \phi_{21,0} \sigma_{12}) tr(A'S^{-1}) + (\lambda_{21,0} \sigma_{12} - \lambda_{22,0} \sigma_{11}) tr(A'W S^{-1}) \\ (\sigma_{11} + \phi_{21,0} \sigma_{12}) tr(A'W S^{-1}) + (\lambda_{21,0} \sigma_{12} - \lambda_{22,0} \sigma_{11}) tr(A'W^2 S^{-1}) \\ (\sigma_{12} + \phi_{12,0} \sigma_{11}) tr(A'W S^{-1}) + (\lambda_{12,0} \sigma_{11} - \lambda_{11,0} \sigma_{12}) tr(A'W^2 S^{-1}) \\ 0_{2 \times 1} \end{bmatrix}$$
where $A$ is either $\hat{\Xi}_r$ or $\hat{\Xi}_L$. As $n^{-1}\text{tr}(AS^{-1})$, $n^{-1}\text{tr}(AWS^{-1})$ and $n^{-1}\text{tr}(AW^2S^{-1})$ are $O(1)$, we have $n^{-1}\hat{D}_2 - n^{-1}D_2 = o_p(1)$, where $\hat{D}_2$ is an estimator of $D_2$ by replacing unknown parameters in $D_2$ by their consistent estimators. It follows by Lemma C.3 of Liu (2014) that, if $q/n \to 0$, $n^{-1}\omega'\hat{P}Z_k = n^{-1}\omega'F_k + n^{-1}\omega'(I_n - \hat{P})F_k + n^{-1}\omega'\hat{P}V_k = n^{-1}\omega'F_k + o_p(1) = O_p(1)$. Hence, $n^{-1}\hat{D}_2 = n^{-1}\hat{D}_2 - n^{-1}[(\hat{\mu}'\hat{\Sigma}^{-1}) \otimes \omega'\hat{P}]Z = n^{-1}D_2 + o_p(1)$. The desired result follows. \qed
Table 1: 3SLS and GMM Estimation ($n_g = 10$, $G = 30$)

<table>
<thead>
<tr>
<th></th>
<th>$\phi_{21.0} = 0.2$</th>
<th>$\lambda_{11.0} = 0.1$</th>
<th>$\lambda_{21.0} = 0.1$</th>
<th>$\beta_{1.0} = 0.6$</th>
<th>$\gamma_{1.0} = 0.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{12} = 0.1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.198(.069)[.069]</td>
<td>.096(.046)[.046]</td>
<td>.104(.045)[.045]</td>
<td>.603(.062)[.062]</td>
<td>.603(.057)[.057]</td>
</tr>
<tr>
<td>3SLS-2</td>
<td>.222(.069)[.072]</td>
<td>.087(.035)[.037]</td>
<td>.098(.035)[.035]</td>
<td>.598(.060)[.060]</td>
<td>.603(.052)[.052]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.194(.070)[.070]</td>
<td>.098(.037)[.037]</td>
<td>.104(.036)[.037]</td>
<td>.604(.061)[.061]</td>
<td>.602(.053)[.053]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.201(.069)[.069]</td>
<td>.095(.038)[.039]</td>
<td>.101(.036)[.036]</td>
<td>.602(.062)[.062]</td>
<td>.600(.055)[.055]</td>
</tr>
<tr>
<td>GMM-2</td>
<td>.225(.069)[.074]</td>
<td>.088(.032)[.035]</td>
<td>.097(.031)[.031]</td>
<td>.597(.060)[.061]</td>
<td>.600(.052)[.052]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.197(.070)[.070]</td>
<td>.097(.034)[.034]</td>
<td>.103(.032)[.032]</td>
<td>.603(.061)[.061]</td>
<td>.600(.053)[.053]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.199(.069)[.069]</td>
<td>.096(.044)[.045]</td>
<td>.102(.045)[.045]</td>
<td>.605(.058)[.058]</td>
<td>.603(.055)[.055]</td>
</tr>
<tr>
<td>3SLS-2</td>
<td>.224(.069)[.073]</td>
<td>.087(.035)[.037]</td>
<td>.095(.035)[.035]</td>
<td>.605(.056)[.056]</td>
<td>.608(.050)[.050]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.195(.073)[.073]</td>
<td>.099(.036)[.036]</td>
<td>.103(.037)[.037]</td>
<td>.604(.058)[.058]</td>
<td>.602(.052)[.052]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.200(.071)[.071]</td>
<td>.094(.039)[.040]</td>
<td>.101(.039)[.039]</td>
<td>.603(.057)[.057]</td>
<td>.602(.053)[.053]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.197(.073)[.073]</td>
<td>.097(.035)[.035]</td>
<td>.102(.034)[.034]</td>
<td>.603(.058)[.058]</td>
<td>.601(.052)[.052]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.200(.070)[.070]</td>
<td>.096(.041)[.041]</td>
<td>.102(.044)[.044]</td>
<td>.606(.047)[.047]</td>
<td>.605(.046)[.047]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.198(.075)[.075]</td>
<td>.098(.035)[.035]</td>
<td>.103(.038)[.038]</td>
<td>.605(.049)[.049]</td>
<td>.604(.049)[.049]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.198(.075)[.075]</td>
<td>.097(.035)[.035]</td>
<td>.102(.037)[.037]</td>
<td>.605(.049)[.049]</td>
<td>.604(.048)[.049]</td>
</tr>
</tbody>
</table>

Mean(SD)[RMSE]
Table 2: 3SLS and GMM Estimation ($n_g = 15$, $G = 30$)

<table>
<thead>
<tr>
<th></th>
<th>$\phi_{21,0} = 0.2$</th>
<th>$\lambda_{11,0} = 0.1$</th>
<th>$\lambda_{21,0} = 0.1$</th>
<th>$\beta_{1,0} = 0.6$</th>
<th>$\gamma_{1,0} = 0.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{12} = 0.1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.201(.053)[.053]</td>
<td>.099(.034)[.034]</td>
<td>.099(.033)[.033]</td>
<td>.598(.051)[.051]</td>
<td>.603(.046)[.046]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.199(.055)[.055]</td>
<td>.100(.028)[.028]</td>
<td>.100(.029)[.029]</td>
<td>.598(.051)[.051]</td>
<td>.602(.043)[.043]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.204(.052)[.052]</td>
<td>.099(.027)[.027]</td>
<td>.098(.025)[.025]</td>
<td>.597(.051)[.051]</td>
<td>.601(.042)[.042]</td>
</tr>
<tr>
<td>GMM-2</td>
<td>.219(.053)[.057]</td>
<td>.096(.024)[.024]</td>
<td>.095(.023)[.024]</td>
<td>.595(.050)[.050]</td>
<td>.600(.041)[.041]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.201(.054)[.054]</td>
<td>.099(.024)[.024]</td>
<td>.099(.024)[.024]</td>
<td>.597(.050)[.051]</td>
<td>.601(.041)[.041]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.202(.053)[.053]</td>
<td>.099(.033)[.033]</td>
<td>.099(.033)[.033]</td>
<td>.599(.047)[.047]</td>
<td>.603(.043)[.043]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.199(.056)[.056]</td>
<td>.100(.028)[.028]</td>
<td>.099(.029)[.029]</td>
<td>.598(.047)[.047]</td>
<td>.602(.041)[.041]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.204(.052)[.052]</td>
<td>.099(.028)[.028]</td>
<td>.098(.027)[.027]</td>
<td>.598(.047)[.047]</td>
<td>.602(.040)[.040]</td>
</tr>
<tr>
<td>GMM-2</td>
<td>.219(.053)[.057]</td>
<td>.096(.024)[.024]</td>
<td>.094(.024)[.025]</td>
<td>.601(.046)[.046]</td>
<td>.605(.038)[.039]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.201(.055)[.055]</td>
<td>.100(.025)[.025]</td>
<td>.099(.025)[.025]</td>
<td>.598(.047)[.047]</td>
<td>.601(.040)[.040]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.203(.052)[.052]</td>
<td>.100(.030)[.030]</td>
<td>.098(.032)[.032]</td>
<td>.602(.037)[.037]</td>
<td>.603(.036)[.036]</td>
</tr>
<tr>
<td>3SLS-2</td>
<td>.218(.055)[.058]</td>
<td>.095(.026)[.026]</td>
<td>.093(.028)[.028]</td>
<td>.608(.038)[.039]</td>
<td>.610(.037)[.039]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.200(.057)[.057]</td>
<td>.100(.026)[.026]</td>
<td>.099(.029)[.029]</td>
<td>.601(.039)[.039]</td>
<td>.603(.038)[.038]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.203(.053)[.053]</td>
<td>.099(.028)[.028]</td>
<td>.097(.030)[.030]</td>
<td>.602(.037)[.037]</td>
<td>.603(.036)[.036]</td>
</tr>
<tr>
<td>GMM-2</td>
<td>.219(.055)[.058]</td>
<td>.096(.024)[.025]</td>
<td>.094(.026)[.027]</td>
<td>.608(.038)[.039]</td>
<td>.610(.037)[.038]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.201(.057)[.057]</td>
<td>.100(.025)[.025]</td>
<td>.098(.027)[.027]</td>
<td>.601(.039)[.039]</td>
<td>.602(.038)[.038]</td>
</tr>
</tbody>
</table>

Mean(SD)[RMSE]
Table 3: 3SLS and GMM Estimation ($n_g = 15, G = 60$)

<table>
<thead>
<tr>
<th>$\sigma_{12} = 0.1$</th>
<th>$\phi_{21.0} = 0.2$</th>
<th>$\lambda_{11.0} = 0.1$</th>
<th>$\lambda_{21.0} = 0.1$</th>
<th>$\beta_{1.0} = 0.6$</th>
<th>$\gamma_{1.0} = 0.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3SLS-1</td>
<td>.200(.038)[.038]</td>
<td>.099(.023)[.023]</td>
<td>.100(.025)[.025]</td>
<td>.600(.035)[.035]</td>
<td>.600(.032)[.032]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.198(.039)[.039]</td>
<td>.099(.020)[.020]</td>
<td>.101(.021)[.021]</td>
<td>.600(.035)[.035]</td>
<td>.601(.031)[.031]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.201(.037)[.037]</td>
<td>.099(.019)[.019]</td>
<td>.100(.019)[.019]</td>
<td>.600(.035)[.035]</td>
<td>.600(.030)[.030]</td>
</tr>
<tr>
<td>GMM-2</td>
<td>.216(.038)[.041]</td>
<td>.096(.017)[.017]</td>
<td>.097(.018)[.018]</td>
<td>.598(.035)[.035]</td>
<td>.600(.029)[.029]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.200(.038)[.038]</td>
<td>.099(.017)[.017]</td>
<td>.101(.018)[.018]</td>
<td>.600(.035)[.035]</td>
<td>.600(.030)[.030]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.201(.038)[.038]</td>
<td>.099(.022)[.022]</td>
<td>.100(.024)[.024]</td>
<td>.600(.033)[.033]</td>
<td>.601(.030)[.030]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.201(.038)[.038]</td>
<td>.098(.019)[.019]</td>
<td>.100(.021)[.021]</td>
<td>.600(.033)[.033]</td>
<td>.601(.029)[.029]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.199(.040)[.040]</td>
<td>.099(.017)[.017]</td>
<td>.101(.019)[.019]</td>
<td>.599(.033)[.033]</td>
<td>.601(.029)[.029]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.201(.038)[.038]</td>
<td>.099(.021)[.021]</td>
<td>.100(.024)[.024]</td>
<td>.601(.027)[.027]</td>
<td>.602(.026)[.026]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.201(.039)[.039]</td>
<td>.098(.020)[.020]</td>
<td>.100(.023)[.023]</td>
<td>.601(.028)[.028]</td>
<td>.602(.026)[.026]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.200(.042)[.042]</td>
<td>.099(.018)[.018]</td>
<td>.100(.021)[.021]</td>
<td>.600(.028)[.028]</td>
<td>.601(.027)[.027]</td>
</tr>
</tbody>
</table>

Mean(SD)[RMSE]
Table 4: 3SLS and GMM Estimation ($n_g = 10$, $G = 30$)

<table>
<thead>
<tr>
<th></th>
<th>$\phi_{21,0} = 0.2$</th>
<th>$\lambda_{1,0} = 0.1$</th>
<th>$\lambda_{21,0} = 0.1$</th>
<th>$\beta_{1,0} = 0.3$</th>
<th>$\gamma_{1,0} = 0.3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{12} = 0.1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.195(.143)[.143]</td>
<td>.089(.100)[.101]</td>
<td>.108(.097)[.097]</td>
<td>.304(.065)[.065]</td>
<td>.305(.060)[.060]</td>
</tr>
<tr>
<td>3SLS-2</td>
<td>.267(.149)[.164]</td>
<td>.076(.051)[.056]</td>
<td>.093(.052)[.053]</td>
<td>.294(.061)[.061]</td>
<td>.298(.052)[.052]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.212(.144)[.144]</td>
<td>.093(.059)[.060]</td>
<td>.099(.053)[.053]</td>
<td>.302(.062)[.062]</td>
<td>.299(.054)[.054]</td>
</tr>
<tr>
<td>GMM-2</td>
<td>.275(.151)[.169]</td>
<td>.078(.044)[.049]</td>
<td>.089(.042)[.043]</td>
<td>.293(.061)[.061]</td>
<td>.294(.052)[.053]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.192(.150)[.150]</td>
<td>.099(.048)[.048]</td>
<td>.103(.046)[.046]</td>
<td>.305(.062)[.062]</td>
<td>.300(.052)[.052]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.201(.147)[.147]</td>
<td>.091(.096)[.097]</td>
<td>.103(.095)[.095]</td>
<td>.307(.061)[.061]</td>
<td>.305(.057)[.058]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.208(.149)[.149]</td>
<td>.091(.065)[.065]</td>
<td>.100(.061)[.061]</td>
<td>.305(.058)[.058]</td>
<td>.303(.052)[.052]</td>
</tr>
<tr>
<td>GMM-2</td>
<td>.281(.152)[.172]</td>
<td>.077(.045)[.050]</td>
<td>.086(.044)[.046]</td>
<td>.305(.056)[.057]</td>
<td>.305(.048)[.049]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.198(.160)[.160]</td>
<td>.099(.052)[.052]</td>
<td>.103(.053)[.053]</td>
<td>.306(.060)[.060]</td>
<td>.302(.053)[.053]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.205(.149)[.149]</td>
<td>.088(.087)[.088]</td>
<td>.100(.089)[.092]</td>
<td>.310(.049)[.050]</td>
<td>.309(.048)[.049]</td>
</tr>
<tr>
<td>3SLS-2</td>
<td>.283(.154)[.175]</td>
<td>.076(.050)[.055]</td>
<td>.081(.052)[.055]</td>
<td>.324(.050)[.055]</td>
<td>.324(.048)[.053]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.205(.193)[.193]</td>
<td>.092(.250)[.250]</td>
<td>.107(.100)[.101]</td>
<td>.311(.064)[.065]</td>
<td>.310(.053)[.054]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.203(.158)[.158]</td>
<td>.088(.075)[.076]</td>
<td>.102(.080)[.080]</td>
<td>.309(.049)[.050]</td>
<td>.309(.048)[.049]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.207(.159)[.159]</td>
<td>.092(.157)[.158]</td>
<td>.105(.082)[.082]</td>
<td>.311(.052)[.053]</td>
<td>.310(.049)[.050]</td>
</tr>
</tbody>
</table>

Mean(SD)[RMSE]
<table>
<thead>
<tr>
<th>$\sigma_{12} = 0.1$</th>
<th>$\phi_{21.0} = 0.2$</th>
<th>$\lambda_{11.0} = 0.1$</th>
<th>$\lambda_{21.0} = 0.1$</th>
<th>$\beta_{1.0} = 0.3$</th>
<th>$\gamma_{1.0} = 0.3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3SLS-1</td>
<td>.203(.108)[.108]</td>
<td>.097(.070)[.070]</td>
<td>.100(.068)[.068]</td>
<td>.299(.052)[.052]</td>
<td>.304(.046)[.047]</td>
</tr>
<tr>
<td>3SLS-2</td>
<td>.251(.113)[.124]</td>
<td>.092(.040)[.041]</td>
<td>.088(.044)[.046]</td>
<td>.294(.050)[.050]</td>
<td>.299(.040)[.040]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.193(.115)[.115]</td>
<td>.101(.043)[.043]</td>
<td>.100(.046)[.046]</td>
<td>.299(.051)[.051]</td>
<td>.302(.041)[.041]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.210(.104)[.104]</td>
<td>.098(.040)[.040]</td>
<td>.097(.037)[.037]</td>
<td>.298(.051)[.051]</td>
<td>.300(.040)[.040]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.200(.114)[.114]</td>
<td>.100(.033)[.033]</td>
<td>.100(.034)[.034]</td>
<td>.298(.050)[.050]</td>
<td>.301(.039)[.039]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.5$</td>
<td>.205(.107)[.107]</td>
<td>.097(.068)[.068]</td>
<td>.097(.067)[.067]</td>
<td>.300(.048)[.048]</td>
<td>.305(.044)[.044]</td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.254(.113)[.125]</td>
<td>.091(.040)[.041]</td>
<td>.087(.043)[.046]</td>
<td>.302(.046)[.046]</td>
<td>.306(.037)[.038]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.257(.112)[.126]</td>
<td>.091(.033)[.034]</td>
<td>.088(.035)[.037]</td>
<td>.302(.046)[.046]</td>
<td>.305(.036)[.036]</td>
</tr>
<tr>
<td>GMM-2</td>
<td>.200(.117)[.117]</td>
<td>.100(.035)[.035]</td>
<td>.099(.038)[.038]</td>
<td>.299(.047)[.047]</td>
<td>.302(.038)[.038]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.9$</td>
<td>.208(.105)[.106]</td>
<td>.096(.063)[.063]</td>
<td>.095(.066)[.066]</td>
<td>.305(.038)[.038]</td>
<td>.306(.037)[.037]</td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.257(.114)[.127]</td>
<td>.089(.039)[.040]</td>
<td>.086(.043)[.045]</td>
<td>.314(.039)[.041]</td>
<td>.316(.037)[.040]</td>
</tr>
<tr>
<td>3SLS-2</td>
<td>.201(.118)[.118]</td>
<td>.102(.047)[.047]</td>
<td>.099(.047)[.047]</td>
<td>.304(.039)[.040]</td>
<td>.305(.037)[.038]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.208(.110)[.111]</td>
<td>.098(.049)[.049]</td>
<td>.096(.055)[.055]</td>
<td>.304(.038)[.038]</td>
<td>.305(.036)[.036]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.258(.114)[.128]</td>
<td>.090(.035)[.037]</td>
<td>.087(.039)[.041]</td>
<td>.315(.039)[.041]</td>
<td>.316(.036)[.040]</td>
</tr>
<tr>
<td>GMM-2</td>
<td>.203(.117)[.117]</td>
<td>.101(.039)[.039]</td>
<td>.098(.043)[.043]</td>
<td>.304(.039)[.039]</td>
<td>.305(.037)[.037]</td>
</tr>
</tbody>
</table>

Mean(SD)[RMSE]
### Table 6: 3SLS and GMM Estimation ($n_g = 15$, $G = 60$)

<table>
<thead>
<tr>
<th></th>
<th>$\phi_{21,0} = 0.2$</th>
<th>$\lambda_{11,0} = 0.1$</th>
<th>$\lambda_{21,0} = 0.1$</th>
<th>$\beta_{1,0} = 0.3$</th>
<th>$\gamma_{1,0} = 0.3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{12} = 0.1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.201(.077)[.077]</td>
<td>.098(.047)[.047]</td>
<td>.101(.050)[.050]</td>
<td>.300(.036)[.036]</td>
<td>.301(.032)[.032]</td>
</tr>
<tr>
<td>3SLS-2</td>
<td>.246(.080)[.092]</td>
<td>.090(.028)[.030]</td>
<td>.092(.031)[.032]</td>
<td>.297(.035)[.035]</td>
<td>.299(.029)[.029]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.195(.080)[.080]</td>
<td>.099(.029)[.029]</td>
<td>.102(.032)[.032]</td>
<td>.301(.035)[.035]</td>
<td>.301(.029)[.029]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.205(.073)[.074]</td>
<td>.098(.028)[.028]</td>
<td>.100(.027)[.027]</td>
<td>.300(.035)[.035]</td>
<td>.299(.029)[.029]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.198(.079)[.079]</td>
<td>.099(.023)[.023]</td>
<td>.102(.024)[.024]</td>
<td>.300(.035)[.035]</td>
<td>.300(.028)[.028]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-2</td>
<td>.247(.081)[.094]</td>
<td>.089(.028)[.030]</td>
<td>.091(.030)[.032]</td>
<td>.303(.032)[.032]</td>
<td>.306(.026)[.027]</td>
</tr>
<tr>
<td>BC3SLS</td>
<td>.196(.084)[.084]</td>
<td>.099(.029)[.030]</td>
<td>.102(.033)[.033]</td>
<td>.303(.033)[.033]</td>
<td>.301(.028)[.028]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.203(.075)[.075]</td>
<td>.097(.030)[.030]</td>
<td>.100(.031)[.031]</td>
<td>.300(.033)[.033]</td>
<td>.301(.028)[.028]</td>
</tr>
<tr>
<td>GMM-2</td>
<td>.249(.080)[.094]</td>
<td>.089(.024)[.026]</td>
<td>.092(.025)[.027]</td>
<td>.303(.032)[.032]</td>
<td>.305(.026)[.026]</td>
</tr>
<tr>
<td>BCGMM</td>
<td>.198(.083)[.083]</td>
<td>.099(.025)[.025]</td>
<td>.102(.027)[.027]</td>
<td>.300(.033)[.033]</td>
<td>.301(.028)[.028]</td>
</tr>
<tr>
<td>$\sigma_{12} = 0.9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3SLS-1</td>
<td>.203(.077)[.077]</td>
<td>.097(.042)[.042]</td>
<td>.100(.048)[.048]</td>
<td>.302(.027)[.027]</td>
<td>.303(.026)[.026]</td>
</tr>
<tr>
<td>3SLS-2</td>
<td>.250(.083)[.097]</td>
<td>.087(.028)[.030]</td>
<td>.090(.031)[.032]</td>
<td>.311(.028)[.030]</td>
<td>.313(.026)[.029]</td>
</tr>
<tr>
<td>GMM-1</td>
<td>.201(.081)[.081]</td>
<td>.096(.036)[.036]</td>
<td>.101(.042)[.042]</td>
<td>.302(.028)[.028]</td>
<td>.303(.026)[.026]</td>
</tr>
</tbody>
</table>

Mean(SD)[RMSE]